A cross-linguistic study on the phonetics of dorsal obstruents

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor philosophiae
(Dr. phil.)
im Fach Allgemeine Sprachwissenschaft

eingereicht an der
Philosophischen Fakultät II
Humboldt-Universität zu Berlin

von
Dipl. Psych. Christian Geng
geboren am 08.11.1968 in Friedrichshafen

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Dr. h.c. Christoph Markschies

Dekan der Philosophischen Fakultät II:
Prof. Dr. Dr. h.c. Christoph Markschies

Gutachter:
1. Prof. Dr. Bernd Pompino-Marschall
2. Prof. Dr. Daniel Recasens

eingereicht am: 20 November 2007
Tag der mündlichen Prüfung: 23 Januar 2008

Abstract

This work investigates place of articulation in a cross-linguistic perspective: While most articulatory and in particular perceptual studies within phonetics are refined to the study of the three major - labial, alveolar and velar - places of articulation, the present dissertation aims at the addition of the palatal place of articulation in obstruent production and perception, with special focus on the Hungarian palatal stops. The self-limitation of phonetics to deal only with the three major stop places of articulation in part has practical reasons: Phonemes like the palatal Hungarian obstruents [c] and [ɟ] are not members of the sound systems of languages like English which can be regarded as the drosophila of experimental phonetics. The guiding idea of this research was that the incorporation of such additional phonemes into the planning and design of the experimental studies might, by increasing the category density, drive the categories into a "tug of war" for phonetic "resources" - articulatory or perceptual spaces. Put differently, the architecture of the current dissertation is centered around ways in which the category distance defined in some phonetic - or potentially also phonological space - can be utilized to derive hypotheses which are best tested in a crosslinguistic design. Such designs can for example be helpful to test whether the phoneme inventory of a language leaves its traces in patterns of velar coarticulation. Concerning the so-called loops,[1] there have been numerous publications dealing with the influences of speech physiology, aerodynamics, general movement principles or articulatory biomechanics in a fairly systematic fashion, while no research efforts have been made so far to investigate possible influences of the system of linguistic contrasts of a given language. The same holds for the domain of speech perception: In speech perception, there are open questions which have not been answered yet or only been touched in a rudimentary fashion: How many stop place categories can be implemented on the basis of formant transitions alone? An arbitrary amount or is there an upper limit? In phonetic research on vowels similar questions are fairly common, in the domain of consonants, I am not aware of any empirical efforts.

Experiments will be devised starting from opposite ends of the rope: speech production and speech perception - which do not necessarily have commensurable theoretical startpoints. Working in both the domains of production and perception is costly and redundancy hard to avoid if one wants to arrive at a coherent theoretical treatment of the necessary conceptual ingredients. The solution chosen was to start with a theoretical introductory part (part I) with a separate treatment of (i) perceptual and (ii) articulatory matters. This main thread is split later into separate experimental chapters - chapters 2 and 3 - both of which already belong to the experimental part (part II). These separate additional theoretical chapters provide the theoretical building blocks which were deemed necessary for the experimental work on perception (chapter 2) and articulation (chapter 3). Also hypotheses are formulated in these sections.

Take as an example justifying the repeated treatment of theoretical building blocks the description of the Gestural Approach: It is only described in very moderate depth in the introductory part (see section 2.2). It only contains the description of the lossless undamped spring-mass system but not of the damped system which forms an integral part of the actual movement generation device of Task Dynamics. This limitation was strategic in the sense that the only purpose in the introductory part was to highlight the alternative conception and

[1] "Loops" denote the loop-shaped trajectories of velar consonants during oral closure, in particular in front vowel context.

role of time and timing in this particular theoretical apparatus - of Ecological Psychology[2] - in comparison to more conventional phonologies. The treatment of perceptual theories can be seen somehow in analogy: While the introductory chapter just aimed to elaborate the approach aspired - in the form of a "Howto" for the application of a cross-linguistic usage of the CP paradigm for place of articulation studies (section 2.3), the information given there is supplemented by theories more closely related to the derivation of hypotheses like for example the locus theory of Sussman et al. (1998) in section 3.2. The division between articulation and perception pertains to the presentation of results: The results for perceptual experimentation (section 6) are separated from the articulatory results (sections 9, 10 and 11), i.e. presented together with their respective theoretical motivation. The results obtained are viewed in a more aggregate fashion in a separate part (part III) containing the general discussion which also concludes this dissertation.

Summary of results - Perceptual Studies
As mentioned, the research on the perception of place of articulation in oral stop consonants has almost exclusively focused the main places of articulation which are also most common in the sound inventories of the world's languages, and the palatal place of articulation has been disregarded in this respect. One language that phonemically has a palatal stop is Hugarian. The aim was to compare a language with and without such a palatal stop phoneme in the inventory under deprived conditions. As a language without such a phoneme French was chosen because of its better match with Hungarian with respect to voicing implementation as compared to e.g. German. The deprivation mentioned was achieved by generating synthetic CV syllables as stimuli where V is the neutral vowel which deprives the listener of the possibility to make perceptual adjustments to vowel context. The listener was further deprived of burst information which was not synthesized. Results suggest that

(i) territorial mapping of the responses in stimulus space for palatal and alveolar place together in a four-category language like Hungarian resembles that of the alveolar place alone in a three-category language like in the present case French, and

(ii) the velar region of the territorial map for Hungarian is shrunk in comparison to the velar region of French.

Summary of results - Articulatory Studies
The dorsal obstruents of Hungarian and German were studied by means of Electromagnetic Articulography (EMA). The target material contained the velar and palatal stops in initial and medial position for Hungarian. For German, target material consisted in (a) medial palatal, velar and uvular fricative allophones and (b) initial and medial velar stops. Results are presented according to the division in (i) intralinguistic Hungarian, (ii) intralinguistic German findings, and (iii) crosslinguistic comparisons:

ad (i) The Hungarian palatals are true dorsopalatals. The fronting of the Hungarian velar converges to a back palatal articulation. Palatals can exhibit large movement amplitudes during oral closure and also during the whole VCV sequence. This implies that contextual velar fronting is not optimally treated in the sense of an underspecification scenario such as that advocated by authors like Keating. Velar stop production in Hungarian shows some patterns of deviance from the patterns observed for other languages like German or English. These patterns partly run contrary to the claim that velar "loops" arise of biomechanical and/or aerodynamic origin exclusively (e.g Kent und Moll, 1972; Perrier et al., 2003).

[2]Articulatory Phonology is the particular instantiation of Ecological Psychology within Phonetics.

ad (ii) Variation generated by the German fricative allophones is distinguishable from within-phoneme variation for the German stops. The stop variants exhibit less contextual variation. Interpretations in terms of phonological instantiation of phonemic or allophonic contrast or in terms of a particular articulatory control regime are rejected due to theoretical considerations.

ad (iii) Crosslinguistic tongue shape comparisons of articulatory profiles indicate that Hungarian velar stop production is in fact reactive to the presence of a phonemic palatal stop in Hungarian. This was evidenced by the crosslinguistic comparison of static tongue configurations as well as by kinematic analyses.

Zusammenfassung

Gegenstand dieser Arbeit sind crosslinguistische Differenzierungen der Artikulationsstelle: Während sich die - sowohl artikulatorisch wie auch perzeptuell ausgerichtete - experimentalphonetische Forschungsliteratur zum Thema Plosive in den allermeisten Fällen auf die Untersuchung der drei Artikulationsstellen labial, alveolar und velar beschränkt hat, hat sich die vorliegende Arbeit zur Aufgabe gemacht, diese Beschränkung aufzuheben und eine vierte, palatale Artikulationsstelle in der experimentellen Planung mitzuberücksichtigen: Die akustischen und artikulatorischen Daten hierzu wurden zum ungarischen palatalen Plosiv erhoben. Die angesprochene Selbstbeschränkung der Disziplin Phonetik ist zum Teil sicher praktischen Erwägungen geschuldet: Phoneme wie die ungarischen palatalen Plosive [c] und [ɟ] sind in den Sprachen der Welt sehr selten und in den Lautinventaren von Sprachen wie dem Englischen schlicht nicht enthalten. Die grundlegende Idee der vorliegenden Arbeit war es dementsprechend, solche Phoneme in die experimentelle Planung mitzuintegrieren, um Phoneme gewissermassen hinsichtlich "phonetischer Ressourcen" konkurrieren zu lassen, also in einen "Kriegszustand" zu versetzen. Anders ausgedrückt, die Grundidee der vorliegenden Dissertation besteht darin, phonetische Kategorien um artikulatorische oder perzeptive "Räume" konkurrieren zu lassen und sich diese Konkurrenzsituation zur Ableitung crosslinguistisch überprüfbarer Hypothesen zunutze zu machen. Ein solcher Versuchsaufbau ist zum Beispiel hilfreich, um zu überprüfen, ob Einflüsse des Phoneminventars einer Sprache auf Muster velarer Koartikulation nachweisbar sind. So existieren im Hinblick auf velare "loops"[3] Forschungsarbeiten, welche systematisch aerodynamische, physiologische, bewegungsökonomische oder biomechanische Einflüsse untersucht haben, wohingegen der Einfluss des Systems linguistischer Kontraste in bisher keiner Arbeit thematisisert wurde. Auch im Bereich der Sprachwahrnehmung ergeben sich Fragestellungen, welche bisher in der Forschungsliteratur nicht oder nur unzureichend beantwortet wurden. Sind beliebig viele Plosivkategorien durch Formanttransitionsparameter allein kodierbar? Oder gibt es hier eine Obergrenze? In der Forschungsliteratur im Bereich der Vokalperzeption sind Arbeiten zu solchen Fragestellungen durchaus gängig, für Konsonanten jedoch sind mir keine empirischen Arbeiten bekannt.

Entsprechend wurden sowohl artikulatorische als auch perzeptive Experimente durchgeführt, was sich allerdings in gewisser Weise als kostspielig herausstellte, da damit auch theoretisch nur zum Teil kompatible Ausgangspunkte notwendig wurden, was im Gegenzug den Gesamtaufbau der Arbeit aufwändiger gestaltete: Ein allgemein theoretischer Teil wurde notwendig (Teil I), in dem bereits sowohl perzeptive als auch artikulatorische Aspekte thematisiert werden. Dieser Hauptstrang wird später in getrennte experimentelle Kapitel (Kapitel 2 und 3) aufgespalten, welche gemäß Gliederung jedoch bereits zum experimentellen Teil II gehören. Diese getrennten theoretischen Bausteine wurden für die eigentliche experimentelle Arbeit zur Perzeption (Kapitel 2) und Artikulation (Kapitel 3) notwendig erachtet. Auch die eigentliche Ableitung experimenteller Hypothesen / Erwartungen befindet sich in diesen Kapiteln.

Die Motivation für dieses Vorgehen sei am Beispiel verdeutlicht: Der gestische Ansatz wird im Einleitungsteil nur wenig detailliert beschrieben (siehe Unterabschnitt 2.2). Hier wird lediglich das - theoretisch unbedeutsame - verlustfreie ungedämpfte Feder-Masse-System beschrieben, jedoch nicht das gedämpfte Gegenstück, welches im Rahmen der eigentlichen Task Dynamics Bewegungstrajektorien generiert. Diese Beschränkung ist jedoch

[3] Unter "loops" versteht man schleifenförmige Bewegungstrajektorien velarer Plosive während der oralen Verschlussphase, insbesondere nach Vorderzungenvokalen.

strategisch motivierbar, da im Einleitungsteil lediglich die alternative Konzeption und Rolle zeitlicher Organisation ("Timing") der Ökologischen Psychologie im Vergleich zu "herkömmlichen Phonologien" verdeutlicht werden sollte. Ähnliches gilt für die Darstellung perzeptiver Inhalte: Im Einleitungsteil sollte lediglich der generelle Ansatz herausgearbeitet werden, - als "Howto" für die Anwendung eines Ansatzes im Rahmen der Kategorialen Wahrnehmung auf crosslinguistische Experimente (Abschnitt 2.3). Die dort gelieferte Information wird dann im experimentellen Teil durch Theorien ergänzt, die sich näher am empirischen Gegenstand befinden wie beispielsweise der Locus-Ansatz von Sussman et al. (1998) in Abschnitt 3.2. Die Trennung von Artikulation und Perzeption besteht in der Präsentation experimenteller Befunde fort: Die Ergebnisse der Perzeptionstests (Abschnitt 6) sind von den artikulatorischen Befunden getrennt (Abschnitte 9, 10 und 11), das heißt, Ergebnisse werden gemeinsam mit ihrer theoretischen Herleitung präsentiert. Die experimentellen Befunde werden in einem separaten Teil (Teil III) erneut in aggregierterer Form abschließend diskutiert.

Zusammenfassung - Perzeptionsstudien

Wie bereits erwähnt, hat sich die empirische Forschung zur Artikulationsstellenwahrnehmung beinahe ausschließlich auf die drei in den Sprachen der Welt häufigsten (Haupt-)artikulationsstellen von Plosiven beschränkt. Die palatale Artikulationsstelle wurde hingegen beinahe vollständig vernachlässigt. Eine Sprache, die solch ein palatales Plosivphonem im Inventar hat, ist Ungarisch. Ziel der Studie war, eine Sprache mit und eine Sprache ohne ein solches Phonem zu vergleichen. Als Vergleichssprache zum Ungarischen wurde Französisch gewählt, da Französisch hinsichtlich der phonetischen Implementation der Stimmhaftigkeitsunterscheidung besser mit dem Ungarischen übereinstimmt als beispielsweise das Deutsche.

Es wurden synthetische CV - Silben als Stimuli generiert, wobei die vokalische Zielposition immer den Fomantfrequenzen des Neutralvokals entsprachen. Den Perzipienten wurden ferner ebenfalls Verschlusslösungsgeräusche vorenthalten, welche nicht synthetisiert wurde. Die Ergebnisse lassen sich im Wesentlichen wie folgt zusammenfassen:

(i) Fasst man in den Identifikationsaufgaben für die ungarischen Perzipienten die palatalen und alveolaren Kategorien zusammen, besetzen sie im Hinblick auf eine Kartierung der Antworten im Stimulusraum mit Hilfe logistischer Modelle ein ähnliches Gebiet wie die alveolare Kategorie alleine für die französichen Perzipienten, und,

(ii) die velare Region des Ungarischen ist im Vergleich zum Gegenstück verkleinert.

Zusammenfassung - Artikulatorische Studien

Es wurden die dorsalen Obstruenten des Deutschen und Ungarischen mittels Elektromagnetischer Artikulographie (EMA) verglichen. Das Zielmaterial für das Ungarische enthielt velare und palatale Plosive in initialer und medialer Position. Für das Deutsche wurden (a) mediale allophonische palatale, velare und uvulare Frikative und (b) initiale und mediale velare Plosive erhoben. Die Präsentation der Ergebnisse wird untergliedert in (i) intralinguistische Ergebnisse für das Ungarische (ii) intralinguistische Ergebnisse für das Deutsche und (iii) crosslinguistische Befunde:

ad (i) Die ungarischen Palatale sind Dorsopalatale. Die Vorverlagerung von Velaren vor Vorderzungenvokalen konvergiert mit der Artikulation von hinteren Palatalen. Palatale weisen während der Verschlussphase und während der gesamten VCV-Bewegung zum Teil große Bewegungsamplituden auf. Dies impliziert, dass kontextuelle velare Vorverlagerung nicht optimal durch ein phonetisches Unterspezifikationsszenario erfasst werden kann, wie es von Autoren wie Keating vertreten wurde. Die Produktion velarer Plosive des Ungarischen weicht von der für Sprachen wie das Deutsche oder Englische ab. Die beobachteten Muster widersprechen teilweise Behauptungen aus der Forschungsliteratur, nach denen velare "loops" alleine aus biomechanischen und/oder aerodynamischen Mechanismen erklärbar sind (z. B. Kent und Moll, 1972; Perrier et al., 2003).

ad (ii) Die durch die Frikativallophonie des Deutschen generierte Variabilität ist von der kontextuellen Variabilität der dorsalen Plosive des Deutschen unterscheidbar: Die Plosive variieren weniger mit dem Vokalkontext. Eine Assoziation dieser Korrelation zwischen phonologischer Spezifikation und artikulatorischen Variabilitätsmustern mit artikulatorischen Kontrollstrategien wird jedoch aufgrund theoretischer Überlegungen verworfen.

ad (iii) Der crosslinguistische Vergleich von Zungenformen ergab, dass die Produktion velarer Plosive im Ungarischen tatsächlich durch die Gegenwart eines palatalen Plosivs beeinflusst ist. Dies konnte sowohl mittels crosslinguistischer Vergleiche statischer Zungenkonfigurationen als auch mittels kinematischer Analysen gezeigt werden.

Contents

Part I.

Theoretical Part

1. Background

Liljencrants und Lindblom (1972), in their seminal paper, were the first to relate *phonetic* principles of perceptual contrast (or auditory distance [1]) and articulatory economy to the structure of vowel inventories and their sizes: In short, languages prefer vowels which are maximally distinct for the perceiver and to be produced with the least effort for the producer. More explicitly, the theory often referred to as Adaptive Dispersion Theory (ADT) makes following predictions: With respect to vowel dispersion in vowel systems, it expects that the maximal range between peripheral vowels should increase with an increase in the number of vowels in the inventory, or, put differently, the area spanned by the peripheral vowels of the vowel system gets larger with an increasing number of vowels in the vowel system. With respect to the distinctiveness of individual contrasts, it predicts that adjacent vowels ought to be quantitatively roughly equidistant in identical vowel systems, and, vice versa become less distinct as the number of vowels in the vowel inventory becomes larger.
Among the criticisms against Liljencrants und Lindblom (1972) are the following: For example, formants are an acoustic measure and not an auditory one, thus other authors have improved on this using more plausible auditory representations like Bark-transformed values (e.g Diehl; Lindblom und Creeger, 2003). This and the usage of background noise in order to augment the ecological validity improved the problematic empirical predictions of too many high non-peripheral vowels and the impossibility to predict [i,y,u] within the set of high vowels of the original version. Furthermore, acoustic vowel spaces cannot arbitrarily increase, there is an upper limit (Maximal vowel space, see Boë; Perrier; Guérin und Schwartz, 1989) given by the capabilities of the vocal tract. In short, the approach has undergone many modifications and reconceptualizations in phonetics and phonology (e.g Flemming, 2002; Schwartz; Boë; Vallée und Abry, 1997) with respect to the endeavour of a numerical simulation of vowel inventories. Given that each language has its own categorical inventory of phonemic units - in this case vowels - it is also possible to *empirically* compare languages with different numbers of phonemes, and the vowel space dispersion hypothesis has undergone more or less explicit testing several times, although with varying success:
Jongman; Fourakis und Sereno (1989) report a trend for larger vowel inventories such as German (14 vowels) and American English (11) to display more peripheral vowels than smaller vowel inventories such as Greek (five vowels) for the vowels /i,a,u/. Other studies could not confirm the predictions made by Adaptive Dispersion Theory though. In Bradlow's 1995 study, the area of the F1/F2 space covered by /i,e,a,o/ was roughly the same in English, Spanish and Greek although the English system has 11 vowel contrasts while the Spanish and Greek systems have only 5 vowels. Livijn (2000) compared 28 vowel systems of different sizes, the results

[1] Auditory distance or perceptual contrast D was conceptualized in terms of the Euclidean distance between any two vowel sounds *i* and *j* in a space defined by the frequencies of the first two formants - in the simplest case: $D_{ij} = ((\Delta F_1)^2 + (\Delta F_2')^2)^{1/2}$.

indicating that the distance between vowels does not increase unless the language under analysis has more than about 8 vowels. It is neither possible nor necessary to give a comprehensive account of all studies dealing with more or less explicit testing of predictions made by the ADT. Rather, it seems more productive to ask why this study had such a large impact on theorizing in phonetics and phonology. The answer can be found in the original paper and deserves more detailed quoting:

> "It is proposed that predictions of phonological facts be derived as consequences of the structure of the mechanisms available for human speech communication and the optimization of their use. Such an extension would constitute a theory that would be different from "Saussurean" linguistics in several respects; e.g., it would be quantitative, and deliberately substance-based." (Liljencrants und Lindblom, 1972, p. 839)

This is obviously an attack against central tenets of traditional generative phonology which postulates that the phonetic realization of phonological representation is universal and non-grammatical, and the paper by Liljencrants und Lindblom (1972) was an early signal of the softening of borders between phonology and phonetics nowadays observed. The topic stimulated to a high degree by Liljencrants und Lindblom (1972) is often referred to as the Phonetics-Phonology-Interface today. A broad classification of linguistic models differing with respect to the division of labour between phonology and phonetics is described in Beckman (1999): The first one of these approaches proposes a strict separation between these disciplines, and the phonological representations are objects encoding "relational information" between the paradigmatic contrasts of a given natural language. In contrast, phonetic events are "quantitative", non-cognitive models of physical events considered as real-world referents of phonological objects (Beckman, 1999). A second approach, reflecting the aforementioned and more recent convergence between phonetics and phonology assumes discrete categories, and speech production is expressed as rewrite-statements by means of derivational rules (e.g. Keating, 1990a). Other approaches emphasize the physical and conceptual constraints which form the common features of phonological organization and contrast, and theories of phonological representation must be directly evaluated as models of human speech production and perception (e.g Browman und Goldstein, 1995). Despite these differences in the theoretical layout, there seems to be agreement on the mechanisms in shaping the paradigmatic contrasts of a language, articulatory economy (or ease), perceptual discriminability and the maintenance of contrast have been agreed on as the key phonetic factors shaping the sound inventories of the world's spoken languages. An example is the following quote: "In recent years, this approach has been adopted directly into generative phonology by Optimality Theorists such as Steriade and Flemming. The factors of acoustic contrast and ease of articulation are treated as constraints in the grammar, and their interaction captures some of the ways in which articulation and acoustics shape phonology." (Pierrehumbert, 2001)

The preceding paragraph emphasized the tremendous influence of simulation approaches to vowel systems on the nowadays agreed key factors shaping the Phonetics-Phonology Interface - articulatory economy, perceptual discriminability and the maintenance of contrast. Then one question seems to be inevitable: Why are there hardly any comparable approaches to consonantism?[2] According to Abry (2003), Ohala, at ICPhS 79 put forward the false prediction of a dispersion theoretic account for a putative 7-consonants systems consisting of the seven consonants

(1.1) ɗ k' ts ɬ m r |

Although embedded into a thought experiment for the purpose of displaying the limitation of dispersion-theoretic reasoning with regard to its *direct* applicability to consonantism, Abry (2003) points out that "in search of consonant-place-space (CPS), we have to leave aside manner, as it is done in predicting oral but not nasal vowel space together." (Abry, 2003, p.727). In the following he develops a triangular representation while actually leaving aside manner which resembles the corner vowel space containing the "corner consonants" [b],[d] and [g]. Thereby, his reasoning is rooted in considerations of the "Dispersion-Focalization-Theory" (DFT) by Schwartz; Boë; Vallée und Abry (1997), which is in the first place aiming at similar explananda as Liljencrants und Lindblom (1972), but not limited to vowel systems as their precursor. DFT theory is based on two auditory-perceptual principles, (a) the dispersion principle in the sense of Liljencrants und Lindblom (1972), i.e. articulatory gestures should provide acoustically sufficiently heterogeneous output, and (b), the focalisation principle: gestures should provide salient spectral patterns, easy to process by the listener. As an example, for vowels, /u/ and /i/ are focal vowels due to the convergence between first and second and second and third formant respectively. Without going into greater detail, this approach is, as mentioned, intended to hold for vowel and consonant spaces likewise. This works by drawing on acquititional reasoning mainly inherited from McNeilage (1998), whose frame-content theory establishes a syllable-based *frame* attributable to the jaw-cycle, whose main acoustic correlate is $F1$ change, or, put more linguistically, vowel height. Then, place of articulation is established by the articulators lips and tongue as the syllable *content*. Its correlates are second and third formant frequencies, and Abry (2003) derives a universal consonant triangle consisting of labial, coronal and dorsal onsets as "corner consonants" as acoustically optimally dispersed/distant members in the F2-F3 space. Nevertheless, apart from Abry (2003), there are at best few other approaches attempting to establish a comparable *substance*-based approach for consonantism, and Abry (2003) himself identifies the major tendencies in phonological feature theory as the driving force: "Jakobson; Fant und Halle (1952) conceived acoustically these places in parallel with the vowels, as a triangular binary representation, until Chomsky and Halle switched to articulatory features." (p. 727)

This alludes to features as layed out in "Preliminaries to speech analysis" (PSA) by Jakobson; Fant und Halle, which are unique (a) in their dual articulatory and acoustic correlation and (b) in the scope of features being valid for vowels and consonants likewise. Of particular interest for the development of parallel universal vowel and consonant triangles are the oppositions *acute/grave* and *compact/diffuse*. The acute/grave opposition is subsumed together with the flat/plain opposition among the tonality features.

> "Acoustically, this feature means the predominance of one of the non-central regions of significant part of the spectrum. When the lower end of the spectrum predomi-

[2] A very obvious answer to this questions relies on speech perception: consonant perception is strongly categorical, vowel perception to a much lesser degree.

> nates, the phoneme is labeled grave; when the upper end predominates, we term the phoneme acute." (Jakobson et al., 1952, p. 29)

Acute/Grave distinguishes front from back vowels and "peripheral" from "central" consonants. Grave sounds include back vowels and labial and velar consonants. Acute sounds include front vowels and dental, alveolar and palatal consonants. Productionally,

> "the gravity of a consonant is generated by a larger and less comparted mouth cavity, while acuteness originates in a smaller and more divided cavity." (Jakobson et al., 1952, p. 30)

Of minor relevance here is the flat/plain opposition, which contrasts rounded and unrounded vowels. More important in this context is the opposition *compact* vs. *diffuse*:

> "Compact phonemes are characterized by the predominance of one centrally located formant region (or formant). They are opposed to diffuse phonemes in which one or more non-central formants or formant-regions predominate." (Jakobson et al., 1952, p. 27)

In production, compact consonants include velar and palatal consonants (as well as open vowels), diffuse consonants are articulated "in the front part of the mouth", i.e. dentals and alveolars. Starting with these featural representations, it becomes possible to construct the triangular and quadratic consonant and vowel spaces. These are motivated as follows for consonant systems (see figures 1.1 and 1.2):

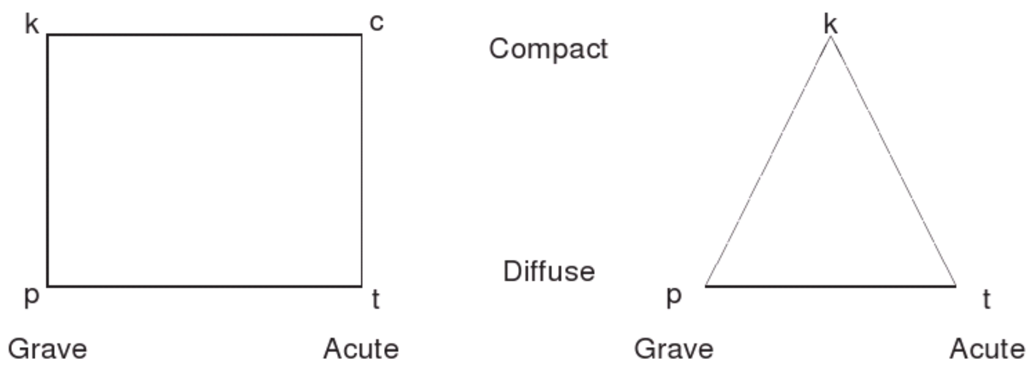


Figure 1.1.: Schema for quadratic and triangular consonant spaces (after Jakobson et al., 1952, p. 33). Czech (left) and French (right) are provided as prototypical languages in the original text.

> "Consonants almost universally possess a tonality feature. As a rule, the diffuse consonants exhibit the opposition grave vs. acute, which often is found also in the compact consonants. In other words, the consonant patterns usually include

> both labial and dental phonemes and frequently also mutually opposed velars and palatals. Such is, for instance, the case in several Central European languages - Czech, Slovak, Serbocroatian and Hungarian. Their consonant phonemes form a square pattern, while in languages such as English and French, which do not split their compact consonant into grave and, ceteris paribus, acute phonemes, this pattern is triangular." (Jakobson; Fant und Halle, 1952, p.32)

In the same vein, Jakobson; Fant und Halle (1952) derive representations for vowel spaces. Additionally to the corner vowels, the vowel /ae(ɛ)/ combines the features compact and acute (see figure 1.2).

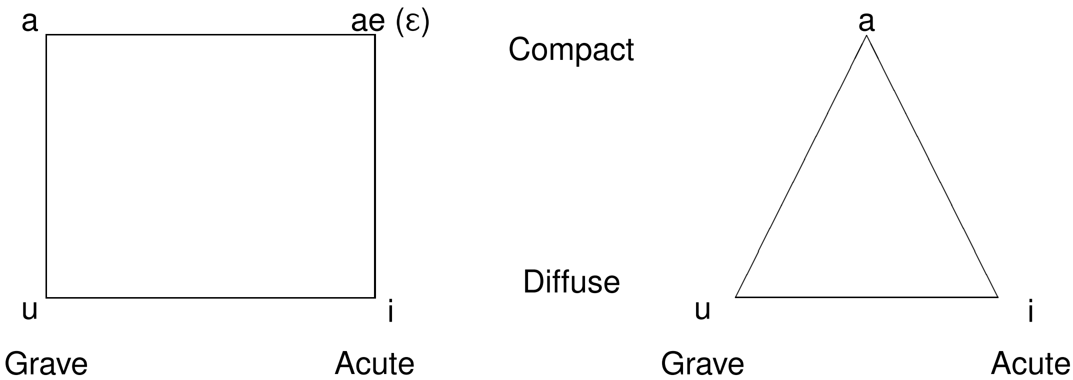


Figure 1.2.: Schema for quadratic and triangular vowel spaces (after Jakobson et al., 1952, p. 34). The original text lists Wichita (left) and Arabic (right) as prototypical examples.

In summary, the partial description of the feature system of Jakobson et al. (1952) served the purpose of illustrating that there had been a conception of a consonant-space isomorph in structure to the vowel space as present in contemporary phonology / phonetics, which was based on a conception of features with a compound definition derived from speech production, acoustics and perception. And here it becomes possible to allude to the aims of the present work the first time: The crosslinguistic, substance-based comparison of the consonantisms of languages of the types as shown in figure 1.1, i.e. languages with three resp. four places of articulation.

1. A coronality mismatch in phonological representations

Maybe the features proposed in Jakobson; Fant und Halle (1952) were too general and poorly quantified (cf. Vallée et al., 2002), however: Generative Phonology, as layed out in "The Sound Pattern of English" (SPE) of Chomsky und Halle (1968) kept from PSA only the idea of a universal system of binary features. The acoustico-perceptual specification was replaced by a universal phonetic representation, but expressed in terms of articulatory features, Trubetzkoy's markedness theory was reintroduced. Maybe the most substantial impact of SPE on the discipline was

the two following decades' sole reliance on the power of segment's relative position within the system to define phonological entities.
On a featural level, SPE replaced the [grave/acute]-feature with the feature [coronal]. Coronal sounds are sounds "produced with the blade of the tongue raised from its neutral position" and a crossclassification with the main places of articulation looks like this:[3]

	lab	dental	alv	retr	pal.alv	alv.pal.	pal.	velar
[cor]	-	+	+	+	+	+	-	-

Table 1.1.: Cross-classification of the feature [coronal] with the main places of articulation.

In order to economically compare [coronal] with PSA, the crossclassification is repeated for the feature [acute] instead of [grave].

	lab	dental	alv	retr	pal.alv	alv.pal.	pal.	velar
[acute]	-	+	+	+	+	+	+	-

Table 1.2.: Cross-classification of the feature [acute] with the main places of articulation.

The differentiation is identical for the labials, velars and all coronal places with the exception of the most posterior palatal place. And indeed, in the phonological literature, a revision of the palatal place specifiying it as a coronal has been proposed by Lahiri und Evers (1991).[1]

Other authors have promoted the position that the phonological patterning of alveolars and palatals has an *articulatory* motivation and constitutes a natural class and therefore does not constitute evidence for the *acoustic* feature [-grave] (Pagliuca und Mowrey, 1980, after Hall, 1997). Definitively though, a mismatch between the feature systems of the PSA and the SPE can be stated. This mismatch is aggravated by a differential treatment of continuants and noncontinuants, in particular stops and fricatives, in their coronality specification. Hall classes palatal *fricatives* as noncoronal:

> Abundant evidence exists that palatal fricatives such as [ç,ʝ] pattern phonologically as noncoronal. (Hall, 1997, p. 15)

Part of the evidence given in Hall (1997) comes from German. [+labial][+coronal]$_\sigma$ is a positive wellformedness condition which allows [t,s,ʃ] in final condition after a labial, but not [ç] because it is not coronal.

(1.2) [+labial][+coronal]$_\sigma$

(a)	pt	Abt	'abbot'	mt	Amt	'office'
	ps	Gips	'plaster'	ms	Sims	'ledge'

[3]Following Hall (1997). At this point, the differential treatment of continuant and noncontinuant obstruents remains to be seen.

[1]Hall (1997) cites other references which argue for [-grave] (=[acute]), because it describes a natural class (Vennemann und Ladefoged, 1973; Hyman, 1973; Vago, 1976; Odden, 1978).

(b)	pʃ	hübsch	'pretty'	mʃ	Ramsch	'junk'
	pç	*		mç	*	

For *stops*, the existence of noncoronal palatal stops is universally denied, they are always coronal and universally their articulatorily specification is alveolopalatal, i.e. they are conceptualized as palatalized alveolars. This is exemplified in the description of Hungarian justified by articulatory evidence.

> "The phonological evidence that Hungarian [c,ɟ, ɲ] are [+coronal] correlates with the phonetic evidence that these sounds are alveolopalatal. (Recall that phonetically Hungarian [c,ɟ,ɲ] are similar to alveolopalatals [ɕ,ʑ], in the sense that both sets of sounds are articulated with a lamino-predorsal constriction in the alveolo-palatal zone [...]. Thus, the fact that Hungarian [c,ɟ,ɲ] are [+coronal] is hardly surprising. (Hall, 1997, p. 14)"

To summarize this paragraph, there exists an inconsistency between the the PSA and the SPE feature systems concerning where [-grave] or [+coronal] sounds end traveling backwards from dental/alveolar places of articulation in direction of the velar rendering the specification of "true" palatals under question. The same inconsistency arises in the specification of "true" palatals with respect to the [+coronal]-specification comparing fricatives and plosives. While the fricatives are considered as [-coronal] in Hall (1997), the existence of true palatal stops is denied on articulatory grounds. They are analyzed as alveolopalatals and therefore conceived as [+coronal]. This is already a question which can be tackled by methods from experimental phonetics and its operationalization would go as follows: Compare realization of true palatal fricatives (for example from German) with realizations of sounds which are potential candidates for "true" palatal stops. Candidate languages here comprise Hungarian or Czech. If the stop, measured where appropriate at a temporarily comparable instance in time, appears to be substantially more front in comparison to the fricative, the asymmetric feature specification specification for coronality in the sense of Hall (1997) can be held valid. If this does not appear to be the case, i.e. no greater fronting for the stop in comparison with the fricative is experimentally verifiable, no match of the phonological specification as put forward in Hall (1997) with the phonetic facts is observed. Without anticipating too much of the discussion of the data situation building the empirical basis of the coronality assignments just discussed, here it can already be ascertained that the data basis mainly consists of one shot, one-speaker measurements from static X-rays or linguo- and palatograms. As will be seen later, accounting for this kind of data on a phonetically satisfactory level, i.e. on an inter-speaker and inter-language level, is formally far from trivial. Apart from that, the question concerning the coronality of palatals calls for a representation of phonological units which has been challenged by theoretical phoneticians.

2. Phonetic representations

One of the criticisms Port und Leary (2005, p. 927) put forward was the opinion that "a fundamental mistake of the generative paradigm is its assumption that phonetic segments are formal symbol tokens". This formalism requires a discrete representation of time to get the formal apparatus of generative theory working. This representation of time is the serial order of discrete symbols, i.e. an integer-valued representation, rather than a representation as rational or real numbers generally necessary for the study of biological systems. Therefore, it seems wise to move in the direction of the question raised at the end of the preceding section, but starting from a different point, namely from phonetic detail.

If one views the mature phonological system as the equilibrium state of a perception–production loop, this involves in a first step a gross synthesis of evidence from the phonetic main disciplines articulatory, acoustic and perceptual phonetics.

2.1. Acoustics

Acoustic invariance: burst spectra

The theory of acoustic invariance (Blumstein und Stevens, 1979; Stevens und Blumstein, 1978) can be seen as an operationalization of the PSA features. It makes the two major claims that (a) there is acoustic invariance in the speech signal corresponding to the phonetic features of a language. (Blumstein, 1986, p. 178) and (b) that the perceptual system is sensitive to these invariant properties. One of the targets of the theory of acoustic invariance was linguistic as to provide means for some natural rules in phonology and why certain assimilations are more likely to occur than others. One of the major domains of application has been the issue of palatal consonants (Blumstein, 1986; Keating und Lahiri, 1993). The analysis put forward in this series of papers achieves a formally simpler reanalysis of Chomsky's analysis of Slavic palatalization analysis by reviving the Jakobsonian featural descriptions as already mentioned in figure 1.1, i.e. the features [compact/diffuse] and [acute/grave].

The feature [compact/diffuse] is there to distinguish between front and back (post-alveolar) for consonants. The name of the feature is associated with its acoustic characterization: [Diffuse] sounds have energy spread widely (diffusely) across the spectrum, while in the case of compact sounds the energy is concentrated in the central area of the auditory spectrum and is therefore termed compact.

The feature [acute]/[grave] is what Chomsky und Halle (1968) call [±anterior]. It distinguishes peripheral from central consonants. Again, the definition is acoustical: [grave] sounds are characterized by low frequency and include back vowels and labial and velar consonants. Acute sounds will display higher frequencies and include dental, alveolar and palatal consonants.

The novelty of the "Acoustic-Invariance"-approach concerning the palatalization example was to derive invariant cues from the acoustic signal and resulted in the definition of "spectral templates" correlating with the above-mentioned phonetic features: Alveolar consonants are characterized by the "diffuse-rising" spectral template: "Diffuseness" in this case means that there are higher amplitudes in higher frequencies. The opposite holds for bilabial stops which are characterized by the diffuse falling template, i.e. lower frequencies have lower amplitudes. The condition of acceptance for the diffuse-falling spectral template is that there has to be (a) a peak below 2400 Hz and (b) a second peak between 2400 Hz and 3600 Hz. There are no restrictions for peaks below 1200 Hz. Velar consonants are characterized by their conformity to the 'compact' template. It is assumed that peaks which are closer together than 500Hz are perceptually integrated into a single peak. Velar consonants exhibit such a behaviour and show only one but prominent spectral peak. The definition of prominence is given as follows: “A peak is ’prominent’, if there are no other peaks nearby and if it is larger than adjacent peaks, so that it stands out, as it were, from the remainder of the spectrum. In this sense, the spectrum is compact.” (Blumstein und Stevens, 1979, p. 1004ff)

In order to view the above mentioned palatalization process as a true assimilation, Blumstein (1986) reports data comparing true palatal stops and velars in different vowel contexts. This involves the acoustic description of palatals before high front vowels in terms of burst spectra: The velar before [i,e,a] exhibits a broad mid-frequency peak in about the region between 2 and 3kHz, whereas before [o,u] there are two distinct peaks, one at about 900 and the other about 4200Hz. The palatal before [i,e,a] shows a high-frequency peak at about 3500Hz, and before [o,u] a mid frequency peak at in the region between 2 and 2.5 kHz. The conditions for interpreting these data as an assimilation are that the two contiguous segments must have similar acoustic properties and the original sound and the modified sound must share a number of acoustical properties. This is the case for [k] before front vowels and [c] before back vowels both exhibiting a broad mid-frequency peak. These data are shown in tabular form in table 1.3.

	velar	**palatal**
[i,e,a]	broad mid-frequency peak ≈ 2-3 kHz	high-frequency peak ≈ 3.5 kHz
[o,u]	peaks at 900 & 4200 Hz, hole in between	mid-frequency peak ≈ 2-2.5 kHz

Table 1.3.: Spectral properties of velar and palatal stops as a function of vowel context (after Blumstein, 1986, p. 183).

Given these acoustic properties, the fronting of the velar can be stated as an assimilation of the gravity feature of the vowel [i] to that of the preceding consonant which changes the velar into a palatal stop (Blumstein, 1986, p. 181):

[k]	→	[c]	/ _	[i]
[+compact]		[+compact]		[-compact]
[+grave]		[-grave]		[-grave]

Lahiri; Gewirth und Blumstein (1984) undertook an acoustic study testing the promise of the Acoustic Invariance approach with respect to crosslinguistic validity. They applied the template-fitting techniques as described in Blumstein und Stevens (1979) to the diffuse, i.e. labial and dental/alveolar consonants of Malayalam and French. The results were dissatisfying, in particular, it was frequently the case that the diffuse-rising pattern for bilabials frequently resembled the diffuse-falling spectral template for the alveolar. Therefore, Lahiri et al. (1984) altered their analysis strategy, now focusing on the change in the distribution of spectral energy between burst onset to voicing onset as the critical invariant. This move is borrowed from the (spectral tilt) metric developed in Kewley-Port et al. (1983) which also exploits spectral changes over time. The form of invariance these authors postulate is called “dynamic relative invariance” (Blumstein, 1986, p. 182) : “This form of invariance is dynamic in the sense that the invariant properties are determined by comparing the spectral properties of portions of the signal across the time domain [...]. It is relative in the sense that the invariant properties are derived on the basis of relative spectral changes in regions of high information". Taken together, spectral characteristics of burst portions are identified as the invariants sought for in the acoustic signal to map onto the phonetic features, and, from the analysis of the phonological process, it becomes evident that transitions are not considered as likely candidates for acoustic invariants: The invariance approach takes a lot of its attraction from the fact that it avoids a classical problem associated with transitional cues for velar consonants, the existence of distinct loci for velars in the context of front versus back vowels.[2] In the Acoustic Invariance analysis of the phonological process mentioned, transitional cues are seen as context-dependent variations, and many of these variations will have minimal perceptual consequences:[3] "Speech perception is seen as less context-dependent than has been generally assumed" (Blumstein, 1986, p. 179). This fits well in the generative paradigm mentioned above, which makes it desirable if language is conveivable as a discrete sequence of feature vectors. As seen in the papers of Kewley-Port; Pisoni und Studdert-Kennedy (1983) and especially Lahiri; Gewirth und Blumstein (1984), this position gradually got weakened by experimental data, and therefore, it seems not to be too surprising that no approach was developed testing how close the fit of *transitional* parameters with the Jakobsonian features is. This might as well have to do with the rivalry between the leading heads in acoustic and articulatory phonetics at MIT and Haskins Laboratories, but clearly represents a gap.

Transitions

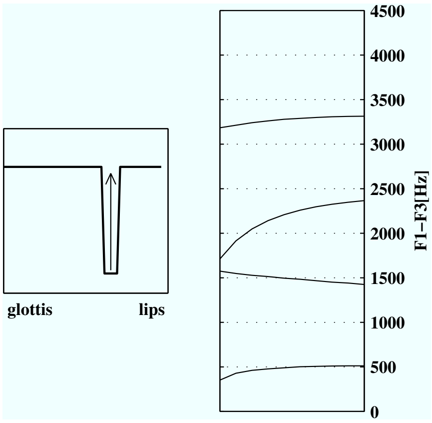
The aim here is to get a gross picture of the behavior of formant transitions at the borderline between velar and palatal articulations, or the borderline between dorsal and coronal articulations if a blunt featural description of the problem is preferred. It seems useful to start off with area functions as discussed in *Acoustic Phonetics* (Stevens, 1998). The simulations will be carried out here to gain insight into the behaviour of the second and third formants when shifting the place of articulation of a velar towards the lips. An additional parameter that has to be dealt with is the length of the constriction,[4] as this is a potential correlate of the "true" palatal, as suggested in some papers. The simulations performed here do not exactly conform to the computational procedures as described in Stevens (1998) which are mostly based on three-tube models with a Helmholtz-resonator in order to model constrictions. Rather the area function to formant conversion is adopted from Sondhi und Schroeter (1987).[5]
Stevens (1998, p.366) estimates formant movements in the vicinity of a velar stop consonant released into a neutral, schwa-like configuration. He centers the location of the constriction for

[2]This topic will be covered in more detail later.

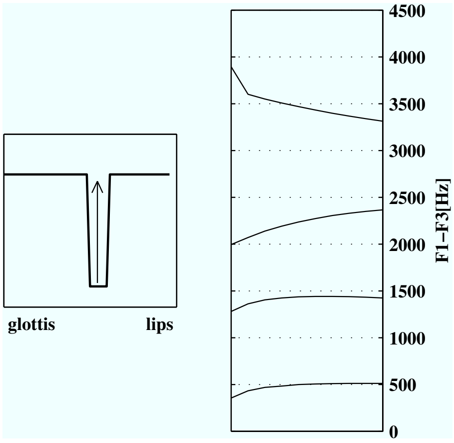
[3]Note that an evaluation of human perception performance, with the exception of experiment 3 in Lahiri et al. (1984) has not been evaluated within this approach. Rather generally, "spectral templates" were developed which served the purpose of the construction of a classificator for measured acoustical data. This is one of the reasons why the distinction between speech acoustics and speech perception is partly abandoned in this introductory section on acoustic phonetics.

[4]Still, the simulations are gross simplifications: No attempt is made to provide realistic area functions of the regions close to the lips and at the larynx.

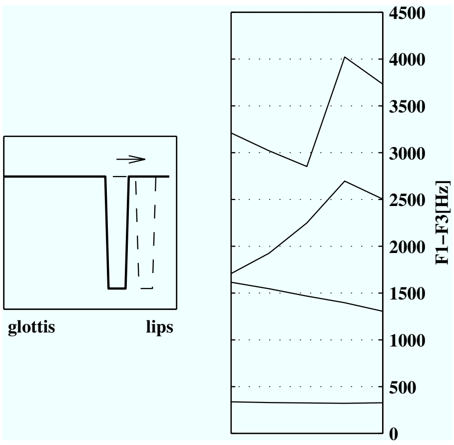
[5]With minor departures from the original work, e.g. the calculation of radiation impedance is not derived from Flanagan (1972) as in the original paper, but rather the approach from Wakita und Fant (1978) is adopted.



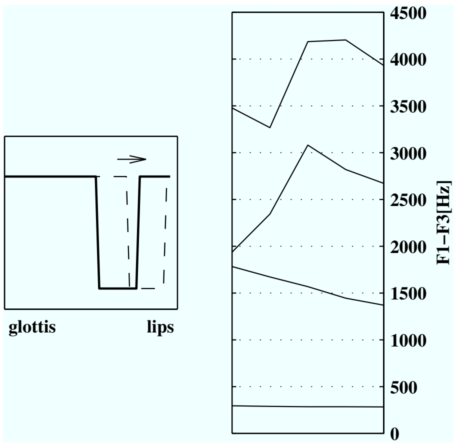
(a) Velar-Schwa: Stevens data



(b) Velar-Schwa: back constriction



(c) Fronting velar constriction



(d) Fronting *long* velar constriction

Figure 1.3.: Acoustic simulations: Simulation for velar-schwa sequences are shown with consonantal constrictions similar to the context of front vowels (a) and back vowels (b). (c) shows the gradual fronting of a velar constriction using a short constriction (3cm) and no vocalic target, (d) shows the same simulation with a long constriction (4cm).

the area function at two thirds of the vocal tract and specifies the constriction length as 2cm.[6] In a first step, a replication of this CV simulation was performed. The results for this replication are shown in figure 1.3a. It has been frequently reported in the literature on velars that as the constriction is moved from posterior (as before a back vowel) to more anterior locations (as before a front vowel), F2 goes from low to high values reaching a plateau near F3. This sometimes is referred to as the velar pinch. This pattern is observed in figure 1.3a which qualifies Stevens' and the simulation in figure 1.3a as belonging to a CV sequence in which the consonantal closure is rather associated to a front vowel. In order to also simulate the onset frequencies in a back vowel context, a second CV sequence was synthesized in which the center of the constriction at the onset of formant transition $2cm$ farther back than in 1.3a, everything else equal. The results of this simulation are shown in figure 1.3b. These two simulations do not appear to be spectacular in the first place at all, but if held against the light of an arbitrary selection of data from the literature - published mainly in perception experiments and synthesis instructions - there appear to be some inconsistencies between the literature and the simulations achieved by Stevens and in this section. The exact values are compiled in table 1.4. Kewley-Port (1982) reports values for the first three formants at transition onset and at the steady state of the following vowel measured from natural /CV/syllables. The values for the second formant onsets vary between 1275 Hz in the /u/-context and 2322 Hz in the context of the front vowel /i/. The low vowel /a/ takes an intermediate value (1733 Hz). The general picture for the transition to the following vowel is that there is very little formant change if the following vowel is /u/ and there is more movement if the following vowel is /i/. The most drastic change is observed if the following vowel is /a/. In articulatory terms, an interpretation is straight-forward: The tongue configuration for /u/ is more similar to the tongue configuration to be made for a velar occlusion. Staying with the same authors, it is possible to interpolate values for the formant values from the KLTEXC synthesis manual which by many researchers in the field is referred to as the "Klatt Cookbook" (Kewley-Port, 1978). Stevens und Blumstein (1978) is one of the few sources that give information about F2/F3 onset values for the rounded back vowel. In this perceptual study, two different continua are synthesized, one as a full /b-d-g/-continuum and one without the central /d/, These continua differ in their exact numerical onset values, so the values of the velar corner stimuli for both continua are reported in table 1.4. Nguyen; Gibbon und Hardcastle (1996) note that they heavily draw on the synthesis specifications reported in the study already mentioned (Stevens und Blumstein, 1978). Nevertheless, the exact numerical values differ from their predecessor. Their perception experiment is refined to the synthesis of initial stops in /i/ and /a/-contexts. Lindblom und Sussman (2004) also report data for second formant onsets in front and back vowel contexts. They can be interpreted from their figure 4. The F2-onset roughly varies between 800 to 1200 Hz for back vowels and between 1800 and 2000 Hz for front vowels. Other material limited to the second formant is obtained from a study by Modarresi; Sussman; Lindblom und Burlingame (2005), although the onsets are not presented separately for different vowel contexts. The striking observation now is the following: Most of the data shown in table 1.4 show much higher values for the F2 onsets in the front vowel context than would have been expected from the simulations as obtained from Stevens (1998) or from the present work. Interestingly, Stevens

[6] Furthermore, the increase in area is normed to $25cm^2/s$ in Stevens' simulations, which is not adopted here, because the shape of the formant trajectory is not relevant here.

himself is coauthoring the paper reporting the highest F2 onsets in front vowel context (2400 and 2500 Hz respectively). Furthermore, this phenomenon is not fully related to the architecture of the studies: Kewley-Port (1982) reports a quite high F2 in measured CV sequences, so the interpretation that the phenomenon could be seen as a synthesis strategy for perception experiments is not justified. Another possibility could be the paradigm shift from FFT-based methods (as in Kewley-Port's paper) to LPC-based methods for formant extraction, although a reason for such drastic changes seems hard to conceive. The fact that the more recent study by Modarresi; Sussman; Lindblom und Burlingame (2005) reports F2 onsets closer to the expectation built from the synthesis reported would be in line with this hypothesis - at the present point it remains mere speculation. In the next step, a basic illustration of *palatal* consonants' transitional patterns

Study Overview:			
CV specification	**F2 onset**	**F3 onset**	**source**
/gu/ spoken	1275	2203	
/gi/ spoken	2322	2835	Kewley-Port (1982)
/ga/ spoken	1733	2367	
/k/ + front vowel:	1815	2565	"Klatt Cookbook"
/k/ + back unrounded vowel:	1340	2375	Kewley-Port (1978)
/k/ + rounded vowel:	1193	2158	
/g/ + front vowel:	1906	2531	
/g/ + back unrounded vowel:	1565	2225	
/g/ + rounded vowel:	1448	1977	
/ga/-synthesis	1610	1750	(Stevens und Blumstein, 1978)
/gi/-synthesis	2400	3000	(full continuum)
/gu-synthesis	1400	2000	
/ga/-synthesis	1580	2000	Stevens und Blumstein (1978)
/gi/-synthesis	2500	3000	(partial continuum)
/gu/-synthesis	1500	2000	
/ga/-synthesis	1640	2100	Nguyen et al. (1996)
/gi/-synthesis	2400	3000	
g-back vowel	800-1200		Lindblom und Sussman (2004)
g-front vowel	1800-2000		
various vowel contexts	≈1770	-	Modarresi et al. (2005)

Table 1.4.: F2/F3 onset values as obtained from the literature (for further explanations see text).

is required. It seems justified to borrow from a non-obstruent, again following illustrations in Stevens (1998): The palatal glide. The high-frequency characteristics of the palatal glide are a high second formant which lies close to the third and fourth formant (Stevens, 1998, p. 526). Although he mentions that a simple acoustic model of the palatal glide can be achieved in terms of perturbation theory - i.e. requiring only a single constriction - Stevens only furnished the more complex articulatory descriptions with exact specifications. In order to derive a simple configuration for a palatal glide, the first logical step consists in an excessive fronting of a velar configuration similar to the velar stop configuration in figure 1.3a. The following simulations will not result in "audible CV transition", are rather undertaken in order to illustrate the effects of the fronting of the velar constriction by means of acoustic nomograms. This can easily be envisioned by looking at the arrows in figure 1.3c and 1.3d. Only closure fronting is modeled, there is no convergence to any open, vowel-like configuration. The simulation shown in figure 1.3c takes the same constriction length which Stevens applies for a glide consonant, i.e. $3cm$ (Stevens, 1998, p. 516).[7] The results for this simulation as shown in figure 1.3c are far from convincing. Although the fronting causes a rising of the third formant, the assumed pattern of the second formant is not obtained. According to Stevens, a relatively long palatal constriction contributes to high values of *both* F2 *and* F3. In order to test this, the simulation was carried out again, but this time using a constriction length of $4cm$, all other parameters equal as in the fronting example just reported. Figure 1.3d shows the results. Although there are minor differences to the preceding simulation, the basic picture is still the same: The situation is slightly better for the area function with the long constriction, but the second formant of about 1750 Hz is still quite low compared to data from the literature: Inspection of figure 9.30 in Stevens (1998) suggests a second formant of about 2000 Hz and a third formant of about 3000Hz for [j]. For palatal stops, the only study found in the literature reporting an F2 onset frequency value determined a value of 2157 Hz for the Komi-Permyak language, the consonant preceding the vowel /a/ (Kochetov und Lobanova, 2003).

This leads to the intermediate summary that the fronting alone is not sufficient, and that the alternation of constriction length is not the remedy either in the construction of an articulatory configuration suitable for generating high second and third formant frequencies simultaneously. Therefore the simulations had to be extended to configurations where a high second formant is warranted: alveolar stops. In order to model the transitions of alveolar closure, Stevens (1998, p. 355) makes use of "a modest narrowing of the tube in the anterior or oral region and a widening in the posterior or pharyngeal region" in his model. Unfortunately, the exact details of this pharyngeal widening are not given in Stevens (1998), but visual inspection of the figures suggests following parametrizations to be reasonable: a widening of the pharynx to about 5.5 cm^2 with a length of the wide section of about 5 cm, the wide section starting at about 2 cm from the glottis. So, the next step consists in extending the velar fronting example to contain a pharynx widening inspired by Stevens' alveolar model, the results of which are displayed in figure 1.4. A further modification to the simulation was necessary though: As mentioned in the description of Stevens' velar model, the constriction was centered at two thirds of the length of the vocal tract. The constriction was centered one cm further back in this simulation. The results are more convincing than before: The second formant reaches a plateau near 2000 Hz and convergence between F2 and F3 is also observed.

Summary

The front constriction alone is not sufficient for obtaining the desired formant configuration, better results were achieved by a supplementary widening of the pharyngeal region. As the last simulation has shown, the result of the high second formant is not so much an effect of the fronting of the velar but more the introduction of a wide pharyngeal region. Of course, with the data just presented, the question whether the rising of the formants is an **automatic** conse-

[7] In contrast to the stop CV simulations with 2cm constriction length.

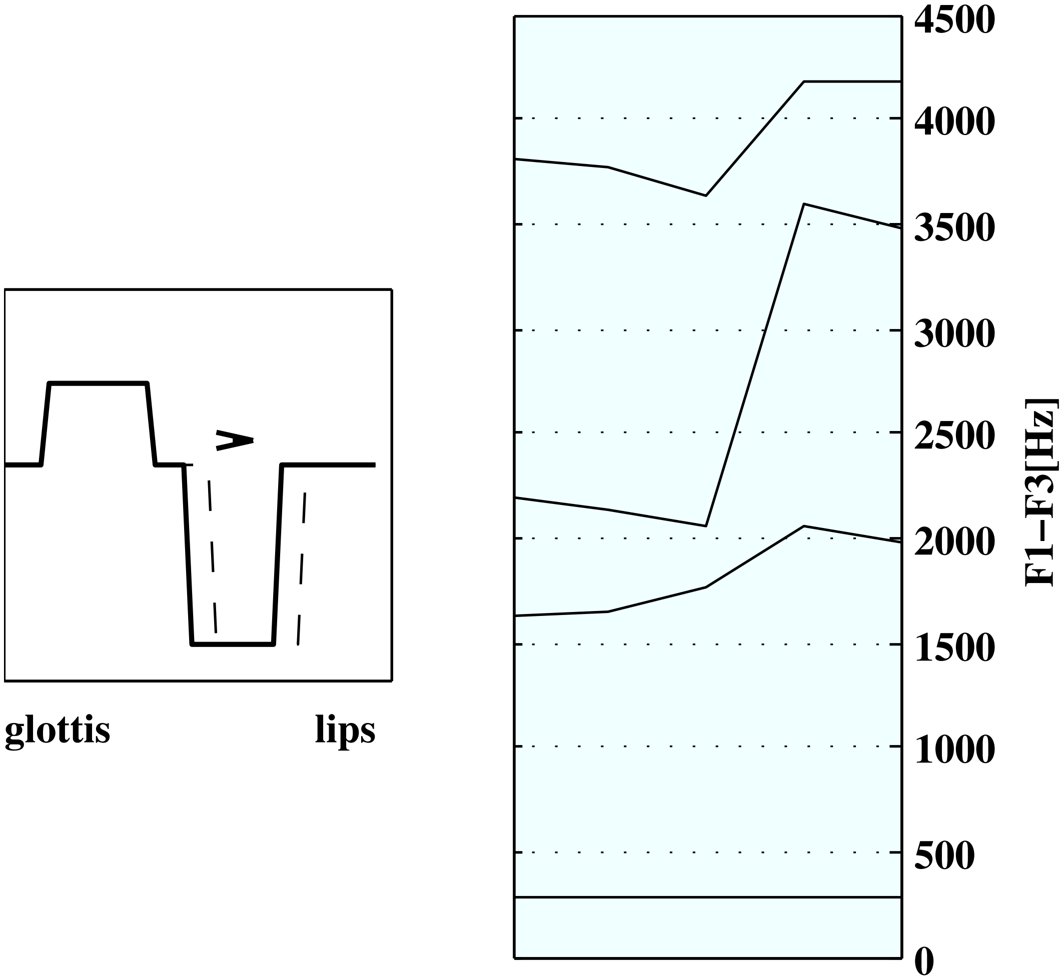


Figure 1.4.: Result of fronting a velar constriction with simulatneous wide pharynx.

quence of fronting cannot be answered, Stevens assumes the following: For both the palatal and the labiovelar glides, “the tongue root is advanced to maximize the volume in the pharyngeal region” (Stevens, 1998, p.516). A second plausible scenario - contrasting with the one advocated by Stevens relying on articulatory control - exists in which the predorsal gesture used for forming the long constriction at the palate is *automatically responsible* for the pharyngeal widening, in turn causing the rising of the second formant close to the third and fourth. By making the constriction longer, the rising of F2 is further enhanced. This implies that the acoustically relevant aspects of the palatal closure do not take place at the actual palatal place of articulation, but are better seen as a side-effect. This goes together well with the standard assumption that F3 is associated with the front (oral) cavity for /i/ and F2 with the back (pharyngeal) cavity (Johnson, 1997, p. 93-97).[8]

The widening refers to the conceptualization of the tongue as a hydrostat: A hydrostat in biological systems is a structure consisting mainly of muscles - and no bony structures - in a separate compartment. The main principle is that water is effectively *incompressible* at physiological pressures. Therefore the tongue will act as a *constant-volume* system.[9] In view of these facts, the creation of the long palatal constriction could be serving the widening of the pharyngeal region as the acoustic simulations of this paragraph suggested the pharyngeal widening as the acoustically more relevant contributor to the rising of the second formant, but surely the back enlargement, not only the constriction fronting is observed for the second formants of palatal segments.

2.2. Production

The question of whether the widening of the pharyngeal constriction is automatic or a planned action - as Stevens seems to propose - can of course not be answered by simulations like the ones presented in this section. A speculative remark is in place though: The book by Archangeli und Pulleyblank (1994) was a corner stone in establishing phonetic motivations of markedness in post SPE Phonology. Their key argument is that the combination of high vowels and an andvanced tongue root is unmarked and provide broad evidence from Edoid languages in order to underpin this claim. Still it has to be conceded that Archangeli und Pulleyblank (1994) are solely concerned with vowels. The question raised whether - or how - the phenomenon of biomechanical coupling on the physiological side and an organizational level of speech movement planning calls for a sketch of the *coarticulation* and related concepts from articulatory phonetics and speech production. The goal of this section is to briefly introduce the necessary concepts. One of the most fundamental questions in articulatory phonetics in general is which of the aspects of speech production are under central nervous control of the speaker and which aspects are just peripheral effects, which can arise from, for example, aerodynamic or biomechanical sources and even their complex interactions. In particular, for velar stops, the debate about the origin of the phenomenon referred to as “loops” has been a paradigmatic test case for such a debate on the origin of overt articulatory behaviour. This debate will be briefly sketched in the next section. Apart from short digressions to tongue physiology and the Phonetics-Phonology-Interface, further sections introduce to the basics of coarticulation theory and Articulatory Phonology.

[8] However, for a critical discussion see Tabain und Perrier (2007).
[9] For further reading with a specific background in speech production see Stone (1995)

Loops

Since the pioneering work of Houde (1968), *articulatory looping patterns* have often been observed in V1-[Velar]-V2-sequences by many researchers. When the velar contact occurs during a foward-directed vowel-to-vowel transition, during the stop the tongue slides in a forward direction towards the second vowel target. These sequences are theoretically unproblematic and can directly be interpreted as a natural consequence of the vowel-to-vowel movement.
However, when the contact occurs during a rearward movement, the direction often is temporarily reversed, i.e. during oral closure the movement goes in forward-direction resulting in small elliptical movement trajectories, or no movement at all is observed. Subsequent to Houde, congruent observations have been made in several studies, for example Kent & Moll (1972), Mooshammer; Hoole und Kühnert (1995), Löfqvist und Gracco (1994), Löfqvist und Gracco (2002), and several competing explanations given for the phenomenon: The looping patterns as a passive forward movement of the tongue due to airstream mechanisms (Kent & Moll, 1972), as a result of an active gesture aiming at the maintenance of voicing (Houde, 1968), synonymously 'cavity enlargement' (Ohala, 1983). Counterevidence against the active planning or cavity enlargement hypothesis consist in the data on German collected by Mooshammer et al. (1995): Articulatory loops during the voiceless stops [k] were larger than during the voiced counterpart [g], but this does not rule out the potential effect of airstream mechanisms completely. Hoole, Munhall & Mooshammer (1998) contrasted normal versus ingressive speech, and ingressive speech resulted in size reduction of the looping pattern: As a consequence aerodynamic influences seem to be at work, but it is not clear when and how they operate. Löfqvist und Gracco (2002) try to explain looping patterns in more general principles of motor control, postulating the entire movement to be planned in terms of cost minimization principles. In a recent modelling study, Perrier, Payan, Zandipour and Perkell (2003) focus on tongue biomechanics moderated by place of articulation. In contrast to Löfqvist und Gracco (2002), Perrier et al. (2003) conclude, looping patterns can be explained in terms of biomechanics alone and the trajectory as a whole does not have to be preplanned. Perrier et al. rather prefer a target based planning reference frame. Furthermore, another finding of this study is the partially high sensitivity of looping patterns to place-of-articulation: Perrier et al. observed that for velars following a front vowel - in contrast to the context of a back vowel - "a small forward shift of the consonant target, associated with very small changes in muscle commands, was enough to reverse the direction and the orientation of the loops, which are now forward directed and counterclockwise oriented." (Perrier et al., 2003, p. 1596). Another reason to expect a sharp distinction on a front-back-dimension lies in a potential peculiarity observed for the dorsal articulator by Alfonso und Baer (1982) consisting in differential initiation times of horizontal and vertical components of movements for the back vowels in contrast to front vowels. Not researched at all in this respect is the potential influence of the structure of sound inventories on the size of articulatory loops: The additional palatal in the paradigm of speech sounds as for example in the Hungarian language could have an effect on the characteristics of the loops in velar stop production. On a gross operational level, one would expect the defining sliding movements during the closure interval to be smaller in a more crowded inventory as Hungarian and to show "normal" variation in a less crowded inventory as German.

It is quite easy to conceive that the basic idea is borrowed from researchers like Sharon Manuel: She examined vowel-to-vowel *coarticulation* comparing languages with more crowded and languages with less crowded vowel inventories, in which neighbouring vowel phonemes are less distinct acoustically (Manuel, 1990). Coarticulatory variation in vowel quality was measured in the genetically related languages Shona (5 vowels), Ndebele (5 vowels), and Sotho (7 vowels). The results showed that the low vowel varied less across vowel contexts in Sotho than in the 5-vowel languages, which is consistent with the hypothesis that contextual variation is subject to a distinctiveness constraint. What remains unclear is whether the main factors underlying loop formation are on the same level of - theoretical - aggregation than the coarticulatory factors which are the major theme in Manuel's investigation: The tiny front movements during the closure of e.g. /ika/ could be seen as finer-grained phonetic details than coarticulatory phenomena and therefore linguistically fully irrelevant - and only a reasonable topic for motor control studies. However, there are researchers which have located looping patterns and coarticulatory phenomena on the same level of resolution. It seems sufficient to take notice of such linguistic conceptions of looping patterns at this place and not to go into further detail. Rather, I will turn to the promised sketch of the coarticulation concept now.

Coarticulation

Broadly speaking, coarticulation refers to the fact that adjacent phonological segments are not realized in the same way in different environments, but rather there is a mutual influence in their realization. Although similar in spirit, coarticulation is not to be confused with *assimilation* in phonology. Kühnert und Hoole (2004, p. 559) note that

> "much of the discussion in the literature concerns the dichotomy between a discrete change of 'segments' implied by a phonological account of assimilation and the phonetic continuum of articulatory reductions that has been observed in experimental studies of assimilated utterances (Barry, 1992; Ellis und Hardcastle, 2002)."

Then, for example, the anticipatory rounding observed during the velar in /ku/ involves no segmental change, and therefore is to be treated as an instance of coarticulation between these two segments, but the realization of /n/ as a bilabial nasal in ‹anpacken› has to be treated as an assimilation. The clearcut separation between coarticulation and other context-dependent phenomena, such as assimilations, is inherited from classical generative theory: Here, coarticulation deals with "transitions between vowels and adjacent consonants, the adjustments in the vocal tract shape made in anticipation of a subsequent motion" (Chomsky und Halle, 1968, p. 295, cited from Farnetani und Recasens, 1999). In contrast, assimilations involve operations on phonological features, and are accounted for by phonological rules, which map lexical representations onto phonetic representations. As the next brick for setting up the concept is the question of the direction of influence of coarticulation of segments according to the segments' temporal order. If one segment influences the realization of the preceding, one speaks of *anticipatory* coarticulation (as in the anticipatory lip rounding example /ku/). If the reverse is true, the term *perseverative* or *carry-over* is used. Alternatively, terms which are presumably easier to use but are biased towards alphabetic writing are *right-to-left* and *left-to-right* coarticulation, respectively. Assimilations again have a similar pair of concepts, the terms right-to-left/left-to-right or anticipatory/perseverative are not normally used here: Concerning assimilations, one speaks of *progressive* and *regressive* assimilation.[10] A last building block for the delineation of the concept concerns the question of whether the *spatial* or *temporal* properties of coarticulated segments are concerned, consequently the literature speaks of temporal and spatial coarticulation.

Another important question is the question about the "richness of the plan".
Kühnert und Nolan (1997, p.62) take the perspective that there must be some invariant, discrete units underlying the variation observed on the behavioural surface:

> "It is essential to the concept of coarticulation that at some level there be invariant, discrete units underlying the variable and continuous activity of speech production. If this were not the case, and, for instance, the mentally stored representation giving rise to a production of the word 'caw' were a fully detailed articulatory plan, then when that word was spoken (in isolation at least) there would be no question of a process of coarticulation - the word would simply correspond to a set of instructions for the time-varying activity of the articulators, and the sub-word segments would not exist in any sense, and could therefore not undergo 'coarticulation'." (Kühnert und Nolan, 1997, p.62)

This is at least compatible with the standard generative approach with its underlying representations and surface manifestations like coarticulation and other properties of phonetic implementation following from universal principles of speech physiology. Here, it becomes interesting to discuss an approach popular in the 1970's where the underlying representation is enriched. Wickelgren (1972, 1969) moves coarticulation away from the periphery to a more central level of speech motor control. In this approach, the speaker does not have a set of phonemic segments represented as the immediate phonetic representation of words, rather an inventory of "context-sensitive allophones". These are allophonic versions of segments specifically designed according to their left and right context in which they are to occur. For example, for a description of the sequence /kik/ one would require the context-sensitive allophones listed in example 1.3.

(1.3) $_{\#}k_i$ - initial /k/ before a front vowel
$_k i_k$ - front vowels surrounded by two voiceless velar stops
$_i k_{\#}$ - final voiceless velar stop after the front vowel.

This approach was discussed in order to show that there have been authors attempting to build a workable solution for the problem of coarticulation based on an enrichment of the underlying representation, and to discuss the problems associated with this enrichment: According to Kühnert und Nolan (1997), it has problems dealing with coarticulatory influences extending beyond adjacent sounds, and with effects caused by prosodic effects, which both would require the definition of additional context-sensitive allophones the number of which, as a consequence, would exponentially grow. In other words, most of the criticism against this approach evolved against its lack of parsimony.[11] Apart from this particular approach, all other approaches to

[10] In regressive assimilation, a feature spreads to the preceding segment. As an example take German nasal place assimilation. The place of the nasal assimilates to the place of the following velar in /baŋk/.

[11] Further critique is that it misses generalizations, for example, all English vowels are nasalized preceding a nasal consonant (Kent und Minifie, 1977).

coarticulation agree in the postulation of a parsimonious underlying representation, and that the integration of segments is rule-goverened and productive, although the exact configuration of the approaches differ considerably, in particular with regard to the nature of the internal representation. The debate about the nature of these representations has, to a large degree, configured the relationship between the disciplines phonetics and phonology or the "Phonetics-Phonology-Interface" and deserves attention in a separate excourse.

Excourse: the Phonetics-Phonology-Interface

It appears hard to excel the formulations on the distinction between three different approaches to the Phonetics-Phonology-Interface with respect to the nature of the disciplines' representations and the division of labour as achieved by Beckman (1999, p.199f), so I can refine myself to a shortened summary of these formulations:

The *first* one of these three proposes "a strict separation between phonology and phonetics. Phonological representations are characterized by discrete symbolizable entitites associated with different nodes in prosodic structures. These entities 'encode relational information' about the language's system of paradigmatic contrasts, and hence are mere 'algebraic objects appropriately formulated in the domain of set theory'. Phonetic representations, on the other hand, involve 'quantitative, non-cognitive' models of 'physical, temporal events'." Phonological and phonetic events are connected by mapping the phonological representations onto the phonetic "real-world referents" (Local, 1992; Pierrehumbert, 1990).

The *second* approach is in quite radical opposition to the first one, and recognizes neither different classes of objects nor a clear division between phonology and phonetics. Such approaches emphasize the physical and conceptual constraints, which form the common features of phonological organization and contrast, and theories of phonological representation must be directly evaluated as models of human speech production and perception. These approaches are also classified as being amenable to a "declarative", "simultaneous" or "non-procedural" view to language in the sense that no "procedural" and sequential phonetics are necessary (e.g. Browman und Goldstein, 1990b, see also later sections on Articulatory Phonology in the current work). In short, the first class of approaches can be seen as simultaneous and nondeclarative because phonetics and phonology deal with objects which are essentially different. The second approach is simultaneous and nondeclarative because the existence of substantially different - phonological or phonetic - objects is denied altogether.

The *third* class of approaches Beckman describes though requires procedural, non-declarative elements, involving several consecutive steps in phonological derivation. As a prototypical representative the paper of Keating (1990a) is mentioned, which takes an intermediate position: Phonological representations are algebraic objects to encode 'relational information' between the paradigmatic contrasts of a given natural language. These neo-generative approaches aim at a convergence of generative phonology and speech production and assume discrete phonological categories, which are interfaced with the speech production apparatus as rewrite-statements expressed in terms of derivational rules. Now, coarticulation acts *on the phonological step in the derivational chain and on the phonetic mapping to articulatory movement.* The latter calls for a kind of mapping to articulatory movement which happens in the context of an *interpolation-*

based model of speech production in Keating's case.[12] Such reasoning is embedded in the more general theoretical framework of *Lexical Phonology*. For the moment it only seems necessary to give a rather general impression of Lexical Phonology in the form of the standard diagram depicting its conceptualization of the phonological grammar, shown in figure 1.5. It simply shows that according to Lexical Phonology, rules can apply at various stages of the grammar.

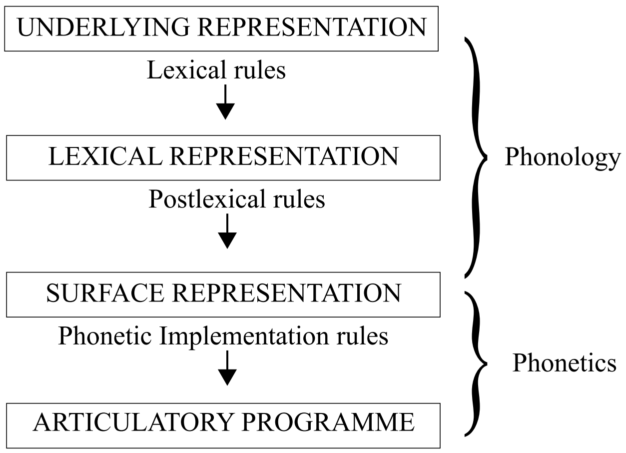


Figure 1.5.: Schematic representation of the common phonological grammar model used in Lexical Phonology, adapted from Gussenhoven und Jacobs (1998, p. 119f.).

Lexical Rules as opposed to postlexical rules are qualified by (i) that they may refer to morphology, (ii) can have exceptions, (iii) are structure preserving, (iv) are accessible to native-speakers' intuition, (v) cannot apply across word-boundaries and (vi) must precede all postlexical rules. This work is not concerned with these levels due to experimental reasons, and because a more striking problem for this kind of approach becomes evident: For example, given the coarticulatory behaviour observed in a corpus study of articulatory e.g. tongue and jaw movement data: How is it possible to decide between phonology and phonetic implementation? This separation appears to be possible under some favourable conditions. Gussenhoven und Jacobs (1998, p. 119f.) give such an example: In English RP, a voiceless plosive is frequently inserted between a nasal and a following fricative resulting in the pronounciation of the word plural of the word "sense" as [sɛnts] rather than [sɛns]. These patterns now can be explained in terms of *details of the phonetic implementation* or be described as the output of a *preglottalization* process to be described in phonological terms. Explaining these patterns in terms of *details of the phonetic implementation* amounts to the following scenario: For the production of the nasal [n], an alve-

[12]The acceptance in the scientific community seems to be larger for models of intonation than for models of supralaryngeal speech production.

olar closure has to be produced and the velum has to be lowered too. The occurence of the [t] depends on interarticulatory timing: To reach an [s], the closure has to be released and the velum raised in order to close the velopharyngeal port. The occurrence of [t] now depends on whether the velum is closed before the alveolar closure release. If this is the case, substantial intraoral pressure is built up behind the alveolar constriction, and the release of the alveolar closure will produce burst cues responsible for the stop impression. Because at the time of the closure of the velopharyngeal port the transglottal pressure difference starts to get neutralized, vocal fold vibrations ceases and the result typically will be voiceless [t]. According to Gussenhoven und Jacobs (1998), this phonetic explanation does not hold though, the pattern rather is an instance of *preglottalization*: Inserted stop segments of the kind reported in this example do trigger phonological preglottalization, just as regular underlying stops do, therefore it must be a *phonological* process. Further, Gussenhoven und Jacobs (1998) mention that for American English (AE) the phonetic implementation hypothesis holds, because AE has no phonological preglottalization. This illustration was primarily intended to point to the problem that distinguishing between the levels within such a highly modular approach is not always trivial.

Further important aspects only mentioned here are (i) that a disentangling of the effects of the phonetic implementation and the effects of the articulatory programme might not be possible in all cases and (ii) the discussion of the notion of *underspecification*. A short illustration of how it functions is given by the observation that the "higher" in figure 1.5, the closer to the underlying form, the more parsimoneous is the representation on its way to be output by the speech apparatus. The underspecification concept is central to the theory of *feature geometry*, which is reviewed in 3, although not touching all of its aspects, rathter concentrating on its role in phonetics as layed out in Keating (1988b). Further, the phonetic instantiation of underspecification targeted in this work requires a more detailed discussion of *interpolation-based models*, which is still to be described in later parts of the current work closer to the derivation of empirical predictions (see part II, section 4 and subsections).

It is interesting here to note that an example closer to the topic of this work is critically discussed in Beckman (1999) referring to Keating (1988b): Here, she uses the V-to-V coarticulation across the voiceless velar fricative /x/ as evidence for a phonological specification of [-back] for a categorically front allophone for /x/ before /i/. There is no phonological contrast between velarized versus palatalized dorsal fricatives in Russian, and presumably no other morphological evidence supporting an underlying specification of any value for [$\pm$back], for this segment, Keating supposes that the feature is inserted by a phonological rule very late in the derivation. The result of this interpretation of F2 coarticulation is an instance of the - controversial - "nontrivial" or "temporary" underspecification. Beckman (1999, p.215) herself doubts though whether Keating's account "casts the burden into the right direction", rather asks whether "we should look for a more complete model of coarticulation to explain the apparition of a cateorical specification". She proposes to take into account two other phonetic phenomena to explain the assimilation of /x/:

- the large quantal region for /i/ in the mapping from constriction location to F2.
- aerodynamic or other (see section 2.2) causes of the forward "loop" trajectories (Mooshammer et al., 1995)

Looping patterns were already mentioned in section 2.2, and it was indicated that there is little hope to "factor" physiological, aerodynamic and biomechanic influences by the movement data which are planned to acquire here. Putting this in the context of what was said in preceding paragraphs, it becomes evident that loops are to be put in the context of the coarticulatory patterns in contextual variation. This has to be done in a way not neglecting or deprecating low-level influences as more or less accidental correlates of the linguistic system, rather these influences are indispensable for arriving at an appraisal of the success of approaches like the one proposed by Keating.

Now, I want to turn back to the ***second*** class of approaches which is opposed to the one just described with due to not recognizing a clear - modular - division between phonology and phonetics. One proponent was already mentioned really at the beginning of this chapter when Port und Leary (2005, p. 927) put forward their reservations against views according to which "phonetic segments are formal symbol tokens". This is not a single person's view, but rather has been qualified as an epistemologically fundamentally different research paradigm, *Ecological Psychology*. Central to Ecological Psychology is a radically different conceptualization of time, as shown in the following quote:

> "the flow of ecological events is distinct from the abstract passage of time [...] The stream of events is heterogeneous and differentiated into parts, where the passage of time is suppoesed to be homogeneous and linear. Issac Newton asserted that "absolute true and mathematical time, of itself and from its own nature, flows equably without relation, to anything external." But this is a convenient myth." (Gibson, 1986, p.100, after Byrd und Saltzman, 2003, p. 150),

Articulatory Phonology I

The question then is: How is an alternative conception of time to be shaped with regard to the particular research domain focused here? Still on a programmatic level, Fowler (1980) argues against speech production theories in general which take phonological features as input. The features used as input for the speech production mechanism are timeless, abstract and static and have to be translated into articulatory movement. As Farnetani und Recasens (1999, p. 51) put it: "In this translation process, the speech plan supplies the spatial target and a central clock specifies when the articulators have to move." Fowler's intention is to overcome this dichotomy and she suggests to modify the phonological units of the plan: The phonological units become dynamically specified phonetic gestures, with an *intrinsic temporal dimension*. The new approach encodes serial order directly in the phonological representation of prosodic organization, in which consonant gestures are phased relative to the cyclic rhythm of stressed and unstressed vowel gestures. Fowler further rejects the phonological treatment of segmentation that this position requires to function. The theoretic approach she envisions here is Gestural Phonology (Browman und Goldstein, 1986, 1990b, 1992), and the concrete instantiation as a model is *Task Dynamics* (Saltzman und Munhall, 1989). In a more recent paper (Goldstein und Fowler, 2003, p. 161), it is layed out that the three key hypotheses of Gestural Phonology are that "vocal tract activity can be analyzed into constriction actions of distinct vocal organs, that actions are organized into *temporally overlapping structures*, and that constriction formation is appropriately

modeled by dynamical systems".
Central to their approach is the definition of the *gesture*. A gesture is defined through the specification of one or more *dynamic equations*, a motion variable (or variables), values for the coefficients of the equation (dynamic parameters) and weightings for individual articulators. Articulators move continuously over time. While in traditional approaches continuous motion is modeled by assuming that there are *targets* for the phonological units, and speech production involves interpolation between these targets (Keating, 1990b), Articulatory Phonology (AP), in contrast, models gestural units as *dynamical systems*. A dynamical system is characterized by an equation (or set of equations) that expresses how the state of a system changes over time. "Crucially, production of a gestural constriction can be modeled as a *mass-spring-model* dynamical system with a *fixed* set of parameter values" (Goldstein und Fowler, 2003, p. 167). The paradigm of dynamical systems as frame of description has been extremely successful for a variety of controlled systems. The formal "ingredients" of such an approach are an (assumed or known) fixed number of degrees of freedom, a phase space, state variables, and a (usually differential) equation of motion governing the temporal evolution of the system, or its movement in phase space along certain trajectories.

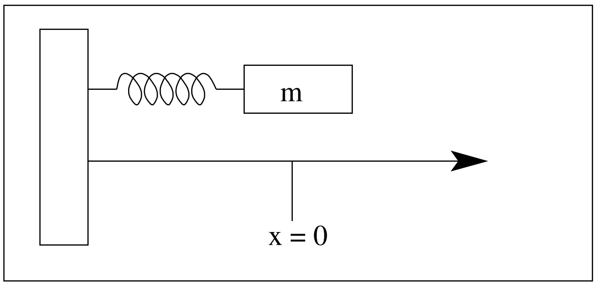


Figure 1.6.: Mass-spring-model, the harmonic oscillator.

As mentioned, the dynamical system inherent in the gestural approach can be described as the familiar problem from elementary physics courses, a mass attached to a spring (see fig. 1.6) - or a *harmonic oscillator*.

If the mass in Fig. 1.6 is pulled, stretching the spring beyond its rest length (equilibrium position), and then released, the system will begin to oscillate. Assuming that the system is without friction, the resulting movement trajectory of the mass can be described by the solution to the equation

$$m\ddot{x} + k(x - x_0) = 0, \tag{1.4}$$

where m is the mass of the object, k is the stiffness of the spring, x_0 is the rest length of the spring, x is the instanteous position of the object and $\ddot{x}$ is the instanteous acceleration of the object. It has the solution

$$x(t) = x_0 \cos \omega t \qquad where \qquad \omega = \sqrt{\frac{k}{m}}. \tag{1.5}$$

If the single parameters are varied, this results in different outcome trajectories, e.g. if the stiffness of the spring is increased, the spring oscillates with a higher frequency, changes in the rest length and the initial position will alter the amplitude of oscillation. This is the equation of motion, and, in this particular case, a second-order, ordinary differential equation. To gain insight about the dynamics of the harmonic oscillator is to consider all possible initial conditions. One way to achieve this is to plot the system in the *phase-plane*. The *state* of the oscillator at time t is the pair of values $(x(t), v(t))$. Given a solution $x(t)$, the velocity v can be found by differentiating yielding

$$v(t) = -\omega x_0 sin\omega t \tag{1.6}$$

Using equations (1.5) and (1.6) and the trigonometric identity $sin^2\omega t + cos^2\omega t = 1$, the phase portrait can be found as

$$x^2 + \frac{v^2(t)}{\omega^2} = x_0^2 \tag{1.7}$$

As easily seen from equation 1.4, gestures within Task Dynamics are autonomous structures which can generate articulatory trajectories in space and time without any additional interpretation or implementation rules (equations 1.5-1.7). But which kind of predictions do Gestural Phonology / Task Dynamics make about the coarticulatory behaviour of velar and palatal segments? At this point, the necessity arises to review the very basic results of one of the most influential studies and the accompanying model of coarticulation: The empirical study by Öhman (1966) and his corresponding conceptual paper (1967) introduced a model of vowel-to-vowel coarticulation, the basic empirical evidence of which was derived from articulatory and acoustic analysis of Swedish VCV utterances produced in isolation, and from similar speech material in American English and Russian. The Russian data was different with regard to (secondary) palatalization. The major finding was that the consonantal transitions (V_1C and CV_2) depend on the identity of the transconsonantal vowel. But: this coarticulatory variability was reduced to almost random fluctuation in the case of Russian. Öhman interprets these findings as follows: The tongue is considered a system of independently operating articulators driven by invariant articulatory command. The apical articulator is involved in the formation of apical consonants, the dorsal articulator in the formation of palatal and velar consonants and the tongue body articulator in the formation of vowels. The reduced coarticulatory variability for the palatalized F2-transitions is seen as the result of conflicting vowel commands on the tongue body, i.e. an [i]-like palatalization commands exerting a blocking effect on the following vowel. And in fact, *coproduction* theory and *Articulatory Phonology* (AP) and Task Dynamics have been elaborated on the basis of Öhman's work.

Essentially, the theory predicts strong variation on place-of-articulation for velar stops as evidenced by the following quote (see also figure 1.7)[13] :

> "In the case where consonants and vowels share the same [tongue body] tract variables [...], the consonant and vowel gestures cannot both simultaneously achieve their target, since they are attempting to move exactly the same structures to different positions. As a result, the location (but not degree) of constriction achieved for

[13]This figure was generated using the software package TADA by H. Nam ("Task Dynamic Application", see Nam; Goldstein; Saltzman und Byrd, 2004).

Figure 1.7.: Illustration of gestural blending (adopted from Saltzman und Munhall, 1989). Velar Consonantal place of constriction (A) varies with following vowel identity (B)

> the consonant will vary as a function of the overlapping vowel." (Browman und Goldstein, 1992, p.165, after Saltzman und Munhall,1989)

But it looks like these specifications should also hold for palatal stops, and therefore would require additional theoretical devices like e.g. a more differentiated notion of gestural blending - which seems to be missing. This to some extent surely is due to the fact that AP was developed for American English in the first place, and there was no necessity to make more fine-grained differentiation of tongue body articulations. In order to fill this gap, Recasens and colleagues have developed the DAC-scale (DAC=degree of articulatory constraint, e.g. Recasens 2002). This scale[14] makes detailed predictions on the coarticulatory behaviour of the segments under consideration and therefore will be reviewed later in a separate section. Instead, the relevant findings from speech physiology are reviewed in the next section, because Recasens, in contrast to AP and Task Dynamics, refers to physiological approaches: Recasens quotes Perkell's (1969) distinction between extrinsic muscles governing vowel gestures and intrinsic muscles governing consonantal constrictions. Recasens conjectures that the production of consonants with a high DAC value - like palatals and velars - involve extrinsic tongue muscles. Consequently, the relation between vocalic and "extrinsic" and consonantal and "intrinsic" is not a perfect one and

[14] Although with a different theoretical emphasis. Simplified, AP has a stronger focus on temporal gestural overlap, while the DAC emphasizes the idea of biomechanical coupling between articulators.

holds only for a subset of consonants. This holds in particular for the - dorsal - consonants under consideration, and it therefore becomes necessary to review the basics of lingual speech physiology.

Excourse: tongue physiology

As mentioned earlier as a result of acoustic simulations (see section 2.1), high formant onsets for velar stops before front vowels - as measured from speech acoustics or as typically used in speech perception experiments - are not achievable by a simple fronting of the place of constriction alone, and additional widening of the pharyngeal cavity is also required. This is due to the fact that there is no extrinsic muscle which would “pull up the tongue in direction of the palate” to form an /i/-like shape. This then has to be achived by other means: The genioglossus is attached to the mental spine of the mandible and its insertions (e.g. Maeda und Honda, 1994). The posterior part (GGP) of it can have the desired effect by contracting in longitudinal direction, which has the effect of moving a considerable amount of tongue tissue in an upward direction. In order to achieve configurations close to the other corner vowels, /u/ and /a/, two other extrinsic muscles are important: The styloglossus (SG) arises from the anterior and lateral surfaces of the styloid process and has the function of pulling the posterior part of the tongue back - and upwards resulting in tongue shapes like those for back vowels, in particular /u/. The hyoglossus (HG), attached to the hyoid bone functions in forming an /a/-like shape pulling the tongue downwards. An important point to notice is that anatomically, unlike in most systems of skelettal musculature, the division in agonists and antagonists is not as easily possible for the tongue. Nevertheless, this has been undertaken forming two *functional* groupings between (a) styloglossus and genioglossus anterior (GGA) and (b) hyoglossus and genioglossus posterior (see figure 1.8, after Maeda & Honda, 1994).[15]

Within the terminology of Hardcastle, this is the “tongue-body” system which is under the regime of the extrinisic tongue musculature as just described. It is supplemented by the “tip-blade system”, which is responsible for finer movements of the tongue and predominantly under the regime of the intrinsic tongue musculature (see figure 1.9). This second system is only of minor importance here, but has been made responsible for the dominance of coronal sounds in sound systems, because only a limited number of articulatory degrees of freedom has to be used in order to generate the desired acoustic output once the gross positioning is achieved (Lindblom und Sundberg, 1971).[16] Further, for stops, the functioning of the mylohyoid has to be mentioned as well: This muscle forms the floor of the mouth and it seems quite generally acknowledged that it assists in lowering the mandible (e.g Perkins und Kent, 1986). The mylohyoid has a specific activation for /k/-like gestures as well. Its role is stiffening the floor of the mouth in order to support the genioglossus for /k/. It thereby is both active for /ki/ and /ku/ (K. Honda, personal communication). The “tongue-body” system has attracted more attention in modelling

[15]The figure is borrowed from course material by P. Hoole.

[16]Here is a short description of the most impotant intrinsic tongue muscles: The superior longitudinalis is a thin layer of muscle below the dorsum that shortens the tongue and curls its tip and sides. The inferior longitudinalis is a paired muscle on the undersurface that shortens the tongue and pulls the tip downward. The vertical fibers of the verticalis found in the sides of the tongue flatten it. The horizontal fibers of the transversalis from top to bottom of the tongue have the function of narrowing and elongating it.

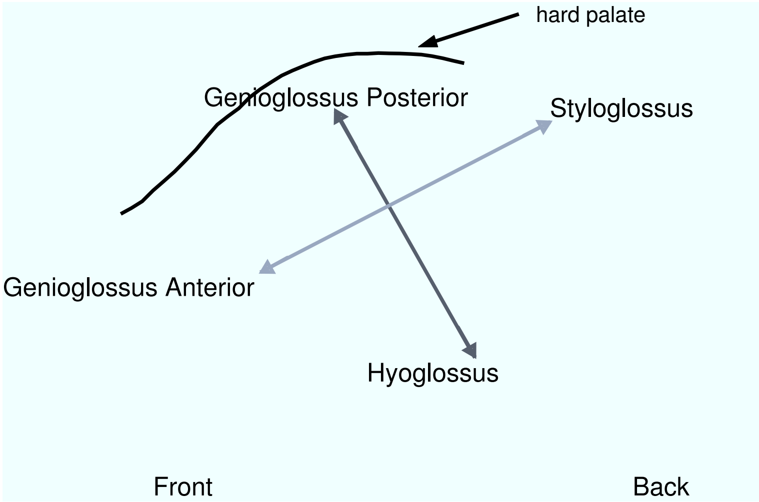


Figure 1.8.: The most fundamental agonist-antagonist pairs of the external tongue musculature (after Maeda und Honda, 1994).

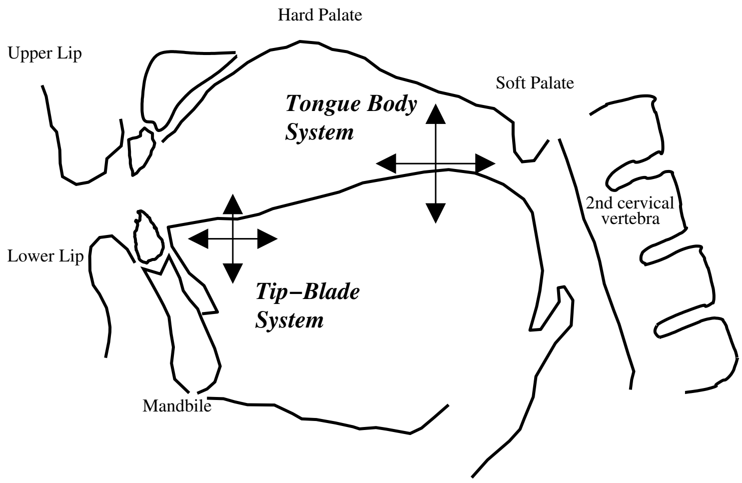


Figure 1.9.: Functional subsystems of the tongue (adapted from Hardcastle, 1976).

studies. Broadly speaking, the most important classes of approaches have been termed "statistical" (e.g Harshman; Ladefoged und Goldstein, 1977; Maeda, 1990; Nix; Papcun; Hodgen und Zlokarnik, 1996; Hoole, 1999; Beautemps; Badin und Bailly, 2001) and "biomechanical" models (e.g Wilhelms-Tricarico, 1995; Payan und Perrier, 1997). This is not more than a taxonomy though, because there also exist hybrid approaches (Sanguineti; Laboissière und Payan, 1997; Sanguineti; Laboissière und Ostry, 1998). Typical results of statistical approaches are those of Harshman et al. (1977), who extracted two movement components of sagittal X-ray data, the first mimicking a high front to low-back movement they termed "front raising" and the second mimicing high-back to low-front movement which was termed back-raising in analogy. Among biomechanical approaches, the model as outlined in Payan und Perrier (1997) is a representative of this class of approaches. This finite-element-model (FEM) is based on the Equilibrium Point Hypothesis (Feldman, 1986), the basic tenet of which is that movements are produced by centrally specified shifts of the peripheral motor system. More precisely, for each muscle a threshold length, λ is defined, which defines where active force starts. The adaption for tongue modelling is based on dividing the tongue into small volumes, which are connected by nodes and provide a representation of elastic properties of tongue tissues. The version of the model as described in Payan und Perrier (1997) controlled following muscles: Anterior and posterior genioglossus, styloglossus, hyoglossus and additionally the most important intrinsic tongue muscles, the verticalis and the inferior parts of the longitudinalis. Muscle forces are a function of muscle lengths, and the model is controlled in this λ-space. Mappings between these different kinds of approaches have to my knowledge first been undertaken in Sanguineti et al. (1997, 1998), making use of the statistical descriptions by Maeda (1990), but provided projections in a muscle space of a biomechanical model.

None of the physiological models hitherto provide an approach of all the muscles involved in the production of speech sounds. For example, the role of the mylohyoid for manner requirements was emphasized because in most of the models it is not implemented. A second important point is that of morphological separability and its accesability to computational approaches like FEM: For example, the genioglossus is regarded as a single muscle by physiologists, and its separation into anterior and posterior functional subgroups as in the functional antagonisms in the Maeda und Honda (1994) model was achieved for purposes of its use for the acoustic modelling of vowels. Similar arguments hold against representation of parts of the intrinsic tongue musculature: Longitudinalis and verticalis are not distinct regions of tissue, but rather are defined by their innervation properties in forming *functional* goals like e.g. tongue grooving.

Summary

What should be evident from physiological modelling is that it is highly complex in itself and does not provide an obvious interface to linguistic units. Therefore, Port und Leary (2005)'s rebellion against phonetic segments as formal symbol tokens and the need for representations did not put forward a replacement of phonetics by physiological approaches, but rather a concept of time more relevant for biological systems was the claim raised by these authors. Articulatory Phonology proposed both an interface to linguistic units and a concept of time suitable for biological systems in the sense of Port & Leary by acknowledging the need to group articulators into *functional* units. These are much more easily related to linguistic units, which might have been responsible for the striking success and impact AP had on the development of the discipline. This section only contained a rough sketch of Articulatory Phonology and Task Dynamics as the primary aim here was to point to alternatives to the "alphabetical" approach for speech production theories in general. A more detailed look at AP and their most relevant concepts for the present purpose, "tract variables" and "gestural score" will be postponed to sections more explicitly concerned with actual empirical work. The criticism of physiological models was not to discredit their usefulness and impact on the discipline, although their problems in mappings to linguistic concepts is recognized across the board in the scientific community.

2.3. Speech perception

Trading relations

Repeating the start of this part which introduced the feature system of Jakobson et al. (1952), there have indeed been feature system which made the attempt to describe distinctive features in acousto-perceptual terms. The experimental approaches mentioned so far have been centered around the idea of "acoustic invariance". Recapitulating, the acoustic invariance approach makes two major claims about the relationship between featural entities and acousto-perceptual mechanism, i.e. that (a) there is acoustic invariance in the speech signal corresponding to the phonetic features of a language. (Blumstein, 1986, p. 178) and (b) that the perceptual system is sensitive to these invariant properties. As was seen, the approach was essentially evaluated in the acoustic domain and one of its major results was the definition of "spectral templates" coding place of articulation. The second claim that the perceptual system is sensitive to such invariants was evaluated to a much lesser degree - with the exception of the study by Lahiri; Gewirth und Blumstein (1984) described below - by these authors themselves. With respect to this special issue under consideration, I also repeat the characterization for the behaviour of burst spectra in velar and palatal stop contexts (see also table 1.3): The velar is characterized by a broad mid-frequency peak in the region between 2 and 3 kHz in the context of front vowels and /a/ - relating to the feature compact, while there is no such peak present in back vowel contexts. Palatals in contrast show a high frequency peak at about 3.5 kHz, and a mid-frequency peak in the region between 2-2.5 kHz in back vowel contexts. To my knowledge, the study coming closest to an empirical test of these ideas is the one by Plauche; Delogu und Ohala (1997). These authors start from known asymmetries in consonant place perception attested from both historical sound change and laboratory confusion studies: A change from /ki/ to /ti/ appears to be much more common than the reverse. Plauche et al. (1997) hypothesize that such asymmetries "arise when two sounds are acoustically similar except for one or more differentiating cues, which are subject to a highly directional perceptual error" (p. 2187). The stimulus design strategy they applied is particularly interesting, therefore it follows a description with a fair amount of detail: The general aim was to process syllables like /ki/ and /pi/ by filtering so that they could enhance the asymmetrical confusion with /ti/, or, as Ohala himself[17] puts it: "The /ki/ contains within itself the "seeds" of its misperception as /ti/". The stimulus generation these authors applied mainly relies on the removal of the "compact" burst portion aforementioned, and the next section presents an informal

[17] Slides of the talk "An interpretive history of phonological science" given at the Centre for General Linguistics, Typology and Universals Research (ZAS), April 2007

evaluation of such a procedure: After segmenting burst and vocalic portions of CV-sequences taken from a female speaker of Hungarian, the burst portions were stop-band filtered in a first step. In the second step the modified burst portion was reconcatenated with the unmodified vocalic part of the same CV-sequence. This procedure results in (a) an original and (b) a modified version of the same signal. The procedure is exemplified in figure 1.10. Informal judgements of several listeners point to a robust alteration from a /ka/ to a /ta/ percept. The top left and right panels of figure 1.10 show the oscillograms of unprocessed and processed signals, corresponding sonagrams are shown in the second row. The ellipse in the right middle panel points to the frequency range of this "surgical" operation. The third row displays the raw spectra, with the circle again pointing to the affected frenquency band. Additionally the bottom panel displays the frequency response of the bandstop filter applied (right).

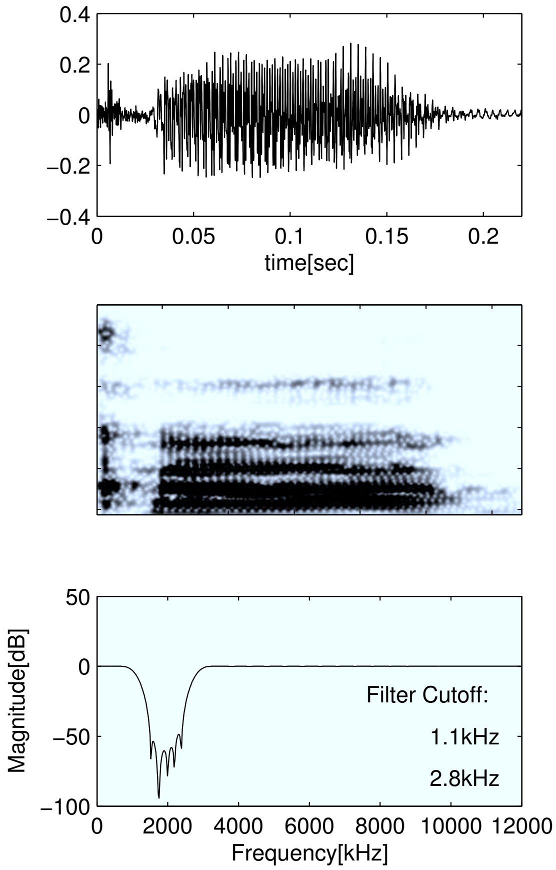


Figure 1.10.: Illustration of the destructive procedure "turning a velar into an alveolar". This is a conceptual replication of the signal processing routines from Plauche et al. (1997). Top panel: Original (left) and filtered (right) versions of the signals; panels second from top: sonagrams of original(left) and filtered (right) versions of the signal; panels second from bottom: raw burst spectra taken at burst portions, original (left) and filtered (right); bottom panel: frequency response of the filter

Some observations are in place though: Most of the time, this destructive processing was much more successful for /a/-contexts. On the contrary, in an /i/-context such a processing much more often either did not result in a convincing /ti/-percept at all or the perception was biased towards [ti] right from the start. This might be explained by the fact that the material taken for this "toy study" was taken from speech material acquired from a single female speaker of Hungarian with a tendency towards weak aspiration, which might be an idiosyncrasy but likewise the results of the phonetic implementation of stop bursts in this language.[18] Furthermore it was much harder to determine the exact cutoff frequencies for the filter. For the perception of place articulation, the study by Dorman et al. (1977) titled "Stop-consonant recognition: release bursts and formant transitions as functionally equivalent, context-dependent cues" can be considered as the hallmark and programme of this branch of perceptual research at the same time, and the phenomenon of a vowel dependency of the perceptual relevance of burst and transitions has been corroborated by several studies in which bursts were either deleted from utterances or presented in isolation. The general picture shows the release burst to determine place of articulation in front vowel context /i/, while formant transitions are dominant in vowel context /a/. This study has stimulated much work in the domain of place-of-articulation perception. In my own view, one of the most diligent studies in this field was the double paper by Smits et al. (1996a,b), which is already evidenced by the impressive amount of cue variables considered in their study (Smits et al., 1996b). Surely, this catalogue is not a managable amount of experimental variables to be subjected to an experimental study if you consider a full-factorial experimental design. Rather, these authors used a two-step strategy to reveal the effects of fine phonetic detail on perceptual judgement: In a first step, stimulus material was designed and presented to listeners in an identification task. This stimulus material was described in terms of the acoustic descriptors shown in figure 1.11. The second part of the experiment then (Smits et al., 1996b) consisted in a mapping from the acoustically measured cues of the stimulus material presented to the actual judgements made by human listeners by means of machine learning algorithms.

Before turning to a description of the results, which are also relevant for palatal segments researched in this work, it seems advisable to clarify the gross cue - detailed cue distinction shown in figure 1.11. Following Smits et al. (1996a,b), detailed cues require a high resolution along the time or the frequency dimension, and the authors list the burst length, burst frequency, formant onset frequencies etc. as examples. In contrast, gross cues are "less clearly visible in the sonagramm." Proponents of such a cue conceptualization make "no explicit distinction between bursts and formant transitions." Instead, "integrative structures" are proposed as being the main cues for perception. The first papers advocating the "gross-cue-apporach" are more recent and have been mentioned already (Stevens und Blumstein, 1978; Blumstein und Stevens, 1979, 1980). Lahiri et al. (1984) aimed at perceptual a evaluation: In this study, labial and coronal initial CV-sequences were generated by means of a parallel formant synthesizer. The design rationale was to generate a conflict between transitions cuing one place of articulation and the *change* in global spectral tilt cuing another place. The intended result was achieved: Participants predominantly classified the stimuli in accordance with the gross cue, the change in spectral tilt. This study was subjected to methodological criticism by Lindholm et al. (1988): They claimed

[18] The description in the handbook of the International Phonetic Association (1999) actually points in this direction: Hungarian voiceless stops are described as unaspirated, the voiced counterparts as fully voiced.

DETAILED CUES

1. L_b: total level of the burst
2. F_{bp}: frequency of the spectral peak of the burst.
3. L_{bp}: spectral level of the peak burst
4. l_b: burst length
5. $F2_O$: frequency of F2 at voicing onset
6. $F2_st$: frequency of F2 in the stationary part of the utterance
7. $F3_O$: frequency of F3 at voicing onset
8. $F3_st$: frequency of F3 in the stationary part of the utterance
9. D_l: distance to the labial locus equation
10. D_d: distance to the dental locus equation
11. D_v: distance to the velar locus equation

GROSS CUES:

1. T_O: spectral tilt at onset
2. ΔT_O: change of spectral tilt after onset
3. T_{st}: spectral tilt in the stationary part of the utterance
4. L_0: level of the mid-frequency peak at onset
5. ΔL_O change of the level of the mid-frequency peak after onset
6. L_{st} level of the mid-frequency peak in the stationary part of the utterance
7. F_O^{mfp}: ERB rate of mid-frequency peak at onset
8. F_{st}^{mfp}: ERB rate of mid-frequency peak in the stationary part of the utterance.

Figure 1.11.: A list of acoustic cues relevant for the perception of place-of-articulation, split by membership to the detailed versus gross class, after Smits et al. (1996b).

that the transitional information was too impoverished to warrant a fair comparison. In the same vein, the results by Smits et al. (1996b, p. 3865) evidence that "in all cases, the detailed cues, such as formant transitions, give a better account of the perceptual data than the gross cues, such as the global spectral tilt and its initial change". Therefore the study by Lahiri et al. (1984) will not be reviewed in greater detail, rather the next paragraph returns to results of the study by Smits et al. (1996a):

According to Smits et al. (1996a), velar bursts generally are more effective than burst cues for other places of articulation. The relative contribution of bursts is larger in the context of high front vowels, the reverse holds for /a/-contexts while generally the contribution of velar bursts is larger than that of alveolar or labial bursts. The contribution of transitions was found to be equal for voiceless /p,t,k/-transitions, while for the voiced counterparts /b/-transitions are a more effective place coder than /d/-transitions. This converges well with results from experimental work by Fischer-Jørgensen (1972) with regard to the fact that (a) burst cues dominate the perception in /i/-context, (b) that the relative contribution of transitions is more important in /a/-contexts, (c) that /k/ always required a velar burst and (d) /p/-transitions could only be overridden by a /k/-burst. Although some other findings by Fischer-Jørgensen (1972) could not be confirmed[19] by Smits et al. (1996a), both papers cast the interpretation in a direction which is well described by the title of a proceedings contribution by Smits: "Context-dependent relevance of burst and transitions for perceived place in stops: It's in production, not perception." Such an interpretation is in fact evidenced by the results of studies by Fischer-Jørgensen (1972) as well as Smits et al. (1996a): Consider as an example the tongue positions at velar closure and at a following

[19] For example, Fischer-Jørgensen (1972) reported that the combination of a /tu/-burst with a /ku/-transition often resulted in a /pu/-percept.

vowel target for the articulation of a CV syllable. It is obvious that the tongue articulator does not have to travel a long - spatial - distance from the closure to the vowel target for /i/. This is - although in a nonlinear way as shown by Quantal Theory (Stevens, 1998) - also mirrored acoustically in the formant transitions being relatively unpronounced in comparison to a CV sequence where the vowel is e.g. /a/. This then is reflected in the perceptual domain in the greater relative importance of transitional cues for the low vowel as described in the studies by Smits et al. (1996a) and Fischer-Jørgensen (1972). A similar line of reasoning can be applied for the greater perceptual importance of transitional cues for the perception of the bilabial: The spectral content of bursts in oral stop consonants is shaped by the cavity between the oral closure and the lips. For the bilabials, there is no such cavity which could affect the spectral properties of the burst. Smits (1996, p. 2473) therefore argues that

> the classification model is not "actively" adjusted per vowel context. Context-dependent cue-weighting mechanisms, [...] are therefore not necessary for an adequate modelling of the classification behaviour. Obviously, these vowel dependent differences in the cue distributions originate in articulatory processes. We therefore conclude that the context-dependence of the perceptual relevance of burst and transition cues is not caused by any perceptual reweighting processes, but by differences in acoustic cue distributions generated in production.

The lesson to learn from this is that any simplistic attempt to compare relative perceptual contribution cross-linguistically will be biased by the language-specific articulatory implementation of the phonemes under consideration. Such cross-linguistic differences in the implementation of the stops systems were already mentioned while discussing the comparatively weak aspiration of the Hungarian stops while introducing the paper by Plauche et al. (1997) in comparison to the implementation of stops in e.g. West-Germanic languages (see for example Jessen, 2000).

Another helpful taxonomic building block is explicitly introduced - to my knowledge for the first time - in Smits et al. (1996a), the distinction in stimulus and therefore experimental design between

- *deleted cue-stimuli* as stimuli from which one or more cues are removed. Experiments with deleted cue-stimuli measure the *necessity* of the deleted cues and the sufficiency of the remaining cues.
- *conflicting cue-stimuli* for which cues are present which point to different response categories. These experiments measure the *relative importance* of the conflicting cues.

Experiments conducted in the *conflicting cue-stimuli* paradigm were discussed presenting the sudies of Smits et al. (1996a) and Fischer-Jørgensen (1972).[20] This kind of experiment is usually evaluated in the context of an "Expanded Factorial Design" (Massaro, 1987; Massaro und Cohen, 1993). An Expanded Factorial Design is derived from a standard full-factorial design in the following way: Consider you have two experimental variables/factors with m and n levels respectively, then the full crossing of these factors will result in m x n experimental stimuli. The

[20] Of course, this is not intended to be a comprehensive summary and several studies have been arbitrarily omitted (e.g. Neagu und Bailly, 1997).

expanded factorial design produces all combinations of the factorial design, but additionlly the m and n levels of the experimental variables in isolation. In this two-factor example then, the total number of combinations in the expanded factorial design will be n x m + n + m. The application of an expanded factorial design for place of articulation recognition in a typical experiment with three places of articulation is illustrated in table 1.5.

		Burst			
		/p/	/t/	/k/	
	/p/	**or**	mb	mb	nb
trans	/t/	mb	**or**	mb	nb
	/k/	mb	mb	**or**	nb
		bo	bo	bo	

Table 1.5.: Expanded Factorial Design and its application to place of articulation studies (adapated from Smits et al., 1996a). Abbreviations: or - original, mb - mixed burst, nb - no burst, bo - burst only.

There are also numerous works in the literature in which the formant transitions have been replaced by silence or a steady-state vowel. These are best seen as one particular instance of the aforementioned *deleted-cue* experiments. Examples include e.g. the studies by Winitz et al. (1972) or Dorman et al. (1977). Recognition from these burst-only studies typically ranges between 50-80%. There are also numerous deleted-cue experiments in which the burst has been replaced by silence, with only transitional information left in the signal (Ohde und Sharf, 1981; Fischer-Jørgensen, 1972; Dorman et al., 1977). Typical recognition rates range from 40-80%, with the general trend that velars were recognized worse than bilabials and alveolars. Wrapping this paragraph up, spectral properties of the burst may be particularly salient in the environment of front vowels where transition motions are minimal, whereas they may play a lesser role in the environment of back vowels where formant motions are larger.

Categorical Perception

Admittedly, the organization of this short introduction to place-of-articulation perception might look awkward at first sight, because the knowledge about context-dependence has historically been discovered much earlier and therefore is not an achievement of the studies reported in the preceeding paragraph about trading relations: The *locus theory* elaborated in a series of papers (e.g. Liberman; Delattre; Cooper und Gerstman, 1954; Liberman, 1957; Liberman; Harris; Hoffman und Griffith, 1957; Delattre, 1969) the concept of the locus as the hypothetical starting point of the formant transitions that characterize stop consonants - at least in CV sequences - acoustically. The approach was sucessful for all the consonant places except velar place: Two loci were needed distinguishing between velar stop place in the context of front and back vowels. The findings were summarized by Delattre as follows:

> The physiological explanation is that velars are favoured by better resonating conditions of the cavity in front of the burst's place of production than the other plosives, and that lip rounding is a further advantage for efficient resonance. (Delattre, 1969, p. 23)

This paper can be seen as the endpoint of the development of the *locus* theory searching for *invariant* place of articulation cues in the perceptual domain. More important than its actual results is its value as an instantiation of the phenomenon of *Categorical Perception* (CP) within phonetics: Categorical perception since then means that a change in some variable along a continuum is perceived, not as gradual but as instances of discrete categories. For example, the classical experiment uses synthetic syllables in which the movement starting point of the second formant is varied in equal steps. Depending on the precise parameter values, the resulting sounds were perceived as 'ba', 'da' or 'ga'. Although the change is gradual, this is not the way it is perceived. Subjects regularly perceived the different stimuli as being instances of either of the three syllable types, 'ba', 'da' or 'ga'. It is just within the locus paradigm that Categorical Perception has first been established within phonetics.[21] The next part briefly reviews this experimental paradigm:

Essentially the application of the CP paradigm to speech stimuli consists in the acquisition of parallel identification and discrimination data on the same stimulus continuum. The following listing gives brief descriptions of the tasks applied in a prototypical CP experiment like for example of voicing or of place perception:

DISCRIMINATION **AX** In the AX task, the perceiver has to give same-different judgements being presented with two stimuli.

- **ABX** In the ABX task, the subject will hear three stimuli in sequence, and has to respond whether X is more similar to the first (A) or to the second (B).
- **4I-oddity** In the four-items-oddity task, the subject will hear four stimuli in sequence, and has to say whether the second or the third is the odd one out of the four. The other three substimuli are identical.

IDENTIFICATION The perceiver has to decide between a given set of response alternatives.

The procedure works in the sense that given that perception is categorical, it is possible to predict discrimination behaviour from identification/labeling responses.[22] The typical effect is that there exists a difference in behaviour at the category boundary, more precisely, that at the category boundary, discriminability between stimuli is maximal. This effect later has been termed the *category boundary effect* by Wood (1976). This effect is shown for a hypothetical continuum in figure 1.12.

It should be noted that almost all of the work within phonetics inspired by CP did not primarily aim at revealing language-specific properties of speech perception but rather were focusing

[21] Inspired by Thurstone (1927)'s *law of categorical judgement* which first introduced the concept of a *psychological continuum*. The fundamental idea of the law of categorical judgement is that it is possible to scale a stimulus set based on simple comparisons between stimuli two at a time, i.e. on the basis of pairwise comparisons.

[22] Assuming that discrimination is underlyingly identification and comparison. For a formal treatment of these prediction models see Pisoni (1971); Pisoni und Tash (1974); Pollack und Pisoni (1971).

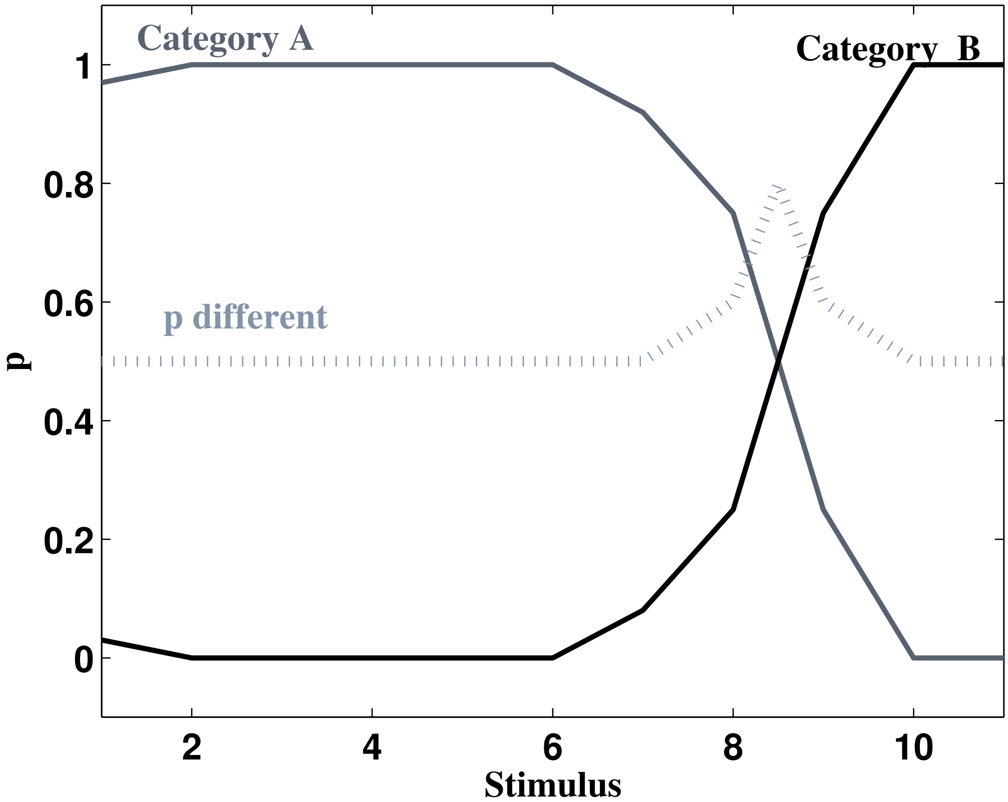


Figure 1.12.: Illustration of results of a Categorical Perception experiment on a hypothetical two-category continuum. For further expanations see text.

on revealing the origins of Categorical Perception in human speech processing itself. A gross taxonomy could divide them into three groups, (a) those in which the phoneme boundaries arise from a perceptual process that is linked to the way sounds are articulated (as in the motor theory of speech perception (e.g Liberman et al., 1967), (b) those in which the phoneme categories arise from category learning like Lane (1965), and (c) those in which categorical phenomena in speech perception arise from inborn sensitivities of the auditory system (Stevens und Blumstein, 1981).
Some further comments are in place here: First, applying the traditional notions of CP for prediction of discrimination behaviour typically resulted in a systematic underestimation of discrimination behaviour which led to further developments in the framework of *Signal Detection Theory*.[23] A second point to note is that the categoricalness of perception typically was stronger in experimental work on consonants than on vowels, which will not be discussed further here. Third, and most important here, Repp und Liberman (1987, p.105) note that

> "unfortunately, cross-linguistic studies using the same stimuli and procedures are not very numerous. Among those that exist, most have dealt with the voicing dimension, as cued by VOT".

Summing up, there has been little crosslinguistic research within the experimental CP paradigm in the domain of place-of-articulation perception, which seems rather surprising in the first place given the advanced methodological apparatus described in this section. Possible reasons for this lack of crosslinguistic research on place-of-articulation could consist in experimental problems for the design of such studies: CP requires stimulus continua, and the choice of continua be related to an arbitrary choice of the subset of cues given Repp's (1984; 1987) operational definition of "cue", according to which a cue is an entity which can be *manipulated by a speech synthesizer*:[24] In other words, experimental results could always be interpreted in terms of cues not manipulated in a particular experiment. Furthermore, the interpretation of such an experiment would be aggravated by the fact that usually different languages have different vowel systems again generating alternative hypotheses about differences in the categorical behaviour. These experimental problems seem to run contrary to the traditionally strong explanatory impetus of the CP paradigm. On the other hand, consider the outcome of a hypothetical *cue-trading* experiment designed to uncover cross-linguistic differences in the relative weightings of different cues in different languages, for example the relative weighting of burst and transitions in the perception of velar stops. The implication of a likewise hypothetical result like for example "the relative importance of burst cues in comparison to transitional cues is bigger in language A than in language B" faces similar problems: An explanation in terms of language specific detail of consonant-place cue-structure and vowel system would always be at hand, making such experiments problematic to defend against objections of triviality if interpretations in terms of human speech perception were desired. In sum, most researchers apparently deemed the dimensionality of potential influences more managable for voicing perception, while cross-linguistic place perception was not researched with the same intensity if not neglected altogether.

[23] Emphasizing the role of *phonemic memory* and *timbre*

[24] More recent approaches based on advanced signal processing techniques are not explicitly discussed.

Summarizing, the perception of place of articulation has, compared to voicing perception, been studied considerably less from a cross-linguistic perspective. I assume that this is primarily due to methodological problems, which were pointed out while discussing studies concerned with trading relations in place perception and by recalling the necessity of a *true* continuum for the demonstration of *category boundary effects*. Finally, it has to be noted that the account of perception studies chosen arbitrarily leaves out all lexical levels: Goldinger et al. (1990) make an acknowledged distinction between *Spoken Word Recognition* and *Speech Perception*. The latter with its strong "bottom-up" orientation has traditionally focused on segmental cues and generally researches aspects "below" the word. It will be exclusively studied in the present dissertation arbitrarily omitting any recourse to the lexicon.

3. Summary and outlook

This work so far has been a coarse summary of the aspects determining the implementation of phonetic categories, with special emphasis on dorsal segments. In this form this introductory chapter was not apt for the purpose of explicitly deriving hypotheses or planning experimental work and needs reworking. This section summarizes what has not been said and that part of what is missing and in some sense is intended to "clean up" for beginning with the experimental chapters. There were many allusions to the basic ideas of the present work, but often distributed over several sections. As the intention is to work in acoustic, perceptual and articulatory domains, the experimental chapters are separated and contain separate theory sections. I found this an acceptable solution because it does not appear reasonable to debate e.g. the contribution of the jaw for place of articulation alternations in the proximity of perceptual data. I am aware though that this strategical decision comes at the cost of some unavoidable redundancy. The empirical studies furthermore are best seen as isolated contributions in the first place, with the attempt of a synthesis in a final, more general discussion concluding this dissertation.

3.1. Perceptual experimentation

This paragraph aims at integrating the information delivered until so far in order to guide the design of perceptual experimentation in the following empirical part: The introduction mentioned that many substance-based approaches for the typology of vowel systems have been proposed in the tradition of Liljencrants und Lindblom (1972), but this kind of research has not attracted the same amount of attention for consonant systems with the exception of the recent attempt by Abry (2003). Approaches for the acousto-perceptual substantiation of distinctive features have evolved at least from the traditions of the "Motor Theory of Speech Production" and the "Acoustic-invariance" attempts. The endpoint of the development of these approaches was in both cases a turning away from their radical initial intent: Delattre (1969) had to concede the relative importance of burst cues in the perception of velars in non-front vowel contexts, later development of the invariance approach had to focus on dynamical aspects of the burst spectrum contrary to the originally static conception of spectral shape as a specifier for place of articulation. In other words, both approaches have been frustrated by the attempt to identify consistent articulatory manuevres underlying speech production or the attempt to identify primitive acoustic-auditory elements in speech perception. Later theoretical development loosened the connection between distinctive features and articulatory or perceptual parametrization by either allowing intermediate levels of representation like the distinction between “cues”, “specifiers” and “invariants” (Fowler, 1994) or by generally assuming weaker relations between signal parameters and features (Nearey, 1996). These have not been discussed in depth yet but will be in discussing experimental results. For the moment the distinction between “deleted” and “conflicting” cue experiments and Repp’s comment that crosslinguistic experimentation in the CP paradigm has almost exclusively focused on voicing perception is sufficient to motivate the aims of the perception study: Is a four-way place-of-articulation distinction implementable on the basis of second and third formant transitions alone?

3.2. Production experiments

Although the coarticulation concept and the Gestural Approach were introduced, their description remained mainly on the surface, and the necessity of a more detailed elaboration remains. This will be indispensable for the design of coarticulation studies aiming at a cross-linguistic comparison of palatal and velar obstruent articulation by means of movement data as aquired by the articulograph. Further theoretical building blocks still missing are the abovementioned DAC-scale, and a taxonomy of palatal articulations as elaborated by Daniel Recasens and colleages (Recasens, 1990, 1997, 2003; Recasens und Espinosa, 2006). Although the scope of this work clearly is articulatory modelling of dorsal articulations, a description of the phonology of the segments is included as well. While introducing Keating’s conceptualization of underspecification in phonetics, the problem of separating different levels of representation in the framework of Lexical Phonology arose. This problem recurs in the phonological description of the *ich*-Laut *ach*-Laut alternation, but between lexical and postlexical levels of the derivation. An attempt at an alternative solution will *not* be made though.

It is hoped that explicit cross-linguistic comparisons between German and Hungarian realizations of velar realizations have theoretical relevance for coarticulation theory: Coarticulation and other properties of phonetic implementation often are assumed to follow from universal principles of speech physiology. However, it has been shown that degrees of coarticulation can be language-specific (Öhman, 1966; Manuel, 1990; Manuel und Krakow, 1984), and there is evidence vice versa that coarticulation is constrained by the phonological system of a given language. The study performed by Manuel und Krakow (1984) showed that the perceptual tolerance for variation in the production of a vowel is lesser in a language with a filled vocalic system than in a language with a less crowded vowel system. Hence, the density of the vocalic system apparently has an impact upon vowel perception. This hypothesis is weakened, however, by the results obtained by Manuel (1990) in a comparative study of three languages which varied little between each other as concerns their vocalic space. This should be understood only as an example, which is illustrating that again, vowels were the primary target of cross-linguistic variational change. An attempt will be made to answer questions like the ones posed by Manuel and colleagues (a) in the articulatory domain and (b) for consonants. This calls for a derivation of speaker-independent methodology in order to derive fair contrasts between speakers from different languages, produced by interindividually different vocal-tract morphologies.

4. Conclusion

Summarizing, the motivation for the experiments planned has mainly been taken from applying the basic concept of increasing the category density in a given phonetic - i.e. physically determined - space. But almost the same experimentation could be derived from a fully disparate theoretical perspective: the perspective of sound change. This perspective has been almost completely left out so far with the sole exception of discussing Acoustic Invariance theory. Still, the topic deserves some attention, although the appreciation in this concluding section might still be argued to be insufficient. It does not constitute the main topic though, and therefore will only be quickly introduced along one of the main lines, i.e. whether sound change is mainly conditioned by articulatory or acousto-perceptual factors.

In line with the Neo-Grammarian tradition, velar palatalization sound changes have long been considered to be the outcome of gradual articulatory changes resulting from the natural phonetic process of contextually induced articulatory fronting. More recent approaches often emphasize the role of perceptual factors driving sound change. From a phonetic perspective, Guion (1996, 1998) tried to account for data on velar fronting in perceptual terms, she proposes and delivers experimental evidence that velar palatalization "is a result of a perceptual reanalysis of fronted velars in fast speech" (Guion, 1998, p. 22).

Phonological approaches attempt to reflect palatalization tendencies in perceptually-oriented low-dimensional markedness scales (e.g. Padgett und Zygis, 2003; Z̊ygis, 2003). This is a reflex of the fact that earlier phonological theory relying on articulation was not able to correctly predict the frequent occurrence of [tʃ] as output of velar fronting sound changes (see the description of feature geometric systems in part II). Furthermore, this tendency surely is stimulated by work by Ohala and colleagues done over the last few decades (e.g. Ohala, 1974; Hombert; Ohala und Ewan, 1979; Ohala, 1983, 1993), who not only advocated the use of experimental phonetic methodology in historical linguistics, but was among the first emphasizing the role of perceptual factors for linguistic processes.[25]

In some sense an intermediate position is taken Kiparsky (1988, 1995) who is interested in the typological taxonomy of sound changes. He organizes sound changes into two major classes. He terms them *major* and *minor* sound changes. The first group, i.e. major sound changes are conditioned by articulation, are gradient in nature, and are driven by articulatory similarity. They can result in the introduction of new segments and their combinations into a language's paradigm, and they are exceptionless.[26] In contrast, minor sound changes can be rooted in perception

[25]In the form of misperceptions or misparsings of auditory input: coarticulation leads to distorted production, which the competent listener is able "factor out" - at least under normal circumstances. This is the default scenario which Ohala calls "correction". But, two further scenarios actually lead to what Ohala calls "mini-sound changes": hypo-correction and hyper-correction. In the hypo-correction scenario, the listener fails to apply reconstruction strategies, and as a result, "the phonetic perturbation, originally just fortuitous results of the speech production process, become part of the pronunciation norm (Ohala, 1993, p. 246)". Such a scenario can occur for different reasons like e.g. the lack of experience to factor out details of the physical properties of speech like in the case children acquiring the language as well as adult second language learners or the failure to perceive the conditioning environment of the phonetic variation. In the second scenario, the hyper-correction scenario, the listener applies reconstruction strategies where this is inappropriate, i.e. correction is applied where it is not in place and generates a new pronunciation which the listener adopts as being a "listener-turned-speaker".

[26]This actually is in defense of what Kiparsky (1988) calls the "exceptionsessness hypothesis"(EH), i.e. sound changes as exceptionless, phonetically conditioned process rooted in the mechanism of speech production in

and/or acquisition, are driven by acoustic similarity, are discrete in nature, structure-preserving and can be sporadic rather than exceptionless. Contrary to the trend observed in current research literature putting the emphasis on perceptual factors, the articulatory hypothesis for the velar palatalization sound change is also upheld by some researchers (see e.g. Recasens, 2003). Recasens (2007, p. 2) further raises a substantial argument against strong perceptual accounts of velar softening: "This process has affected not only aspirated stops with long and salient releases composed of frication burst and aspiration noise (e.g., in the Germanic languages), but also unaspirated stops exhibiting shorter and less perceptible releases with frication noise only (e.g., in Romance)." This discussion will not be furthered here, it rather is intended as the concluding remark: Much of the following experimental work is characterized by a dual interpretability (a) in terms of "category density" effects and (b) as research on phonetic factors contributing to a typologically common sound change.

the neogrammarian tradition: "The existence of an important class of exceptionless sound changes grounded in natural articulatory processes is not in doubt, of course. It is the claim that it is the *only* kind of sound change that is under question." (Kiparsky, 1995, p. 640)

2. The cross-linguistic comparison of formant transitions

If you recall the state to which the material on speech perception was developed in the introductory part, this chapter is designed to conduct crosslinguistic experimentation on place-of-articulation in the CP paradigm. As mentioned, this type of study is fairly rare[1], crosslinguistic research focusing almost exclusively on voicing perception. As also argued already in the introductory part, this at least partly can be attributed to methodological problems of deriving ways to construct experimental stimuli and suitable experimental designs due to problems related to properties of the perceptual system as discussed at length in the literature on trading relations. Another cornerstone of the argumentation which need not be treated at length again is the distinction between deleted cue and conflicting cue experimentation. This distinction will become relevant again when the choice of the neutral vowel will be motivated in section 4.1 while focusing on transitional cues. These are seen as "movement encoders" which hopefully warrant topological (re-)projections of subjects' decision behaviour into physiological stimulus space and therefore will make it amenable to crosslinguistic comparison. This reprojection into stimulus space alludes to the notion of "Territorial Mapping" (as developed in Nearey, 1997, and succeeding publications). The method for deriving Territorial Maps deviates from the Nearey original; therefore, a separate section will be devoted to motivating this departure (section 5.4).

This chapter restates work (Geng et al., 2005) which had a stronger emphasis on speech category formation and development (see sections 1 and 2), but puts the emphasis on the architecture of sound systems, in substance-based fashion, very much in the sense of Liljencrants und Lindblom (1972), but for consonants: Is a stop system in a language robustly implementable in terms of transitional parameters alone, neutralizing much of the cue structure normally available to the listener by experimental design? The particular experiment aspired will make the attempt to deprive the listener of (i) (almost) all cues except transitional information and (ii) the possibility to make perceptual adjustments for different vocalic contexts. In this sense the experiment is a "deleted cue experiment" in the terminology of Smits (1996). Further, in order to avoid problems with potential contextual adjustments made by the perceptual apparatus in different vocalic contexts, the experiments based on synthetic stimuli were conducted in the neutral vowel context. The French data depend on the courtesy of Caroline Bogliotti to allow the reanalysis of data from her PhD (Bogliotti, 2005) in a crosslinguistic fashion. The Hungarian data were recorded in parallel with these already existing materials. The leading question was whether patterns of categorical perception can be influenced by the deprivation of a) burst cues and b) the possibility to make contextual adjustments to vocalic contexts while increasing the category density in a given perceptual space. This was undertaken in order to push the capacity of the perceptual apparatus to its limits. The theoretical part reviews models of speech category development, as

[1] According to Repp und Liberman (1987, p.105). See the quotation in 2.3.

well as the attempts to construct acoustic invariance from transitional or burst cues. This creates the redundancy of repeating parts of the Blumstein (1986) invariance account, but supplements it by their principled arguments against “transitionalists” like Sussman et al. (1998).
After describing methodological peculiarities and results, the discussion shifts the focus on to more recent theories tackling the relation between distinctive features and their correlates. This is achieved by reviewing the idea of a “double-weak” theory of speech perception (Nearey, 1995) in moderate depth for illustrative reasons.

1. Models of speech category development

In order to learn to perceive and produce one’s native language, the pre-linguistic child not only has to isolate and segment the individual units in the stream of speech, but also represent in some way the information which specifies the regularities among various productions of the same phoneme or word, and ignore irrelevant variations. In infant speech perception, one major focus has been the role of linguistic exposure or experience for the developing perceptual apparatus: In early infancy, before 6-8 months of age, infants not only discriminate native phonemic contrasts, but also discriminate contrasts involving syllables which are not used to distinguish meaning in their native language. This behavior distinguishes infants from adults who have difficulties discriminating syllables differing by only a single phoneme under many circumstances if that particular phonemic contrast is not used in their native language. The relevant decline in these abilities occurs late in the first year of life: Most infants between 6-8 months can discriminate non-native contrasts, very few of the 10-12 month old are able to do so. Nevertheless, this decline is not always evident between 10-12 months: Best (1995) shows that this decline only occurs for contrasts which involve sounds which are similar to sounds used in the native language. Best’s interpretation of these findings conceptualizes experience-dependent assimilability to the phonology of the native language rather than linguistic experience per se accounting for this retention of discriminative abilities. But still, undoubtedly, language-general perceptual sensitivities in newborns undergo a change and become more language-specific during the first year, and the most obvious interpretation of these findings was that they arise from loss of perceptual capacity due to the lack of experience.

Counter evidence against such experience-dependent reasoning comes from findings that infants continue discriminating some non-native phonetic contrasts even though they have never heard them (Best et al., 1988) and lose others although they are part of heard speech (Pegg und Werker, 1997). Now, these findings reported can be accommodated in a series of (meta-)theoretical approaches briefly sketched here: For strong nativist models, infant speech perception capabilities reflect the operation of a special-purpose speech processing module. For a particular device, all possible parameters for language-universal rules are present at birth, and the contribution of specific experiences is to select some settings appropriate for the native language while others are deactivated. The probably most influential of these theories has been the “Motor Theory of Speech Perception” (Liberman und Mattingly, 1985) according to which specialized computational routines analyze phonetic input in terms of the potential mode of production. In contrast, evolutionary models reflect the initial speech perception capabilities as the operation of a general auditory perceptual system developed during evolution. Age-related changes in speech perception reflect processes of self-organization of these systems. Therefore, no specialized module or built-in device is required for the perception of speech. Rather, general purpose auditory sensitivities are seen as contributing to speech perception, and language is said to have evolved either through phylogeny or ontogeny to take advantages of these auditory sensitivities.

The most reknowned of these approaches is most probably Kuhl's Perceptual Magnet Model: According to the Perceptual Magnet approach, in early life, there are some regions in the vowel space that are more stable than others. yielding some rudimentary neural/perceptual organization of the vowel space. Repeated experience remodels the perceptual space by setting up the more frequently experienced sounds as stronger attractors, i.e perceptual magnets (Kuhl; Williams; Lacerda; Stevens und Lindblom, 1992).

2. A more data-driven classification

Another, more data-driven classification arises, if the initial perceptual space is related to the perceptual space assumed in the adult listener. Such a classification is offered in a paper by Phillips (2001). According to this paper, different hypotheses are conceivable for the mapping of natural phonetic settings onto phonological ones: One, as advocated by Werker und Tees (1984), postulates that the natural, universal phonetic boundaries stay present in the adult population to various degrees, i.e. a subset of the innate perceptual boundaries is taken to form an additional phonological decoding level. In Phillips' terms, an approach like this is called a "structure-adding" approach based on the central tenet that "developmental change does not involve loss"(Werker, 1994). Contrastively, approaches, in which the change from infant to adult perceptual space involves rather a reshaping of the infant representations and an erasure of "superfluous" non-native boundaries than the selection among existing ones, are termed "structure-changing" approaches: Theorizing in line with Kuhl's Native Language Magnet, prelinguistic representations are replaced by new ones through perceptual magnet effects. In Kuhl's approach, categorical perception phenomena are related to the prototypes: In the vicinity of the phoneme prototype, vowel discrimination is more difficult, i.e. exemplars close to the prototype are hard to distinguish, while the opposite holds for exemplars farther away from the prototype. Note that the prototype model, although primarily invented for vowel perception has been extended to consonant-place-identification (Davis und Kuhl, 1994).[2] A third possibility synthesizes claims from structure-adding and structure-changing approaches (Serniclaes, 1987): In category-development, new category boundaries are created through perceptual couplings between predispositions. However, primitive boundaries still remain discriminable in some conditions, and may be conducive to "allophonic perception", like in people affected by dyslexia (Serniclaes et al., 2004). In this study the authors demonstrated poor discrimination performance for dyslexics in comparison to normal controls at the category boundaries but improved discrimination performance for within-category stimuli, i.e. impoverished categorical perception. Convergent evidence comes from a Mismatch Negativity (MMN) study conducted by (Leppänen und Lyytinen, 1997).[3] These

[2]Significantly, this is only a conference abstract.

[3]MMN is a change-related brain response evoked by auditory event-related potentials. They are measured using electroencephalography (EEG) methods.

authors found an attenuated MMN over the left, but not the right hemisphere of 6-month-old infants with at least one dyslexic parent and one more remote dyslexic relative compared with control infants with no such familial background. I will return to this point later, but first I will motivate the interest in the Hungarian language for setting up place-of-articulation contrasts.

3. Cues for place-of-articulation distinctions: Of invariants and cues

The choice of Hungarian is motivated by the presence of palatal consonants in the sound inventory. Their phonemic status as plosive or affricate has been a matter of debate though: For example, while the Hungarian palatals are treated as affricates in the Handbook of the IPA, other authors treat them as stops. A summary of this discussion will be given in a later chapter (2). A series of papers (Lahiri und Blumstein, 1984; Keating und Lahiri, 1993; Keating, 1993, 1988a) starting from a genuinely phonetic point of view - the "Acoustic Invariance"- approach - performed extensive acoustic analysis of these sounds. An overview, the major intentions and the research rationale of this approach are displayed in the following section.

3.1. Acoustic invariance: The analysis of burst spectra

As already mentioned in a separate section (2.1) in the introductory part, the theory of acoustic invariance (Blumstein und Stevens, 1979; Stevens und Blumstein, 1978) makes two major claims: The first is that there is acoustic invariance in the speech signal corresponding to the phonetic features of a language (Blumstein, 1986, p.178). In particular, this claim postulates invariance across speakers, phonetic features and languages. The second claim is that the perceptual system is sensitive to these invariant properties. One of the hopes of the theory of acoustic invariance was to provide means for some natural rules in phonology, and, more precisely, why certain assimilations are more likely to occur than others. One of the major domains of application has been the issue of palatal consonants (Blumstein, 1986; Keating und Lahiri, 1993). The analysis put forward in this series of papers achieved a formally simpler reanalysis of Chomsky's analysis of Slavic palatalization analysis by reviving the Jakobsonian featural descriptions, more explicitly the features compact/diffuse and acute/grave. The feature compact/diffuse, supposedly common to both vowels and consonants, distinguishes between open and closed vowels[4] and front and back (post-alveolar) consonants respectively. The name of the feature comes from its acoustic characterization. Diffuse sounds have energy spread widely (diffusely) across the spectrum, while in the case of compact sounds the energy is concentrated in the central area of the auditory spectrum and is therefore termed compact. The feature acute/grave is what Chomsky und Halle (1968) call anterior. It characterizes both vowels and consonants and distinguishes back vowels from front vowels and peripheral from central consonants. [+grave] sounds acoustically are characterized by low frequency and include back vowels and labial and velar consonants. Acute sounds will display higher frequencies and include front vowels, dental, alveolar and palatal consonants.

[4] With open vowels always being compact.

The intention of the “Acoustic-invariance”-approach concerning the palatalization example was to derive invariant cues from the acoustic signal - and additional palato- and linguograms - and resulted in the definition of “spectral templates” correlating with the above-mentioned phonetic features: Alveolar consonants are characterized by the “diffuse-rising” spectral template: “Diffuseness” means that there are higher amplitudes of higher frequencies. The opposite holds for bilabial stops which are characterized by the diffuse falling template, i.e. lower frequencies have lower amplitudes. The condition of acceptance for the diffuse-falling spectral template is that there has to be (a) a peak below 2400 Hz and (b) a second peak between 2400 Hz and 3600 Hz. There a no restrictions for peaks below 1200 Hz. Velar consonants are characterized by their conformity to the ’compact’ template. It is assumed that peaks which are closer together than 500Hz are perceptually interpreted as a single peak. Velar consonants exhibit such a behavior and show only one, but prominent spectral peak. The definition of prominence is given as follows: ’A peak is ’prominent’, if there are no other peaks nearby and if it is larger than adjacent peaks, so that it stands out, as it were, from the remainder of the spectrum. In this sense, the spectrum is compact.’ (Blumstein und Stevens, 1979, p. 1004ff). In order to view the above mentioned palatalization process as a true assimilation, Blumstein (1986) reports data comparing true palatal stops and velars in different vowel contexts. This involves the acoustic description of palatals before high front vowels in terms of burst spectra: The velar before [i,e,a] exhibits a broad mid-frequency peak in about the region between 2 and 3kHz, whereas before [o,u] there are two distinct peaks, one at about 900 and the other about 4200Hz. The palatal before [i,e,a] shows a high-frequency peak at about 3500Hz, and before [o,u] a mid-frequency-peak at in the region between 2 and 2.5 kHz. The conditions for interpreting these data as an assimilation are that the two contiguous segments must have similar acoustic properties and the original sound and the modified sound must share a number of acoustical properties. This is the case for [k] before front vowels and [c] before back vowels both exhibiting a broad mid-frequency peak. The form of invariance they postulate is called “dynamics relative invariance” (Lahiri und Blumstein, 1984, p. 182): “This form of invariance is dynamic in the sense that the invariant properties are determined by comparing the spectral properties of portions of the signal across the time domain [...]. It is relative in the sense that the invariant properties are derived on the basis of relative spectral changes in regions of high information”. Taken together, spectral characteristics of burst portions are identified as the invariants sought for in the acoustic signal to map onto the phonetic features, and, from the analysis of the phonological process, it becomes evident that transitions are not considered as likely candidates for acoustic invariants: The invariance approach takes a lot of its attraction from the fact that it avoids a classical problem associated with transitional cues for velar consonants, the existence of distinct loci for velars in the context of front versus back vowels. In the Acoustic-Invariance analysis of the phonological process mentioned, transitional cues are seen as context-dependent variations, and many of these variations will have minimal perceptual consequences: “speech perception maybe less “context-dependent” than has been generally assumed” (Blumstein, 1986, p. 179). Nevertheless, the constitution of transitional parameters as a (partial) solution to the invariance problem has also more recently been undertaken (Sussman et al., 1998).

3.2. Acoustic invariance: Locus equations

The characterization of the different places-of-articulations by the above-mentioned is not perfect though, as considerable variation has been observed already in the original work (Blumstein und Stevens, 1979): Apart from low-frequency peaks corresponding to subglottal resonances immediately following the burst, another problem was encountered:

> "The second characteristic not accounted for by the diffuse-rising template concerned a set of consonants whose spectra contained a substantial peak of energy in the vicinity of 1800 Hz - a peak whose amplitude exceeded the amplitude allowed by the alveolar pattern. This peak of energy corresponds to the starting frequency of the second formant for the following vowel - the so-called hub or locus[...]" (Blumstein und Stevens, 1979, p. 1005)

The diffuse-rising template allows the occurrence of the locus justified by the findings that (a) there are considerable interindividual differences in the frequency distributions and (b) that it concerns only 27% of the total alveolar productions - but locus is not seen as central for the characterisation of place-of-articulation. Instead of starting from burst characteristics, Sussman et al. (1998) take the opposite approach and attempt to build a theory of place-of-articulation perception on the basis of formant transitions: Since the pioneering work of Delattre et al. (1955) Delattre et. al. (1955), the question about formant transitions has been which parameters of the F2 transition the brain actually encodes, and how the different transitions characterizing a particular stop consonant in different vowel contexts are organized into a single percept by the auditory system. Operationally, Sussman et al.'s approach suggests to fit linear regression made to scatterplots of coordinates representing second formant frequencies at the midvowel on the x-axis and second formant transition onsets on the y-axis. Linear regression models are fitted separately for each consonantal place-of-articulation, but across different vowel contexts. Statistically, the F2 at onset is taken to be the criterion, the F2 at the steady portion of the vowel the predictor (Sussman et al., 1998, p. 246). This is the locus equation, which represents a second order feature. At a theoretical level, Sussman et al.'s approach draws on an analogy between human information processing and neuroethical evidence: "The neural unit that serves as the ubiquitous higher-order auditory processor appears to be the combination-sensitive neuron. Combination-sensitive neurons are specifically " tuned to coincidence (synchronization) of impulses from different neurons in the time, frequency and/or amplitude domains" (Suga (1994), p. 143, after Sussman et al. (1998), p. 242). For example, each orientation cry of the mustache bat consists of a long constant-frequency (CF) component followed by a short, faint, upward-sweeping FM (frequency-modulated)-component. The area processing the FM in the predominant harmonic is spatially isolated from the area processing the CF components, and specialized. Furthermore, the tonotopic representation in this area is orderly (Suga, 1977). Note that the relevance of burst information is not denied by this approach, rather it is assumed that, for processing speech sounds, the human auditory system may create many combination-sensitive neurons and functional subdivisions where the relationship between formants - associated with CF components -, transitions - associated FM-components - and parameters characterizing "fills" (Suga, 1996) - associated with burst or silence portions of the acoustic signal - are mapped. Rather, the preference for formant-associated information stems from what has been called the tinkering metaphor: "Engineers design machine-based recognition systems, and evolution designed the human brain: "natural selection does not work as an engineer works. It works like a tinkerer" [...]. Human speech perception is the late-comer with respect to sound processing. It was not designed de novo to handle overlapped speech sounds. What worked so perfectly in ancestral forms was not completely adequate for the task at hand. The computational mechanisms that were evolutionarily conserved had to be tinkered with as these new signal forms necessitated altered combinatorial algorithms using already functioning processors" (Jacob (1977), p. 1163 , after Sussman et al. (1998), p. 287f.) To summarize, the last two paragraphs sketched two computational descriptions of consonant-perception, based on different and - theoretically - non-overlapping portions of the acoustic signal. While the Acoustic Invariance Approach has a stronger focus on modeling phonological processes, Sussman et. al.'s approach makes explicit reference to neural substrates underlying place-of-articulation perception. This point will be resumed later when summarizing the motivation of own experimental efforts.

4. Experimental considerations

4.1. The choice of the neutral vowel

Among many others, Fischer-Jørgensen (1972) and Smits et al. (1996a) demonstrated that the perception of place of articulation in stop consonants depends on both burst and transition cues. After Smits et al. (1996a), the relative importance of burst- and transition cues depends on voicing, place of articulation and vowel context: So first, with respect to cue-trading, by taking stimuli from the same voicing category (i.e. voiced stops) and in the same vocalic context (stop + schwa syllables) one can fix trading relations between burst and transition cues. Second, the schwa has been assigned a prominent developmental role: The first vowel productions of infants are central vowels resulting from an open and neutral shaping of the vocal tract, more peripheral vowel qualities are acquired later, i.e. developmentally, the vowel space expands toward the periphery (Kent und Murray, 1982). Related to this line of reasoning, the central vowel plays a structuring role in the development of consonant-place-perception. Further, the central vowel might also have been a catalyzer in phylogenetic development (Carré et al., 2002). In the neutral vocoïd context, both rising F2 and F3 transitions represent bilabials, both falling F2 and F3 represent coronals and falling F2 and rising F3 represent velars. Place boundaries are then related to natural psychoacoustic boundaries in the schwa context, i.e. boundaries are related to flat transitions as the latter constitute the limit between rising and falling frequency movements. It is then possible that natural boundaries were used as primitives for developing place articulation distinctions in speech communication and that the neutral vocoïd context acted as a reference in this process. Related to this line of reasoning, the central vocoïd plays a structuring role in consonant-place-perception for both phylogenetic and ontogenetic reasons. As became clear in the previous paragraph, the experimental approach taken has to take into consideration developmental issues. In order to further clarify the motivation for the design of stimuli, we will report on some selected experimental findings: First, language-specific shaping of the perceptual space may occur earlier for vowels than for consonants: For instance, Kuhl; Williams; Lacerda; Stevens und Lindblom (1992) report that Swedish-learning infants showed the "magnet effect" at the age of six months when tested on the native Swedish vowels, but not for the non-native English vowels, and, vice versa, English-learning 6-month-olds showed the magnet effect only for the English vowels. This is comparatively early with regard to findings on consonantal development as for instance those made by Werker und Tees (1984) on non-native Hindi and Nthlakampx contrasts. Second, manner distinctions appear to be acquired later than place distinctions as stated by Goldstein und Fowler (2003): Acoustic data of six children aged between 1.1 and 1.9 years taken from the CHILDES database (Bernstein-Ratner, 1984, after Goldstein; Byrd und Saltzman, 2006), were presented to adult raters. The task consisted in classifying the initial consonants of words with known adult targets. Results revealed that raters were better at the classification of the constriction organs, - i.e. the place-of-articulation like tongue tip, lips, tongue dorsum - than at the classification of within-organ differentiation (Constriction degree). - i.e. manner distinctions like stop, fricative, glide (Studdert-Kennedy, 2002). Note that this study was based on infant speech production data, in contrast to the present study aiming at revealing evidence about the original "blueprint" of place-of-articulation-*perception*. Anyway, it seemed advisable to take precautions against possible confounds stemming from a potential "staging" in the acquisition of phonological categories. Further credibility is lent to this argumentation by findings from cue trading studies showing that infants comparatively assign greater importance to dynamic cues - transitions - than adults in syllable-internal formant transitions (Nittrouer, 1992).[5]

4.2. Design of the continua

The assumption concerning the meaningfulness of the neutral vowel for place-of-articulation perception does not exclusively stem from methodological reasoning - neutralization of trading relations- or acquisitional considerations, but has some supporting empirical evidence in adult perception: Carré et al. (2002) made the observation that labeling boundaries in the F2 onset - F2 offset plane for the voicing distinction converge to flat transitions for neutral vowel contexts. The design for place of articulation continua is inspired by an analogy: As mentioned, place boundaries are considered to be related to natural psychoacoustic boundaries in the schwa context in this work, i.e. boundaries are related to flat transitions as the latter constitute the limit between rising and falling (formant) frequency movements, which is in line with the Sussman et al. (1998) hypothesis of different neural processing strategies for steady state and frequency modulated signal components. Place boundaries are then related to natural psychoacoustic boundaries in the schwa context, i.e. boundaries are related to flat transitions as the latter constitute the limit between rising and falling frequency movements. This strategy was underlying the design of the continuum (see figure 2.1). A heuristic starting point was the transition-based description of the place features afforded by the Distinctive Region Model (DRM) of place production (Carré und Mrayati, 1991). The DRM is organized around the neutral vowel (schwa) as a central reference. In the neutral vowel context, place boundaries tend to correspond to flat F2-F3 transitions, the categories being characterized by rising vs. falling transitions. The four possible combinations between F2 and F3 transitions directions generate four Distinctive Regions on the anterior- posterior direction with the following specifications: F2-F3 both rising (R8), F2 rising- F3 falling (R7), F2-F3 both falling (R6), F2 rising-F3 falling (R5). Although there are

[5] This claim recently has been questioned Mayo und Turk (2005).

no clearcut correspondences between Distinctive Regions and articulatory descriptions of place categories, the R8, R7, R6, and R5 regions are usually ascribed to labial, dental, alveolar and velar places of articulation, in that order (see figure 2.1). However, for the purpose of relating the DRM to the four Hungarian stop categories, our working hypothesis in the present study was that alveolar stops would occupy the Distinctive Region with rising F2-falling F3 transitions (R7 in the DRM) and that palatal stops would occupy the Distinctive Region with rising F2- rising F3 transitions (R6 in the DRM). This share-out between place categories and Distinctive Regions is thus different from the usual conception which ascribes dentals to R7 and alveolars to R6. The reasoning behind using such an "artificial approach" was in principle already justified in the introduction, when discussing the place-dependency of the relative weightings of different cues (see section 2.3) in former speech perception studies. For the present purpuse it is not relevant though whether the explanation for place differences in cue-trading is the perceptual apparatus making active adjustments in the assignment of relative weights for different aspects of the signal or whether "it is articulation" as postulated by Smits (1996).

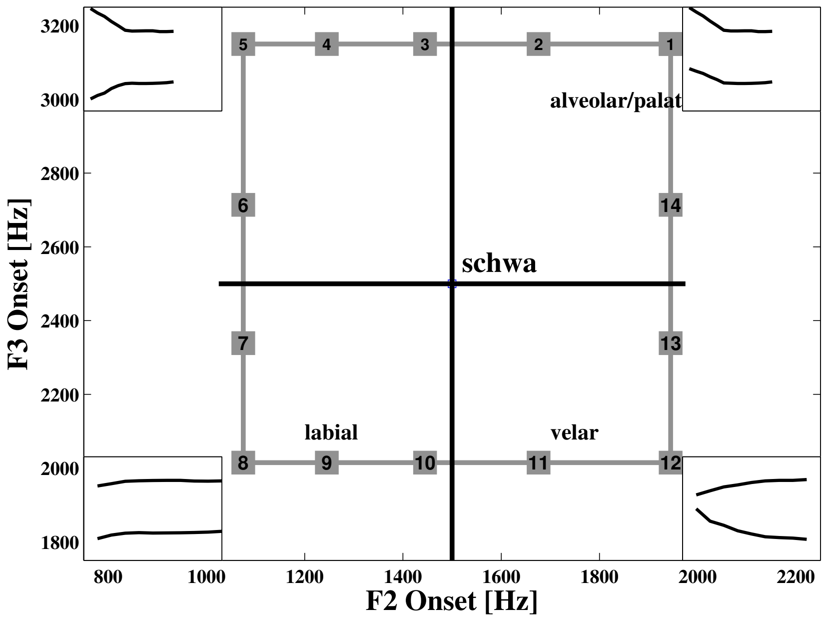


Figure 2.1.: Locations of potential place categories in the F2-F3 transition onset space according to the Distinctive Region Model.

4.3. A previous study

As mentioned, the partitioning of this acoustic space into the three phonologically relevant stop place categories present in the French language was previously examined with perceptual data (Bogliotti, 2005).[6] Results showed that the velar region occupied roughly the lower right quadrant, corresponding to falling F2- rising F3 transitions. The two other categories, labials and

[6]Experimental details will be given in the method section (section 5).

coronals, share the remainder of the space with a perceptual boundary located inside the empty rising F2 - falling F3 region, though not in the middle of that region. Results also indicated that non-phonological boundaries remained discriminable for the French adult subjects, thereby showing that even when natural boundaries are located inside phonological categories, they can still affect consonant discrimination. This result has been interpreted as evidence for a residual perceptive sensitivity for natural boundaries. The perspective taken here is a crosslinguistic one though: How will this perceptual space be divided given an additional category?

4.4. The hypothesis space

Summarizing the above, the study presented here aims at giving answers to questions from two distinct scientific disciplines: *First*, the question about the phonetic substance in the shaping of phonological systems: As mentioned, in technical terms a "deleted cue" experiment measures the necessity of the deleted cues and the sufficiency of the others (Smits et al., 1996a). The reason for the choice of a deleted-cue paradigm due to factors related to category formation and methodology was motivated already. To repeat, from a linguistic point of view the question asked is whether it is possible to have four place-of-articulation categories in a system of paradigmatic contrasts if the place-of-articulation distinctions were to be made on transitional information alone, or whether an upper limit of perceptual identifiability / discriminability is reached.

Second, the question concerns the potential remainder of universal phonetics in adult speech perception. This question is related to the models of speech category formation as sketched in the introduction: Here the Perceptual Magnet approach makes distinctive predictions insofar as the remodeling of the perceptual space leads to a deactivation of natural phonetic boundaries. This results in coherence with the empirical predictions made by the Categorical Perception Model (see Pisoni und Tash, 1974), i.e. a primitive relationship between discrimination and identification performance with no spurious discrimination peaks. With respect to labeling performance, the magnet approach would predict a high degree of modifiability of the category boundaries, i.e. the implementation of the additional palatal place-of-articulation would bear no complications, but rather an additional robust category. In this regard, this approach makes distinctive predictions as compared to structure-adding approaches emphasizing the role of innate perceptual boundaries, where the local increase in category density would lead to a conflict between universal phonetics and the stop inventory of Hungarian, as the quadrant with both falling transitions should lead to perceptual conflict due to "over-population".

5. Method

5.1. Participants

The 22 participants for the Hungarian subset were (a) participants of an undergraduate linguistics course or (b) volunteers contacted via a mailing list. Apart from their first language, Hungarian, all of them were familiar with at least one of the languages German, French or English. Most of them were participants of undergraduate exchange programs. They were between 18 and 53 years old with no reading or hearing impairment reported. The French dataset was similar in age structure: Subjects' age ranging between 17 and 59 years. Likewise, there were no known auditory problems.

5.2. Stimuli

23 stimuli CV sequences were generated with a hybrid formant synthesizer modeling the formants trajectories in series and the noise portions in parallel.[7] F1-F2-F3 transitions ended at 500, 1500 and 2500 Hz respectively after a 27 ms transition. The VOT was set to -95 ms and the stable vocalic portion had a duration of 154 ms. The stimuli differed as to the onset of F2 and F3 transition. 14 stimuli were generated by separate modification of the F2 and F3 onsets along a "phonetic" continuum, normal to the locations of the natural boundaries - corresponding to either flat F2 or F3 transitions - as shown in Figure 2.1. The same amount of stimuli was generated with the same basic data but an additional, constant burst-like signal portion. Successive stimuli were 1 Bark apart on both continua. The present paper only deals with the data of the "phonetic" continuum.

5.3. Procedure

Both continua were presented to each of the participants. The continua with and without burst were presented in alternating order resulting in a between-subject factor (order of presentation) which was used for control purposes. Hungarian participants were told that they would hear one of the four sounds "b","d", "gy" or "g" and were instructed to report which of the four sounds they had heard. They were told that the sounds not necessarily were presented with equal frequency, and to judge each sound separately. For the French participants, the procedure was alike except that for them only three response alternatives b, d and g were available.

5.4. Statistical models

Territorial maps and logistic modeling

Nearey (1997) shows how a topological representation of categories in stimulus space can be generated from the coefficients of the logistic regression models. The approach used here deviates due to reasons to be discussed from the standard Logistic Regression (LR) models. The data here were fitted by Nonlinear Regression with a hierarchical model in which the effect of F2 was nested in the effect of F3 and the latter was nested in the effects of Residual Cues (i.e. the acoustic cues for place which were constant in the stimuli). This model is instantiated by Equations 2.1 and 2.2. Labeling responses depend on a Logistic Regression equation including Residual Cues, a nested LR equation including F3, itself including a nested LR equation including F2. Each LR equation included different variables representing the effects of Burst and Language.

$$\Phi(\gamma(cues)) = \frac{1}{1+e^{-\gamma(cues)}} \tag{2.1}$$

In equation 2.1, Φ is the Logistic function and γ is a linear function. The model used here is composed of three nested Logistic functions, as specified in equation 2.2. Here, R stands

[7] Developed by René Carré.

for “Residual Cues”. In other words, this constitutes a hierarchical logistic model for speech perception[8] “cascading” the contributions of the different transitional cues.

$$\text{Labeling Response} = \Phi(\gamma(R, \Phi(\gamma(F3, \Phi(\gamma(F2)))))) \tag{2.2}$$

6. Results

6.1. Labeling responses

The labelling curves for the stimuli with or without burst in French and Hungarian are presented in figure 2.2. Although there are obvious differences between the labelling curves for the stimuli with vs. without burst, the location of the boundaries (i.e. the stimuli collecting an equal number of responses for two adjacent categories) are only marginally affected. In French, there are three boundaries corresponding to (from left to right in figure 2.2) the alveolar/labial, the labial/velar and the velar/alveolar distinctions. Interestingly, there is a secondary peak of velar responses around the alveolar/labial boundary, mainly for the stimuli without burst. In Hungarian, there are four boundaries corresponding to (from right to left in figure 2.2) the palatal/alveolar; alveolar/labial, the labial/velar and the velar/alveolar distinctions. The distinctions between palatals, alveolars and velars are not very clearcut (figure 2.2). However, the Hungarian palatal and alveolar functions, when taken together, correspond fairly well to the French alveolar function. Given that the alveolar responses additionally were fairly infrequent in the results of the Hungarian dataset, it seemed desirable to lump the alveolar and palatal responses together in order to test whether the underlying categorical space is identical between Hungarian and French, given that the judgements are based on transitional information alone.

6.2. Territorial maps of labeling responses

The data were fitted with Non Linear Regressions (NLR) run on a hierarchical model (see Method, Equations 2.1 and 2.2). A separate NLR model was run for each category boundary of the pooled data, i.e. labial/velar, velar/alveolar-palatal and alveolar-palatal/labial boundaries. NLR was used for testing the effect of language on place identification as well as specific hypotheses on the location of the place boundaries in the F2-F3 onset transition space. The expectation was that the place contrasts which are common to both languages would display the same perceptual boundaries.

For the *labial/ velar* contrast, the model only included an F2 component nested in a Residual cues component. The effect of F3 and its interactions with Burst and Language were not significant. There were 7 significant parameters. Burst and Language biases were not significant. The effects of F2 (bias and slope), the Burst x F2, Language x F2 (all $p<.001$) and Burst x language x F2 ($p<.05$) interactions were significant. The labial/velar boundary corresponds to an almost flat

[8] With a convergence with the conceptual coupling concept postulating perceptual interdependencies in the processing of different features. Accordingly, the model includes interdependencies in the perception of the different acoustic cues which convey these features. However, rather than being symmetrical, couplings are hierarchical in equation 2.1, a working assumption for the sake of parsimony: A symmetrical model would indeed require feedback loops in the processing of the different cues.

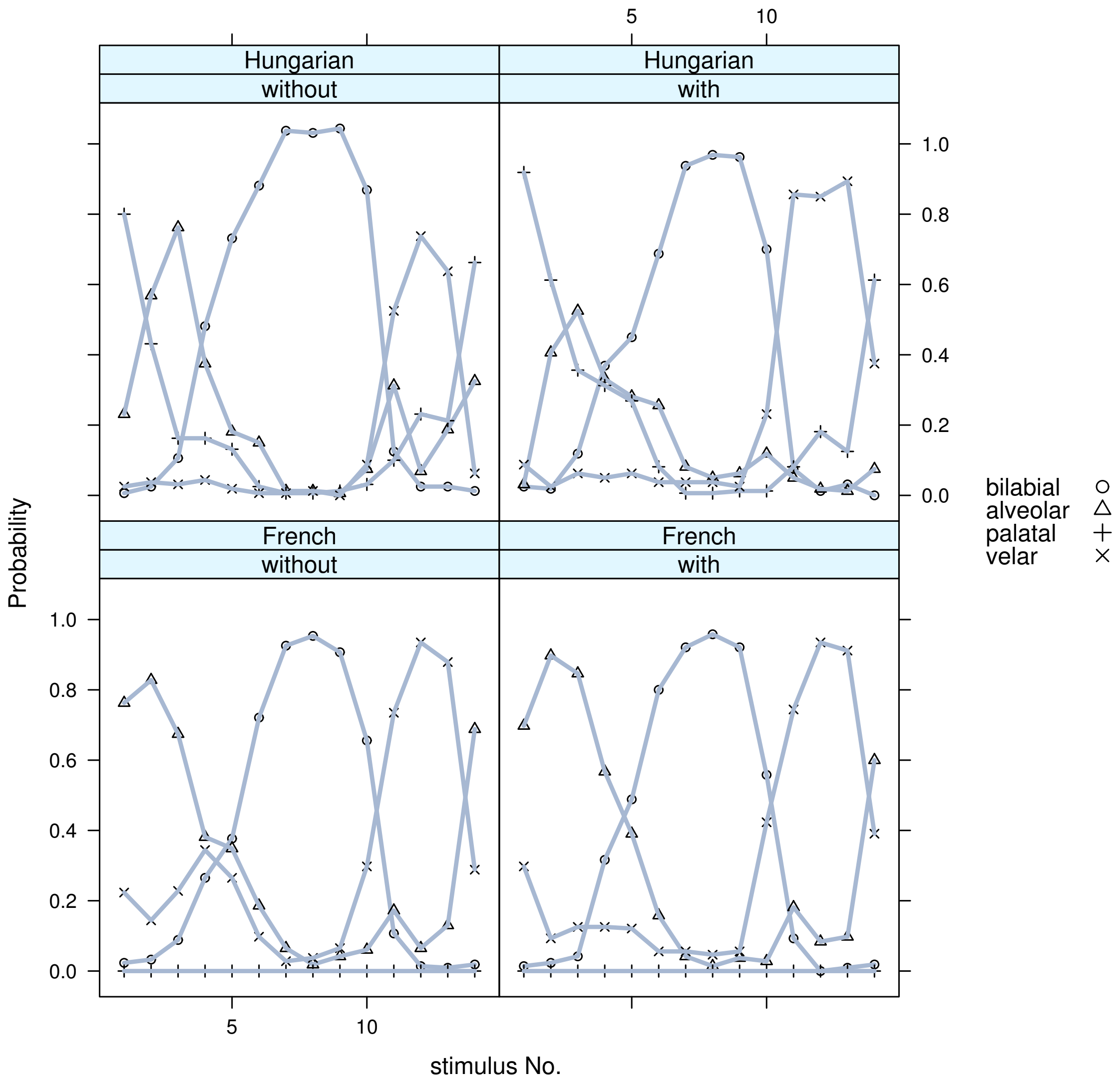


Figure 2.2.: Labeling curves for the stimuli with and without burst in Hungarian and French.

F2 transition in both languages, both for the stimuli with and without bursts (see also figure 2.4). For the *velar/alveolar-palatal* contrast, the model included a F3 component nested in a Residual cues component. The effect of F2 (bias and slope), and its interactions with Burst and Language were not significant. There were 8 significant parameters. The effect of the Residual cues, Burst bias, Language bias and Burst x Language bias were significant (all p<.001). The effects of F3 (p<.001) and the Burst x F3 interaction (p<.01) were also significant. The velar/alveolar-palatal boundary corresponds to an almost flat F3 transition for the stimuli without burst and a slightly falling F3 transition for the stimuli with burst (see table 2.1).

		without burst		with burst	
		Hungarian	French	Hungarian	French
	Observed	11.5	11.3	11.3	11.1
lab./vel.	NLR	11.6	11.1	11.3	11.0
	CI limits	11.2-12.2	10.7-11.6	10.9-11.9	10.7-11.5
.	Observed	14.3	14.7	14.8	14.8
vel./alv.-pal	F3 onset	14.3	14.7	14.8	14.9
	CI limits	13.8-14.6	14.6-14.9	14.7-14.9	14.9-15.0
alv.-pal./lab.	Observed	1.0	0.7	0.5	0.7
F3 ext./F2 ext.	NLR	0.9	0.9	0.7	0.7
	CI limits	0.5-1.3	0.5-1.3	0.4-1.0	0.4-1.0

Table 2.1.: Values of formant transitions at the perceptual boundary for the place contrasts common to both languages, for each burst condition (without vs. with), and for each language (Hungarian vs. French). Each data cell gives the observed values, NLR estimates and 95% CI limits. For the labial/velar contrast, the boundary values are fairly close to the flat F2 boundary transition (11.2 Bark, 1500 Hz F2) in both languages, both for the stimuli with and without bursts.

For the *alveolar-palatal/labial* contrast, the model included an F2 component nested in an F3 component The effects of the Residual cues as well as Burst and Language biases were not significant. There were 6 significant parameters. The effects of F2 and F3 (bias and slope), as well as the Burst x F2 and Burst x F3 interactions were significant (all p<.001). The results for F2 and F3 transition onset values are presented in table 2.1, per language and burst condition. A rising F2 transition is compensated by a falling F3 transition in both languages and both burst conditions, indicating that the alveolar-palatal/labial boundary corresponds to a trade-off between a rising F2 and a falling F3 transition.

Until so far, the usage of the computationally more expensive NLR models has not been justified. To catch up on this, the performances of the NLR models were compared to those of the simple Logistic Regressions with the same number of parameters. The percentage of explained variance amounted to 63.4% with NLR vs. 61.8% with LR for the labial/velar contrast, to 40% with NLR vs. 38% with LR for the velar/alveolar-palatal contrast, to 64.1% with NLR vs. 60.4% with LR for the alveolar-palatal/labial contrast. The NLR models fitted the data better than simple Logistic Regressions although the quantitative differences are fairly small overall. However, these differences are far from being negligible because the differences between expected and observed boundaries are much larger with the LR vs. NLR models. This is illustrated with two different examples in Figure 2.3. Especially in the left panel, it is evident already by visual inspection that the ordinary Logistic Regression procedure models the category boundary defined as the 50% crossing of identification probability substantially differently, i.e. between stimulus 12 and 13, whereas observed data and the NLR suggest a category boundary closer to stimulus 13. The situation is similar in the right panel. The low gain in overall percentage by the usage of the NLR can also be illustrated by the overall higher frequencies of the bilabial responses. Responses clearly in the bilabial region of the continuum will be as easily captured by the standard Logistic Regression model, whereas this statistical approach will break down close to category boundaries. Therefore the decision for the formally more complex NLR is justified. Territorial

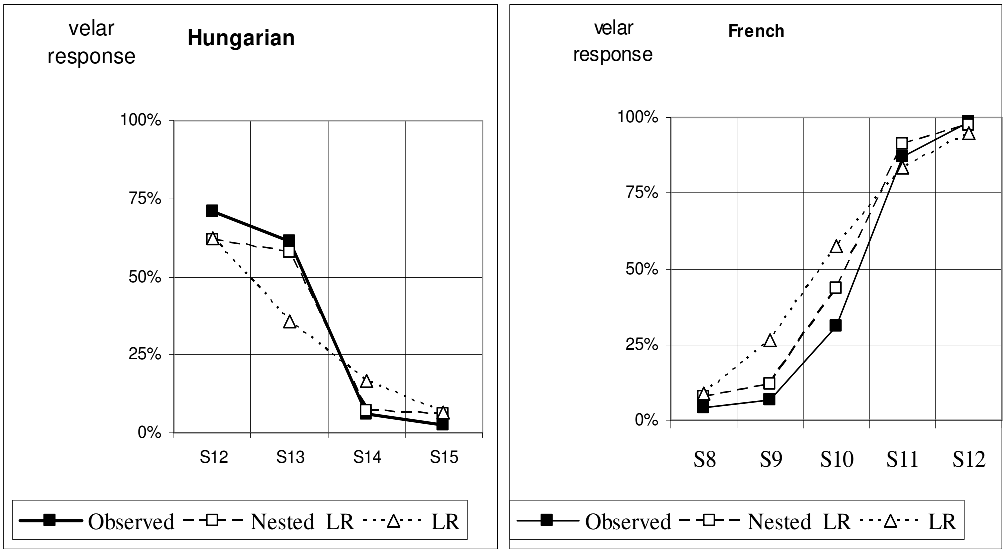


Figure 2.3.: Examples of the relative failure of the Logistic Regression (LR) vs. Non Linear Regression (NLR) for assessing perceptual boundaries (50% response points). Observed and expected response scores for the labial/velar contrast in French (right) and for the alveolar-palatal/velar contrast in Hungarian (left). In both cases the assessment of the boundary is substantially improved by the use of NLR in comparison with LR.

maps of the place categories in the F2-F3 onset frequencies are presented in Figure 2.4. These maps were obtained by calculating the boundaries between categories from the outputs of the Non-Linear Regressions (equation 2.1). For both the stimuli with and without burst, the velar region corresponds to the lower right quadrant with boundaries corresponding to fairly flat F2 and F3 transitions (see table 2.1 for details). The labial/alveolar-palatal boundary corresponds to the tradeoff between a rising F2 and a falling F3 transition. There is some tendency for the velar region to be narrower in Hungarian but differences between languages are fairly small.

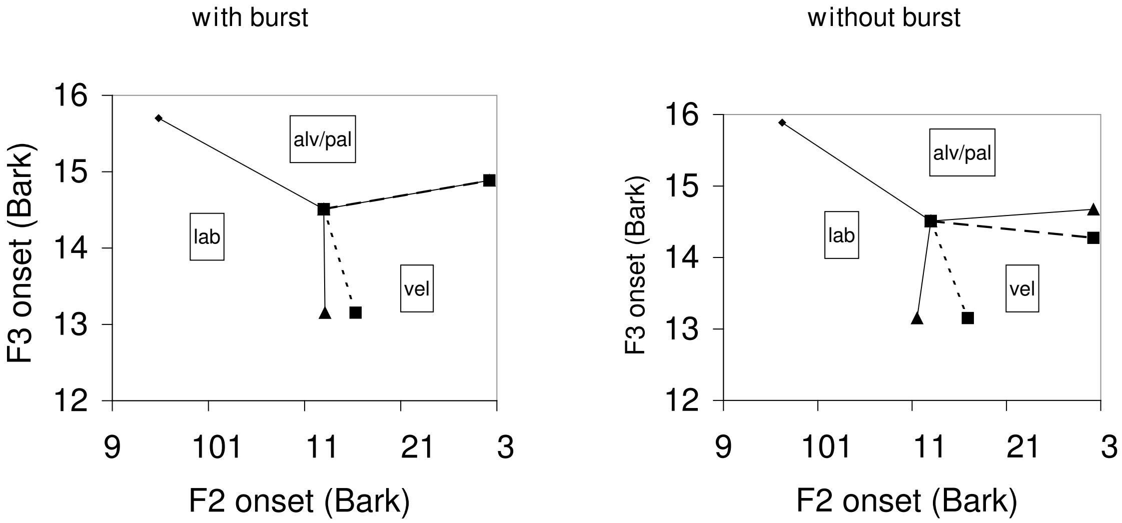


Figure 2.4.: NLR estimates of the territorial maps per burst condition with both French boundaries (plain lines) and Hungarian boundaries (dotted lines). The labial/alveolar-palatal boundaries of the two languages overlap.

6.3. Category boundary effect

An additional question which arises is whether identification and discrimination data collected for *Hungarian* conform to the patterns which have been described in the framework of Categorical Perception, i.e. whether the data exhibit (a) discrimination maxima at the category boundaries and (b) discrimination peaks which are predictable by the identification function. For this purpose, we carried out an analysis of variance with percent correct responses as the dependent variable, burst (two levels, factor BURST), predicted versus observed (two levels, factor PRED) and stimulus number (14 levels, factor STM) as within- subject factors. The order of presentation (2 levels, ORDER) was treated as a between-subject factor. Correct observed discrimination response scores were calculated for adjacent stimuli on each continuum as the mean correct response to both same pairs and different pairs. The predicted discrimination scores were calculated by a formula adapted from Pollack und Pisoni (1971). Discrimination results are shown in Figure 2.5.

Mauchly tests of sphericity yielded significant values for STM and the STM x BURST interactions. Therefore, degrees of freedom were Huynh-Feldt-corrected, and corrected values of significance and degrees of freedom will be reported. The significant effects and the results for the between-subject factor ORDER are shown in table 2.2.

The interaction between PRED and STM gave a significant ($p<.05$) result. This indicates a failure of the prediction for some stimuli. Visual inspection of observed versus predicted

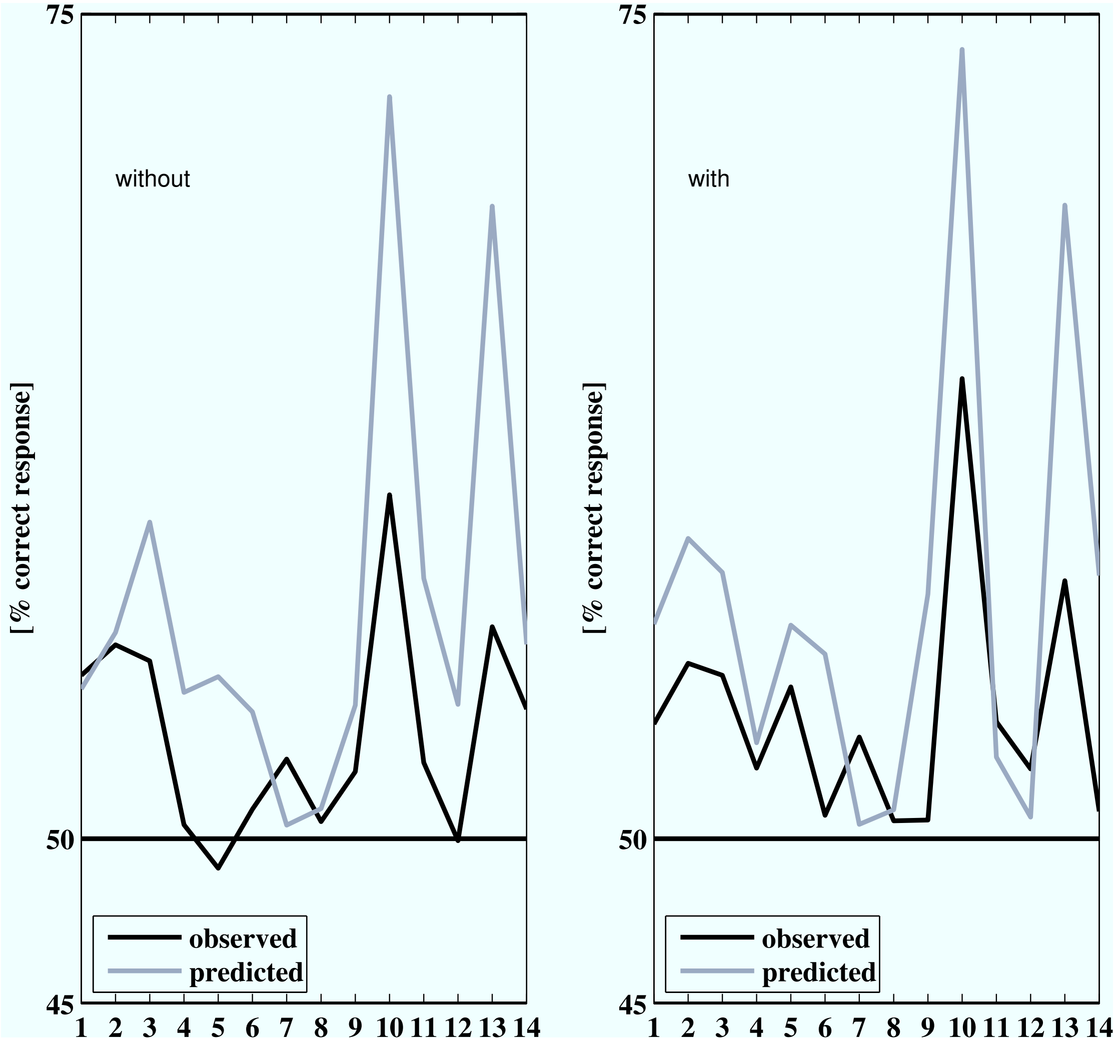


Figure 2.5.: Mean observed and predicted discrimination scores for the stimuli without (left panel) and with burst (right panel).

	Phonetic		
	df	**F**	**p**
Order	1	.003	.960
Pred	1.0	22.0	<.001**
Stm	6.30	8.50	<.001**
Pred x Stm	11.30	1.86	<.05*

Table 2.2.: Repeated measurement ANOVA of predicted and observed scores.

discrimination scores (Figure 2.5) showed that this failure is local to the stimuli with high F3, i.e. the boundaries between palatal and alveolar and alveolar and bilabial are selectively affected. Observed discrimination peaks fall closer to the natural boundaries, i.e. to flat transitions lines in the F2-F3 onset space, than did the peaks expected from the labeling data. The same kind of analysis was carried out with the observed discrimination scores only. The factors are identical to those in the previous analysis except that there is no within-subject factor for PRED. The significant results and the results for the ORDER between-subject factor are reported in table 2.3. Again, some tests of sphericity were significant: STM, STM x PLOS, (1%-level). For the analysis only including observed discrimination scores, i.e. dropping the PRED within-subject factor, generally fewer effects reach the significance level: Only the STM main effect reached the level of significance.

	Phonetic		
	df	**F**	**p**
Order	1	.002	.964
Stm	11	3.94	.000

Table 2.3.: Repeated measurement ANOVA of predicted scores.

7. Discussion

The present results show that transitions are used in much the same way in both Hungarian and French: Strikingly, the contrasts which are common in both languages use almost the same perceptual boundaries, especially for the stimuli with burst. These common boundaries are not selected at random but correspond to qualitative changes in the direction of frequency transitions. Additionally, traditional patterns from Categorical Perception break down as evidenced by the significant interaction between the random factors coding stimulus number and whether data were predicted from identification scores or observed experimentally. These results can be discussed with respect to several not fully related aspects:

7.1. Acquisitional aspects

The results reported a comparison of the perceptual F2/F3 space as partitioned by a language with four place categories in contrast to a language which has only three categories like French. At the moment, we have to restrict ourselves to a qualitative comparison of the territorial maps as shown in Figure 2.4: One possible outcome of the local increase in category density would be a substantially increased region covered by palatal and alveolar places together in comparison to the region covered by the alveolar alone in a three-category language. However, this is not the case: The category boundary between alveolar/palatal and bilabial is at a location similar to the French sample used in Bogliotti (2005). If the coronal space should be considered as enlarged at all, this occurs at the expense of a smaller velar region. We have described theories of speech category development in the introductory section. The approach which postulates the greatest independence of adult percepts from natural boundaries and thus the greatest modifiability of category boundaries is the structure-changing - in the terminology of Phillips (2001) - Perceptual Magnet. The data we have shown in this study are challenging with respect to this claim, if the Magnet approach were to be adopted in a similar fashion for consonants.[9] Phonemic boundaries appear to be remarkably constant over languages, with the exception of an intrusive alveolar-palatal boundary for Hungarian located inside the French coronal region. This appears to go together better with approaches like the strong nativist stances like as advocated by e.g. Werker und Tees (1984) or neo-gestaltist approaches like Serniclaes (1987).

This was also evidenced by the logistic modelling undertaken, which lumped the palatal and alveolar place categories together, resulting in a convergence between the territorial maps for both languages. Interestingly, the examination of the discrimination results reveals four discrimination peaks for four place of articulation categories in the continuum containing the residual burst information (see figure 2.5, right panel) and the complete breakdown of the CP category boundary effect in the continuum without burst information (see figure 2.5, left panel).[10] The decision between theories of speech category formation is not the main focus here, therefore the next section will make a switch towards the discussion of transitions vs. bursts as the main vectors of speech perception.

7.2. Transitions vs. burst as vectors of place perception

The poor representation of the alveolar/palatal distinction between Hungarian consonants in the F2-F3 transition space suggests that broader couplings are necessary for stabilising the alveolar percepts. Other features, among which those provided by the burst spectrum, might be necessary for the addition of the palatal to the three principal places of articulation. To reveal the contribution of the burst-related features one has to use stimuli generated by factorial variation of burst and transitions. There have been several attempts in the past for separating the contributions of burst and transitions to place perception in stop consonants and most of these studies point to the functional equivalence of the two cues across phonetic contexts (e.g. Dorman et al., 1977). However, these results were collected in languages with only three place categories which make it rather difficult for evidencing autonomous contributions of the two kinds of cues because both contributed to all the three possible contrasts. Things might turn out differently in a four-category language like Hungarian in which the different place contrasts might rely on different cues: as shown in the present results, transitions are sufficient as long as there is no alveolar/palatal contrast present. It is then possible that the perception of this contrast might rely on burst properties which are independent of the onset frequencies of the formant transitions. Future experiments with stimuli generated by factorial variation of burst and transitions should allow to clarify this point. The fact that the category boundaries remain relatively stable between both languages under consideration while adding an additional place-of-articulation seems to suggest that there is an upper limit of place categories implementable on transitional cues alone, which is of potential interest for the architecture of phonological systems. Without an

[9] Almost all of the work within the Perceptual Magnet approach was achieved by examining vowels, not consonants. The abstract by Davis und Kuhl (1994) is a notable exception. The fact that consonants are a better testing ground for classical CP was already mentioned in the introductory chapter (see section 2.3).

[10] This effect was not further explored here, because at the moment there is no hypothesis at hand.

in-depth treatment in this place, it has been the declared aim of this study from the onset to determine whether four places-of-articulation were robustly implemented on transitional cues alone, a result which would have questioned the validity of attempts to define systemic, perceptually-oriented low-dimensional markedness scales of the kind proposed in Padgett und Zygis (2003) and Žygis (2003): Essential perceptual parameters would have arbitrarily been omitted from the conception of such a scale - by the authors' exclusive focus on nontransitional information - if the perception of four places-of-articulation could be determined by transitions alone. However, this was not the case, and these authors at least do not seem to be completely on the wrong track. This point will be resumed in the next section.

7.3. Specifiers, invariants, cues

The introduction promised to shift the discussion in the direction of the relationship between distinctive features and their correlates and to review Nearey's "double-weak" theory (Nearey, 1995), which offers an informative taxonomy for the characterization of relationships between articulatory, symbolic and auditory/perceptual entities. At the time of its development proponents of these approaches became increasingly frustrated by the search for invariants in the acoustic signal pointing to features. At the risk of repeating myself, locus-theorists as well as proponents of the Acoustic-Invariance-theory had to make concessions: Earlier work on the locus-theory had focused on low-vowel contexts, later work on additional vowel contexts discovered the two-loci problems for velar consonants and had to draw on burst information which was not the initial intention. Similarly, the Acoustic Invariance approach had to withdraw from earlier, more radical formulations aiming at static frequency characteristics *specifying* the place-of-articulation and therefore *distinctive features*. It is classified by Nearey as an instance of a *double-strong theory*:

> "One important approach to this problem can be referred to as a double-strong theory of phonetic specification. Here the term strong is taken to imply a simple, robust, and transparent relation between physical and symbolic elements. Stevens and Blumstein's 1981 model is double-strong, in that it postulates (i) strong relations between symbols and gestures and (ii) strong relations between symbols and auditory properties." (Nearey, 1997, p. 3241)

Other approaches exhibit less symmetry: postulating strong relations between only one of the physical domains on one side and the symbolic level on the other side. An example for a theory making a strong link between articulation and the symbolic side is the "Motor Theory":

> "Strong-gestural approaches may be exemplified by the motor theory of Liberman und Mattingly (1985). This postulates a strong, transparent relationship between symbols and gestures. There is a more complex and unidirectional path that relates gestures to auditory properties . This reflects the complex, typically nonlinear mapping that psychomotor and biomechanical factors impose between natural units of the gestural domain and their acoustic consequences of the unidirectional mapping postulated [...]." (Nearey, 1997, p. 3241)

Without forgetting to mention that the inverse mapping from acoustics to gestures is not assumed to require analysis-by-synthesis as in Liberman und Mattingly (1985), but instead is said to involve "direct perception" of gestures by listeners, Nearey also subsumes Fowler (1989) under the "strong gesturalists". The other pole of "strong auditory" theories is represented by Diehl und Kluender (1989), Kingston und Diehl (1994) and Kingston und Diehl (1995), which hold approximately the opposite position of strong gesturalists:

> "Relations are assumed to be strong between symbols and auditory properties but only weak and indirect between symbols and gestures." (Nearey, 1997, p. 3242)

Nearey summarizes criticism against both classes of theories on empirical grounds. These are not repeated here, it rather is important that he posits the double-weak standpoint as a *pragmatic* choice given that the claims made by strong theories are too strong and not tenable. This to some extent resembles the terminology applied by Fowler (1994) who defined *invariants* in a strong sense as *specifiers* uniquely determining a property. Understood in a weaker sense, invariants are defined as *cues*, which provide information for stop place, but not sufficient information to *specify* stop place. The results of the present experiment of course render a strong interpretation of transitional parameters in terms of specifiers unlikely, and currently no comparable data on complementary, acoustically aperiodic cues are available to me.

In the preceding section, efforts to construct low-dimensional markedness scales motivated by perceptual mechanisms related to the perceptual relevance of spectral moments were mentioned (e.g. Padgett und Zygis, 2003; Żygis, 2003), and it could be speculated that such efforts would be rejected by proponents of a double-weak theory like Nearey. This is not the case as evidenced by the following quote:

> Strong theories (and especially the double-strong theory) present scientifically appealing asymptotes that no researcher in phonetics can safely ignore. However, viewing such asymptotes from below, with a fuller appreciation of *what might be* and fewer postulates about what *must be*, may lead to a theory of phonological contrast that more closely resembles *what is* (Nearey, 1995, p.38).

But still, if one starts from a perspective emphasizing human information processing and how humans integrate information into a categorical percept rather than from the linguistic system reflected in terms of feature systems, research efforts have to deal with the complication that the perceptual apparatus uses all available information to arrive at a percept - after weighting and integrating these bits of information. In the present work, the statistical model attempted to build such trading and integration processes directly in the statistical approach (see section 5.4). The approach taken started with a simple hierarchical model as a parsimonious working hypothesis and did not allow for feedback loops and more complex architectures. It therefore bears some superficial similarity to simplified Artificial Neural Network architectures, which are often held up as metaphors for biological neural networks,[11] but the emphasis was put on parsimony and the paramorph testing of specific hypotheses about information-processing integration, in particular the idea of "hierarchical" processing, i.e. exploiting first the most informative structure

[11] See Shepherd (1990) for a critical review of the biological plausibility issue.

in the signal and in consecutive steps integrating it with other signal portions in order to arrive at a percept. The experimental work performed here was refined to strongly impoverished experimental manipulations. Further experimentation incorporating a more complex and broad cue structure could be revealing for the substantiation of processing strategies and cue precedences the perceiver is likely to apply. In principle hierarchical models as applied here can be comparatively evaluated in such respect, and it surely would be informative to do so.

Part II.

Experimental Part

3. A cross-linguistic study of dorsal obstruent articulation

This part of the current study is concerned with some in-detail analysis of articulatory patterns of the dorsal obstruents under consideration. Mainly, it is intended to present the results of the articulatory and acoustic analysis of two corpora designed for this purpose.

(1) The first corpus is concerned with the analysis of the German alternation between *ich*- and *ach*-Laut in the German fricative system and the corresponding patterns in the realisation of the velar stops.

(2) The second corpus deals with the analysis of palatal and velar stops in Hungarian.

The dorsal alternation is in some sense the common ground for both studies and designed to facilitate uncovering possible influences of the different language paradigms on categorical organisation. Although not the main topic here, a display of the most important phonological patterns seems in place. These descriptions are achieved separately for the German *ich*- and *ach*-Laut alternations (in section 1) and the palatal stops in Hungarian (in section 2). The relevant parts of the consonant inventories of both languages are displayed in appendix A. Further theoretical building blocks required are more severely inclined towards Articulatory Phonetics, like the data-driven taxonomy of palatal articulations elaborated in Recasens (1990) and succeeding work by the same author and colleagues (e.g. Recasens; Farnetani; Fontdevila und Pallares, 1993; Recasens, 1997, 2003; Recasens und Espinosa, 2006). This work is described in section 6. As this work has its roots in coproduction theory, Articulatory and Gestural Phonology, the concepts which were omitted in the introductory part (section 2.2) are also presented in this part (in section 5). Articulatory Phonology itself was developed in parallel with Feature Geometry. The parallelity between these approaches is depicted in a separate section devoted to Feature Geometry. Patricia Keating developed (i) a distinguished approach to coarticulation and (ii) representations for the palatal segments under considerations, both on feature-geometric grounds. Her work also deserved the treatment in a separate section (see section 4).

After this theoretical reworking and the method section (section 8), articulatory patterns are in a first step delivered separately for each language (sections 9 and 10), and in a second step crosslinguistic comparisons (section 11) for selected contrasts - or better where possible at all - are furnished. The final discussion concludes.

1. The phonology of the German dorsal fricative

As repeatedly mentioned, the objective of the current study are dorsal articulations. Apart from the velar stops, German has such material also in the fricative inventory: the dorsal fricatives ç and x. The following sections (1.1 to 1.3) will give an overview over some of the the most important phonological approaches which have dealt with these segments. This summary proceeds historically with a mere descript.

1.1. Structuralist and generativist analyses

According to Jessen (1988), "the behavior of dorsal fricatives ç and x is one of the notorious problems in German phonology." He also remarks that the most essential facts were already introduced in Paul (1916) and the issue has been a matter of passionate debate since Bloomfield (1930). The principled patterning is not a matter of debate though:[1]

(3.1) (a) velar /x/ after a back vowel

Buch [uːx]	Bucht [ʊx]	Buche [uːx]
hoch [oːx]	Masochist [ɔx]	Bochum [oːx]
nach [ɑːx]	sachte [ax]	Lache [ɑːx]

(b) velar /x/ after [aʊ̯]
Bauch, tauchen [aʊ̯x]

(3.2) (a) palatal /ç/ after a front vowel

riechen [iːç]	ich [ɪç]
Bücher [yːç]	nüchtern [ʏç]
-	Blech [ɛç]
-	Löcher, möchte [œç]

(b) palatal /ç/ after [aɪ̯], [ɔʏ̯]
Eiche, Beichte [aɪ̯ç]
heucheln, Bäuche [ɔʏ̯ç]

(c) palatal /ç/ after tautomorphemic consonant
Milch, Dolch [lç]
Mönch, manche [nç]
durch [ʁ̯ç]

(d) /ç,ʃ,k/ at the beginning of a prosodic word
China [çiːna,ʃiːna,kiːna]
Chemie [çemiː,ʃemiː,kemiː]

Within structuralist tradition, both alternatives, the analyzes as separate phonemes (/ç/,/x/) or as allophones of the same phoneme (/ç,x/) have been pursued as workable alternatives. For example, Moulton (1962) proposed an analysis as separate phonemes remarking that the complementary is not fully consistent at morpheme boundaries, more precisely the suffix *-chen* is

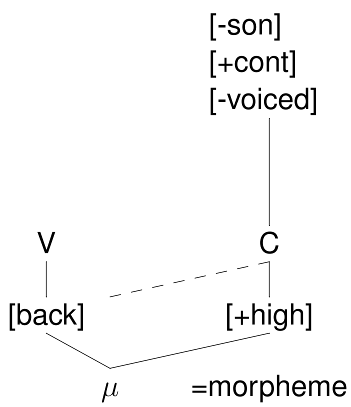
[1] According to the description of German in the IPA-Handbook (Kohler, 1990a), the dorsal fricative is uvular (/χ/) in some contexts. This section deals with the phonological treatment of dorsal fricatives though. As far as I can see, no additional problems would be introduced for the phonological approaches described here, therefore the distinction between uvular and velar cognates will be temporarily collapsed.

always realized as the palatal [ç] regardless of the preceding vowel. This is evidenced by oppositions like *Kuchen* ([kuːxən]) versus *Kuh-chen* ([kuːçən]). Still within structuralist tradition, this claim was not uncontradicted. The alternative analysis as allophones ([ç,x]) would predict the sole occurrence of the palatal in *-chen* from the morpheme boundary. Another attack on the allophonic analysis was launched by Cercignani (1979, after Jessen, 1988) who included data from proper names and foreign words in word-initial position into the analysis. This produces both ç and x: *[ç]arisma* versus *[x]an* or *[ç]irurg* versus *[x]imena*. Although not minimal pairs, this appears to make an analysis as separate phonemes more viable.

In generativist approaches, the central importance of minimal pair tests was abandoned in favour of the construction of *underlying* and *surface* representations. The principal problems remained the same though: The question now was whether the underlying representation is /x/ and [ç] as the surface representation had to be derived by rule or vice versa, whether the underlying representation is /ç/ and [x] as the surface representation had to be derived by rule. Again, both positions were held. Interestingly, the analysis as different phonemes seems to have played no role any more: Ramers und Vater (1992) and Lass (1984) view /x/ as the underlying phoneme for diachronic reasons, while many others take a distributional view arguing that /ç/ is the underlying phoneme because of its occurrence in more different environments as evidenced by the data in examples 3.1 and 3.2. These analyzes will not be presented, rather the next section continues with more recent approaches using the frameworks of Feature Geometry and Lexical Phonology.

1.2. Structure preservation

Hall (1989) tries to view the problem of the distribution of German dorsal fricatives as an instance of an assimilatory process. He deviates from the analysis assuming /ç/ as the basic, fully specified palatal, rather postulates that the dorsal fricative has no specification for [back] in the first place. Fricative assimilation spreads the feature of backness from a vowel onto immediately adjacent voiceless high fricatives. He derives [x] from a previously unspecified state which possibly is taken from a neighboring vocalic segment. Cases where no assimilation occurs are captured by a redundancy (or default) rule which assigns the feature [-back]. This assimilation can be represented like in example (3.3).



(3.3)

Fricative assimilation, i.e. the spreading of backness, occurs in all root-internal fricatives except in words containing a root and a diminutive suffix; there it is blocked. This is the case in words like *Kuh-chen, Pfau-chen, Tau-chen* (dim. cow, peacock, rope), which a successful analysis has to account for. Hall's attempt resorts to stratal morphology, by arguing that the domain of the fricative assimilation rule is level 2. Hall adopts the stratal division from Giegerich (1985) and Wiese (1986) for German shown in table 3.1:

	morphology
level 1:	class I derivation
level 2:	class II derivation, compounding
level 3:	inflection

Table 3.1.: Stratal division of morphology for German

The reason for locating FA (fricative assimilation) at level 2 is that within this stratal division, Umlaut is assumed to be just at this level 2 which must precede fricative assimilation. This is needed in order to account for alternations like Bach [bax] versus Bächlein [bɛçlaen]. Likewise, he argues against locating fricative assimilation on level 3, because then *Kuh-chen* would be incorrectly output as velar fricative as in table 3.2. This can successfully be avoided by the default rule at level 2 as in table 3.3. Note that the capital X in tables 3.2 and 3.3 refers to an unspecified state:

		Kuchen 'cake'	Kuhchen 'little cow'
		/kuːXən/	/kuː/
level 1:		-	-
Bracket Erasure		-	-
level 2:	*morphology*		
	add -chen	-	kuː]Xən
level 3:	*phonology*		
	FA	kuːxən	kuːxən
		[kuːxən]	*[kuːxən]

Table 3.2.: Stratal division of morphology for German: Without default rule, therefore the output is erroneous: *[kuːxən]

Hall (1989) puts forward similar arguments for the case of g-spirantization: In German, /ç/ alternates not only with [x], but also with [g], as for example in [veːnɪç] vs. [veːnɪgə]. g-spirantization only affects [g] among the German voiced stops though, is optional, sporadic,[2] and therefore is not viewed as systematic. Hall argues in a comparable fashion to qualify fricative assimilation to be on level 2 and not on level 3 of the stratal morphology. Now, distinctive segments with restricted distributions are anomalous in theories with structural orientation, and Hall's paper has to be seen as an attempt to render such segments nondistinctive by introducing them via rule. The *Structure Preservation* hypothesis (Kiparsky, 1985) claims that

[2] It only occurs in syllable-final position after /ɪ/, as suffix *-ig* or also as part of the stem.

		Kuchen ’cake’	Kuhchen ’little cow’
		/kuːXən/	/kuː/
level1:		-	-
Bracket Erasure		-	-
level 2:	*morphology*		
	add -chen	-	kuː]Xən
	phonology		
	FA	kuːxən	-
	Default Rule	-	kuː]çən
Bracket Erasure		[kuːxən]	[kuːçən]

Table 3.3.: Derivation of *Kuchen* versus *Kuh-chen* after Hall (1989)

non-distinctive features must be introduced at the *postlexical level*. It expresses the hypothesis that morphologically and lexically conditioned alternations usually are alternations between contrastive units like phonemes. In order to get the idea, go back to the derivation of *Kuchen* versus *Kuh-chen* as shown in tables 3.2 and 3.3 above: It is evident that level 2 is still on the lexical level. In Hall’s view this poses a challenge to the *Structure Preservation* constraint, because locating FA at level 2 introduces non-distinctive features coming with the dorsal allophones, which should be limited to postlexical levels.

This view was challenged in a paper by MacFarland und Pierrehumbert (1991) questioning that fricative assimilation can count as a counterexample to Structure Preservation due to the following argumentation: In the first place they draw on the marking condition which Hall (1989, p.13) uses in order to represent the fact that backness in [ç] and [x] is non-distinctive:

$$\begin{array}{c} * \begin{bmatrix} +\text{high} \\ \alpha\text{back} \end{bmatrix} \\ | \\ [-son] \end{array} \tag{3.4}$$

MacFarland und Pierrehumbert (1991) quote Kiparsky (1985, p.285) who states that marking conditions like the one in example 3.4 must apply not only to “underived lexical representations, but also to derived lexical representations”. They try to rescue Structure Preservation attempting “to reconcile FA and the marking condition” (MacFarland und Pierrehumbert, 1991, p. 174). They refer to Kiparsky’s analysis of Catalan nasal assimilation where he states that “the velar nasal /ŋ/ only occurs before /k/ and /g/ lexically. This restriction is reflected by a marking condition that restricts velar nasals to linked matrices”. MacFarland und Pierrehumbert (1991) emphasize Kiparsky's restriction to linked matrices: "This note calls to our attention the fact that the marking condition holds only if its structural description is matched exactly, that is, when the number of association lines in the structural description in question is equal to the association lines in the filter, namely one. When the structural description of the feature filter is not exactly matched, the filter does not apply, and the assimilation of the nasal to a following velar is possible." (MacFarland und Pierrehumbert, 1991, p. 174f.) They arrive at the conclusion that as a consequence, the marking condition given in example 3.4 does in fact *not* block the fricative assimilation rule, because "in order for an underived or derived lexical representation to violate the marking condition, its structural description must match the condition exactly. This is not the case for the forms created by FA. The rule spreads the feature [back] so that both segments, vowel and consonant, now share the feature, with the results that [αback] is doubly linked. The marking condition, however, specifies that [αback] be only singly linked" (MacFarland und Pierrehumbert, 1991, p. 174f.), and that therefore 3.4 is not applicable to the output of the lexical fricative assimilation rule. From these arguments, the authors conclude that although lexical fricative assimilation introduces a non-distinctive feature, it does *not* violate a marking condition, and therefore it does *not* oppose the Structure Preservation Principle.

Bybee (2001) discusses the same data on Structure Preservation, without explicitly making the attempt to refute MacFarland and Pierrehumbert's objections against Hall's analysis,[3] she rather points out that cases in which morphological status interacts with variable phonetic processes pose a problem for approaches like Lexical Phonology with its strong modular inclination. She brings back in diachronic evidence: The consistent realization of *-chen* as a palatal originated because the suffix had the form containing a palatal vowel earlier which conditioned the palatalization (*-ichiin*). It became lexicalized, then occurring outside the conditioning environment. She reasons that under these circumstances the palatal must already have been lexicalized, otherwise it would not have occurred outside its conditioning environment. Her conclusion expresses Bybee's discomfort with the strongly synchronic formulation of the Structure Preservation constraint, and that it, in her view, should be " be formulated with a diachronic tendency for phonetically conditioned variants to become contrastive and/or morphologically conditioned" (Bybee, 2001, p. 57).

1.3. Other recent approaches: OT

One aspect which became evident from the discussion of the paper by Hall (1989) is that the stratal morphology applied in this work is highly language-dependent, although Structure Preservation is seen as a kind of universal. In more recent years, Optimality Theory (OT, Prince und Smolensky, 1993) with its focus on the investigation of universal principles, linguistic typology and language acquisition has become increasingly popular. The core idea of OT is that the observed, "surface" forms arise from the resolution of conflicts between constraints. The ranking of these determines which of a particular set of candidates surfaces, because it is chosen by a criterion of maximal harmony (optimality). A sketch of this theory will not be provided in this place, for introductory reading refer to Kager (1999). Rather, this section will sketch and briefly discuss selected materials dealing with the topic here, German fricative assimilation: I will (i)

[3] She actually claims that there are strong diachronic tendencies for phonetically conditioned variants to become contrastive and/or morphologically conditioned.

briefly mention an analysis by Féry (in prep.), and (ii) sketch an analysis from a paper by Noske (1997). I will start with the analysis by Féry by a simple listing of the set of constraints her analysis is based on (3.5):

(3.5) (a) DORSFRICASS - Context-sensitive Markedness
A vowel and a following dorsal fricative have to agree in their specification of [back] (phonotactic)

(b) NOBACKFRIC - Context-free-Markedness
*[x]

(c) NOFRONTFRIC - Context-free-Markedness
*[ç]

(d) IDENT(BACK) - Faithfulness
Faithfulness constraint, input and output specifications have to agree in their specifications for [back]

(e) CRISPEDGE(PRWD)
the prosodic word must have sharply defined boundaries

These constraints enable to draw a tableau, the standard tool used in OT to display the functioning of an analysis, exemplified for the standard alternation where fricative assimilation is effective. For details on the notation see the caption of table 3.4:

buX	DORSFRICASS	IDENT(BACK)	*BACKFRIC	*FRONTFRIC
☞ bux			*	
buç	*!	*		*
byç		*!		*

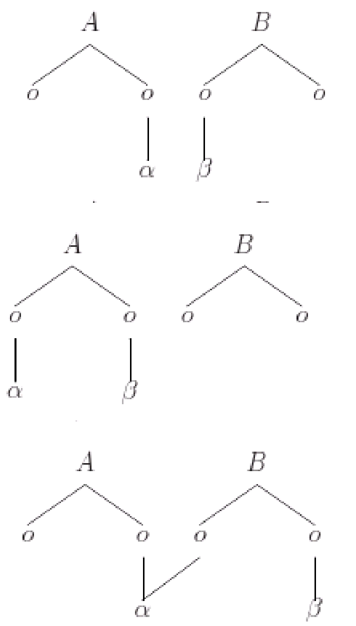
Table 3.4.: Tableau for the standard case of German fricative assimilation. The rows show the different output candidates, the columns the constraints. Their relative ranking is ordered from left to right. As common practice within OT, ☞ indicates the optimal output candidate, constraint violations are indicated by *, decisive constraint violations are indicated by!.

The interpretation is straightforward: The output candidate [buç] violates the highest-ranked constraint DORSFRICASS, because [u] and [ç] do not agree in their specification for backness. This is the highest-ranked constraint, therefore the violation is decisive. The output [byç] violates the faithfulness constraint IDENT(BACK) which militates against altering the back vowel in the input into a front vowel. The winning candidate ([bux]) violates the constraint NOBACKFRIC militating against the occurrence of [x], but these violations are not decisive. Table 3.4 was presented for the purpose of demonstrating a workable analysis in the currently dominating paradigm, OT. Without further discussion of this analysis, one comment is in place though: The postulation of a set of contradictory context-free markedness constraints (NOBACKFRIC and NOFRONTFRIC) appears to be ad-hoc, because providing constraints postulating "one alternative or the other" seems to me to be hard to reconcile with any technically correct scientific theory.

As mentioned the example shown in table 3.4 served illustrative purposes only. More interesting is the constraint from the CRISPEDGE-family. CRISPEDGE constraints were first applied to the fricative assimilation problem in Noske (1997), and I discuss the main argument of her paper instead in the following paragraph.

CRISPEDGE constraints shift the discussion back to the main theme phonologists have adopted in the more recent past, boundary-centered approaches in how to account for compound and diminutive constructions in German fricative assimilation as the main problems for FA patterns. The development of CRISPEDGE is related to an extension of alignment ideas - or better the class of ALIGN-constraints - from the morphology-syntax-interface to the mapping between *two prosodic categories* (McCarthy und Prince, 1993). Itô und Mester (1999) revised alignment theory "to treat multiply linked structures as coincidental" (see Noske, 1997, p.224). This comes at the cost of a problem which can be illustrated by example (6) in Noske (1997) also retyped here:[4]

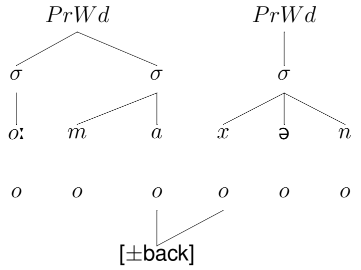
(3.6) (a) (b) (c)



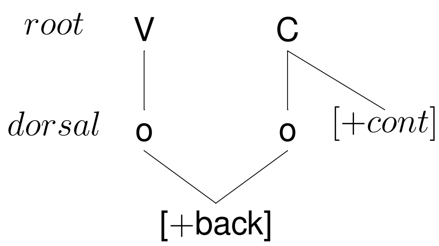
This reinterpretation of alignment - treating multiply linked structures as coincidental as shown above - cannot distinguish between the cases of a sharply defined edge like in 3.6(a) and a blurred edge like in 3.6(c) any more: In example 3.6(c), α is linked to both A and B, the higher order constituents, and Itô und Mester (1999) solve this problem by "introducing CRISPEDGE which requires that all prosodic elements are incorporated into a single higher prosodic unit and so rule out any cross-junctural linkage" (Noske, 1997, p.225). Noske fixes the domain of the CRISPEDGE constraint to the prosodic word. To give an example, CRISPEDGE is responsible for blocking the spreading of fricative assimilation over the two distinct prosodic words like in *Oma-chen*.

[4] She actually types the same representation twice to avoid an ambiguity between morphological and prosodic categories.

(3.7) *[oːmaxən] “grandma(dim)”



The same violation applies to composites like *Indo-china* (*[ɪndoxiːna],“Indo-China”). Her analyses comprise two further constraints, first the faithfulness IDENT-IO([BACK]) and CVLINKAGE. The faithfulness constraint IDENT-IO([BACK])’s purpose is to preserve the input - which Noske (1997) always chooses to be /ç/. CVLINKAGE is an OT equivalent for feature spreading to model the assimilation Noske adopts from Itô und Mester (1995). CVLINKAGE requires “a back vowel and a following dorsal consonant to share a single specification [+back].” Further, CVLINKAGE *requires double linking* of the feature value [+back], i.e. “a sequence whose individual segments are specified as [+back] violates this constraint”. It is shown in example 3.8:



(3.8)

Given an input like *Kuchen* (/kuːçn/, ’cake’) it is obvious that CVLINKAGE has to be ranked higher than IDENT-IO([BACK]) in order to produce the output [kuxːən], therefore it will not be displayed. Of greater importance are cases like *Oma-chen* as presented above, where the CRISPEDGE constraint is active, shown in table 3.5.[5]

The alignment constraint CRISPEDGE rules out candidate a., because a. involves a multiply linked structure. It also becomes evident that the incorporation of the notion that a sequence whose *individual segments are specified as [+back]* is necessary for CVLINKAGE to work: Otherwise candidate c. would not violate CVLINKAGE and rather constitute the optimal output form. The actual decision between candidates b. and c. is established by the lowest-ranked constraint IDENT-IO([BACK]), which prevents the [-back] input /ç/ from changing to the [+back] velar [x]. [6]

[5]Noske (1997) shows the same analysis to work for composites like *Indo-China*, a presentation of a tableau is omitted for the sake of brevity.

[6]The discussion of Noske’s paper was intended only as an instance of contemporary OT-approaches, it has not been the only one: Merchant (1996) proposes a different solution, but will not be discussed here. Noske (1997) further

	CRISPEDGE	CVLINKAGE	Ident–IO([back])
o:max[ən	*!		*
☞ o:ma[çən		*	
o:ma[xən		*	*!

Table 3.5.: Functioning of the analysis by Noske (1997).

At this point, the discussion is shifted away from the boundary back into the centre of phonetics again: In one of the papers discussed (MacFarland und Pierrehumbert, 1991), a phonetic explanation of German fricative assimilation in terms of coarticulation is explicitly ruled out. They put their rejection as follows:

> In the light on the literature on coarticulation [...], one might question whether assimilation and default rules are really needed. The alternative would be that the phonology does not assign any value of the backness features to the fricatives, and that the observed pronunciations are attributable to coarticulation with the neighboring vowel. However, two types of evidence lead us to reject this alternative. As shown below, the specification of the feature [back] is insensitive to morpheme boundaries, which are not available to phonetic rules. Similarly, the default rule has lexical exceptions in the form of two incompletely assimilated borrowings: *Chatschaturjan*, the name of a Russian composer and *Chupze*, a Yiddish/Hebrew expression, surface with a word-initial ach-Laut.

Further, they argue that in their own analyzes [ç] surfaces regardless of the backness of the following vowel, whereas the velars always tend to show the assimilation to the following sound which a coarticulation explanation for fricatives would require. This consistent realization as [ç] has not been left uncontradicted though, [x] is seen as typical for loanwords by most phonologists.

Summary

The section selectively summarized the phonology of German fricative assimilation in roughly chronological order, starting with older structuralist work proceeding to the most recent efforts achieved in the current optimality-theoretic framework via Lexical Phonology, which was presented in slightly more detail. This was necessary because in the papers by Hall (1989) and MacFarland und Pierrehumbert (1991) the currently still used representation of the German voiceless fricative as palatal was introduced, which necessitates the diligent treatment of the morpheme

argues that Merchant's solution fails to treat the schwa as underlyingly epenthetic, which leads to complications. Also omitted is the discussion of a rather recent monograph (Robinson, 2001) which according to Hall (2002) provides a novel analysis "of the same old" problem and a lot of new dialectal data.

boundary that also pervades the current OT-approach by Noske (1997). Typological frequencies further indicate no hard diagnostic for a markedness implication (following DeLacy, 2006), but there are fairly strong typological tendencies that render the prevailing analysis of contemporary phonology with the palatal fricative as input as a kind of special. Turning back to a more phonetic viewpoint, one might wonder about the asymmetry in directionality of the "assimilation" processes comparing [+cont] and [-cont] obstruents in German, i.e. fricatives and stops at the same dorsal places of articulation: While German has a single velar stop phoneme and contextual variability usually has no reflex in an IPA-transcription symbol,[7] the fricatives are transcribed by the allophonic system described in the preceding paragraph ([ç,x,χ]). The remarkable point though is that the direction of progressive assimilation for fricatives is reversed for the stops. If one were to recast this in coarticulatory terminology (see section 2.2), one would speak of *anticipatory* or *right-to-left* coarticulation for the stops and *perseverative* or *left-to-right* coarticulation for the fricatives of German.

2. The phonology of the Hungarian palatal stop

There has been a long-standing debate in Hungarian phonology whether the palatal should be treated as an affricate or a stop. As far as I can see, the argumentation from different authors seems to be a more diachronic one for the substantiation of this phoneme as an affricate, and a synchronic argumentation seems to be applied mainly by the advocates of the stop-classification. In a first step, both argumentations will be summarized in short separate sections.[8]

As palatal segments like the one in Hungarian have had a large impact on the development of *feature-geometric* representations within non-linear phonologies, these will be introduced in a separate section. These representations do not serve the purpose of modelling the phonological processes in the current work. Rather these representations have been used at the Phonetics-Phonology-Interface to derive interesting phonetic representations of the segments under considerations making testable predictions about coarticulatory behaviour.

2.1. The treatment as an affricate

This section first synthesizes the arguments as elaborated in Kálmán (1972) based on sound change. After that, other, synchronic arguments for the classing of the palatal as an affricate will also briefly be mentioned.

Kálmán's notation differs from standard contemporary IPA-notation; where necessary, this will be clarified. The main line of Kálmán's argumentation is that the development of affricates was facilitated by preceding processes of vowel deletion. According to Kálmán (1972), the diachronic development of the palatal began with a reorganization of Hungarian consonantism after the tenth century. Kálmán (1972) notes an increase in the number of affricates with the loss of root-final vowels starting as early as the ninth century as well as to the loss of some medial

[7] This is not fully true: Some authors use [k+] to denote the fronted version of the voiceless velar stop, others use [c] which according to the IPA is reserved for a plain voiceless palatal stop.

[8] Much on the literature on this topic is written in Hungarian and therefore was not accessible to me. Therefore this topic is only treated in a rudimentary fashion altogether.

vowels. (Kálmán, 1972, p. 54) assumes that the consonant cluster [ts] and [ds] were transformed into the affricate c.[9]

> “The affricates continued to be rearranged. In some of the dialects ǯ may have developed into a dˈ[10] as early as the 11th or 12th century [...] In the 14th and 15th centuries the process must have extended to cover, in all likelihood, the overwhelming part of the language area, t’ was restricted to certain local areas where it continued to survive. A more precise delimination of the period of this change, (for example ǯalog > dˈɔlog ’by foot’, orth. gyalog) is impossible, because the spelling of the period fails to indicate the exact quality of this consonant”. (p.54)

The emergence of the voiced counterpart somehow paved the way for the emergence of its voiceless cognate:

> “Following the phoneme change ǯ > dˈ, the phoneme č had no voiced correlate” (Kálmán, 1972, p. 55).

Their development is viewed in parallel:

> “Parallel with the phonetic change ǯ > dˈ, and not earlier than the thirteenth century, a new phoneme, namely tˈ evolved from tj and from the palatalization of t [...] The development of dˈ and tˈ followed parallel courses, although independent of each other”.

Apart from these changes primarily increasing the number of consonants, there are also systematic changes affecting place of articulation:

> “An important change was the development of some spirants and nasals into affricates. Although never assuming overall validity, this change has still affected a great number of words, so that the scope of the affricates was powerfully augmented and, notably enough, a new affricate was also produced [...]. This process can be documented sporadically from the tenth century on, but affricates began to increase in number from the fifteenth century on. In addition to the change z > ʒ mentioned above, the s > c development belongs here also [...].” (Kálmán, 1972, p. 57)

Figure 3.1 summarizes the shifts just described.

Apart from this diachronic foundation of Hungarian palatal segments as affricates, also synchronic reasoning could be applied. For example, Olsson (1992, 1993) defines the palatal segments as affricates on prosodic grounds. He argues that one should take the strongest position, i.e. the position most resistant to lenition as basic: If the oral palatal noncontinuants are phonetically realized as affricates before a stressed vowel and as stops in weaker positions, it follows that they are underlyingly affricates. A counter argument would state that genuine affricates are never realized as stops though. Finally, Nádasdy und Siptár (1989) mention phonotactic arguments which could be made in favour of an affricate classing, i.e.the occurrence of the palatals

[9] Kálmán (1972) uses c for a dental affricate ([t͡s]) like in German <Ziegel>. Similarly he uses č for [t͡ʃ].
[10] the voiced palatal (IPA: [ɟ]). In the same vein tˈ is used for [c].

in initial consonant clusters. For example, /t^y/ and /d^y/ never occur as first members in initial consonant clusters (/pr,pl,tr,kr,kl/ but */t^{y}r,t^{y}l/ and /br,bl,dr,gr,gl/ but */d^{y}r,d^{y}l/). It now could be argued that these gaps arise because /t^y,d^y/ are affricates. The counteragument articulated by Nádasdy und Siptár (1989) goes as follows: First, there are other gaps, not only the palatals are affected: */t^{y}l,d^{y}l/ but also */tl,dl/. Second, Nádasdy und Siptár (1989) argue that further gaps might be accidental: Almost all cluster-initial words in Hungarian are loanwords, and loanwords "will not include segments/combinations that do not occur in the language they are borrowed from."(Nádasdy und Siptár, 1989, p.20)

diachronic development of palatals (after Kálmán, 1972)

- loss of root-final (and some medial) vowels (9-13th century)
- increase of number of consonants (after the 10th century)
- regressive voice assimilation facilitated by vowel deletion
- augmented scope of affricates
  – z > ʒ (IPA: z > d͡z)
  – s > c (IPA: s > t͡s)
  – š > č (IPA: ʃ > t͡ʃ)
- – ǯ > d' (IPA d͡ʒ > ɟ) and
  – t/tj > t' (IPA t/tj > c)

Figure 3.1.: Summary of the chain shifts that led to the development of contemporary palatals (after Kálmán, 1972)

2.2. The treatment as a stop

Contrary to the treatment of palatal noncontinuants as affricates, as described in the last section, Siptár (1989) and Nádasdy und Siptár (1989, pp. 19-20) argue for a classification of these noncontinuants as stops. The reasoning is partly phonetic and partly phonological, whereby the variability of the phonetic manifestation and the participation in synchronic phonological processes are quoted as the respective evidence. Their main points are summarized in 2.2 and 2.2.

Phonetic motivation

Siptár (1989) argue that affrication occurs before stressed vowels and word-finally. Before unstressed vowels the affrication is weaker, before a stop no affrication occurs at all. The fricative component is also absent before /r/. Further, before /l/, lateral release is observed like in stops. /m/ may be preceded by slight affrication, but no affrication is heard before /n/ or /n^y/. Further, the degree of affrication depends on style and rate of speech: In slow, deliberate speech, affrication is stronger than in fast, casual styles. The author argues that true affricates would not exhibit such an extensive amount of variability. The listing in 3.9 summarizes these patterns of variability.[11]

(3.9) (a) contexts with strong affrication:

before stressed vowels	*tyúk*	'hen'
word-finally	*fütty*	'whistle'
	vágy	'desire'

(b) contexts with moderate affrication:

before unstressed vowels d^y	*mágyar*	'Hungarian'
before /m/	*hagyma*	'onion'

(c) contexts with weak/absent affrication:

before oral stops	*ágyba*	'to bed'
before /r/	*bugyrok*	'bundles'
before /l/	*fátylak*	'veils'
before /n/	*hagyna*	'we would leave some'
before n^y	*hegynyi*	'as a large hill'

Phonological motivation

The phonological argumentation for the stop classification is adopted from Siptár und Törkenczy (2000). They argue on the basis of the behavior of pre-stop allophones of stops: In this position, stops can be replaced by their non-released variants. Palatals do obey this pattern as shown in example 3.10, although there are vacillations in some contexts:

(3.10) hegytőll [hɛt^{y}˺töl], *[hɛt͡^{y}çtöl] , "from(a) hill"
hagyd [hɔd^{y}˺d] *(hɔd͡^{y}ɟ d), "leave (imp.)".

Further, affricates are resistant to OCP-driven[12] fusion across word boundaries. Stops are merged to geminates in any style of speech, whereas affricates remain unmerged in careful speech (e.g rác cég) [raːt^{s}-t^{s}eːg] "Serbian firm"). In colloquial speech, the first affricate may lenite into a fricative [raːst^{s}eːg].

Summary The phonemics of Hungarian palatal obstruents had to be touched as a long-standing debate, but apparently is not of much relevance in contemporary phonologies: The discussion about the representation of affricates seems to have been settled since the late 90's. This has to do with their representations as contour segments, i.e. segments which carry both opposite values of a distinctive feature. Such representations often were used for contour tones, but as well for the representation of affricates which are assumed to carry both values of continuancy, i.e. are specified as [±cont]. While there are discrepancies in the underlying representations, most theories agreed that the structure of affricates is more complex than that of simple segments. In

[11] This paper uses t^y and d^y for the palatals.
[12] Obligatory Contour Principle: Adjacent identical segments are ungrammatical.

contrast, Clements (1999) treated the affricates as strident (noncontoured) stops, and apparently the discussion about the correct representation of affricates has even settled since the publication of this paper.[13] Instead of furthering the "affricate problem", it seems more useful to proceed to a description of *feature geometry*. As mentioned, palatal segments and palatalization processes have had a large impact on their development. One of the few phonetic studies concerning the Hungarian language should be briefly mentioned though: Kovács (2002) reports burst-closure ratios, an acoustic marker frequently used to distinguish between stops and affricates. Her result suggest intermediate values for [c] and [ɟ] between plain stops and the other affricates of Hungarian.

3. Feature geometry

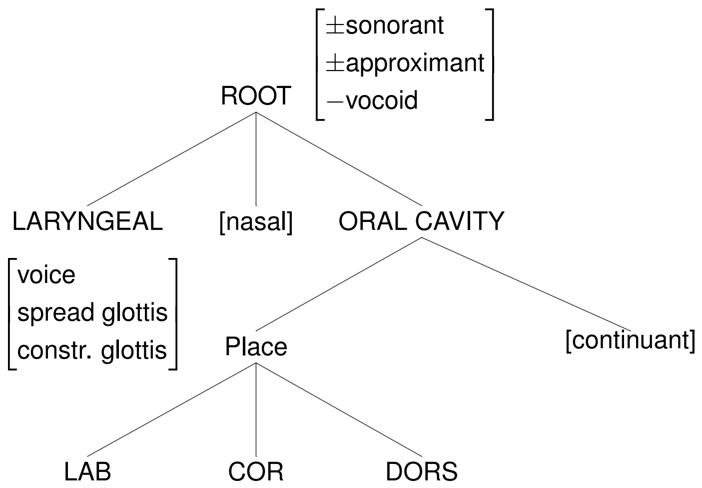


Figure 3.2.: The feature geometry described in Clements & Hume (1995, see p. 292).

Feature geometry (e.g Clements, 1985; Clements und Hume, 1995) can be seen as an extension of traditional featural approaches to segmental representation which allows processes to be analyzed with reference to particular features - of the segment - rather than to the entire segment. Feature geometry emerged from Autosegmental Phonology (Goldsmith, 1976), which had been the first theory in which features are active in tiers that may act independently of each other, i.e. in a nonlinear fashion.

Consider, as an example, homorganic nasal place assimilation as in German /va:gən/ → [vaːgŋ], "Wagen" (Kohler, 1990b, p.83). Here, the nasal assimilates to the preceding velar consonant /g/, but only in place of articulation, whereas other features like voice and nasality are not concerned. If one conceives the features concerned to lie on different autonomous tiers, Autosegmental Phonology already predicts that nasal place assimilation should occur. Feature geometry extends the autosegmental approach by organizing the tiers in a hierarchical system

[13]Marzena Zygis, personal communication.

of *nodes* which have the purpose to predict and constrain how segments are allowed to interact, e.g. in assimilation processes. The skillful organization of such hierarchies allows parsimonious representations of phonological processes like assimilations by grouping features which are frequently affected together as dependent nodes in the hierarchy. By assuming such a hierarchy, feature geometry simultaneously claims that changes involving a major node will cause the adoption of any dependent node.

On the dependent nodes features may be located that are free to *spread* to neighboring segments without affecting the higher structure, i.e. their parent node. Moreover, by locating certain features at a common node, one formalizes the fact that particular features often pattern together in their phonological behaviour. The complementary operation to spreading is called *delinking*. As an example, take the progressive voicing assimilation in English from /hæt+z/ to [hæts]. In order for voicing to spread from the penult voiceless alveolar obstruent to the word-final sibilant, [+voice] has to be delinked in order to enable the spreading of [-voice].

Taken together, feature geometry predicts phonological processes taking place at different hierarchical levels: Some processes will entail whole groups of features organized under a major node, others only a particular feature located under a minor node depending on a particular major mode. Only those features that form a constituent linked under a common node are expected to participate together in natural phonological processes such as assimilations. The topology of one of the most common systems (Clements und Hume, 1995) is shown in figure 3.2, for consonants. Note that the same feature tree, with minor differences is replicated for the representations of vowels.

A feature-geometric model specifically addressing palatalization processes and challenging earlier approaches was proposed by Lahiri und Evers (1991). It was designed with the particular interest of facilitating the description of palatalization processes like the fronting of velars which is of concern for the comparative treatment of palatal and velar stops which draws on a typological treatment by Bhat (1978). This paper is first summarized in short excourse before describing the system by Lahiri und Evers (1991).

3.1. Excourse: palatalization crosslinguistically

Already almost at the beginning of the paper Bhat wonders about the concept of palatalization:

> "palatalization has been considered a single diachronic (or morphophonemic) process by linguists, and it is represented by a single process in traditional terminology". (Bhat, 1978, p.49).

This holds "even though the generative terminology represents it as a two-fold process, i.e. 1) the change of velars to a [+high -back]-position and of the dentals and labials to a [+high] position." (Bhat, 1978, p. 49). Anyway, his approach is typology-driven and palatalization is treated as a single diachronic process while attempting to make generalizations regarding "its occurrence, spread or disappearance in various languages" (Bhat, 1978, p. 49). According to Bhat (1978), this results in at least three distinct (sub-)processes contributing to palatalization, and Bhat's interest lies in the presentation of these processes as independent ones. They are described by Bhat as follows associating their acoustic consequences:

tongue-fronting "raise in the frequency of the second formant"

tongue-raising "lowering of the frequency of the first formant"

spirantization "addition of stridency (or frication)"

Although rather simplistic (see e.g. the association of tongue-fronting with second formant rising and compare it to section 2.1), this association of distinctive features like [±back] and acoustic features like associated second formant movements, and in particular, the terminological division into distinct (sub)processes prooved stimulating for scientific discourse concerning palatalization processes. Bhat lists as an example the misconception of a palatalization process as an instance of a "depalatalization":

> "The change of k to ts has been regarded by some linguists (see Newton, 1972) as a case of depalatalization, i.e. a change of k > tš > ts. Such an assumption would be unnecessary if palatalization is analyzed as made up of three distinct constituent processes." (Bhat, 1978, p. 59).

He also already noticed the division between palatal versus palatalized segments and the crosslinguistic trend of associating - optional! - spirantization to velar palatalization:

> "Both velar and apical stops, when palatalized, generally become spirantized as well. But there do exist instances in which the latter change has failed to take place." (Bhat, 1978, p. 59).

3.2. The system of Lahiri & Evers

Inspired by Bhat's work, Lahiri und Evers (1991) summarized the most frequent processes recurring under the cover term palatalization. These are (i) the fronting of velars already mentioned repeatedly in the current work, (ii) the change of place of consonants, e.g. alveolar and dental consonants becoming palato-alveolar or prepalatal in the context of front vowels and [j] like for example in English <miss you> and (iii) the addition of a secondary palatal articulation to almost any consonant. They devise their own feature hierarchy, tailored to model these different palatalization (sub-)processes. In the first place, the topology of the tree and the moves which differentiate it from other geometries will be described. The hierarchy is displayed in figure 3.3.

First of all, and not evident from figure 3.3, they reject the copying of the whole feature tree for vowels and for consonants as in Clements (1985) and instead argue for a "unitary set of features for vowels and consonants" (Lahiri und Evers, 1991, p. 98). A further point to note is that [high] is conceived as binary unlike in some other geometries. They propose a set of common features among front vowels, the palatal glide and (alveo)-palatal consonants, which are grouped under a single articulator node. This move is assumed to model feature spreading that characterizes the various palatalization processes in a simpler way, as to be shown. In a first step consider the featural representations in figure 3.4 arising from the postulated structure.

Without going into too much detail about the system developed by Lahiri & Evers, it becomes quickly visible that spreading the Coronal Node together with its dependent [-anterior] feature

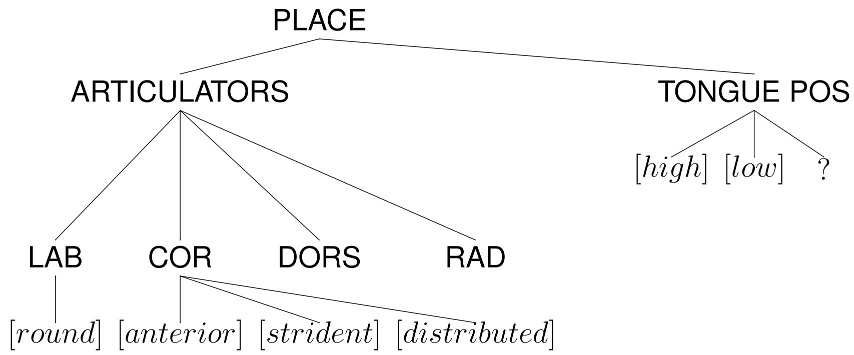


Figure 3.3.: The place node in the feature geometry model of Lahiri und Evers (1991, p. 87).

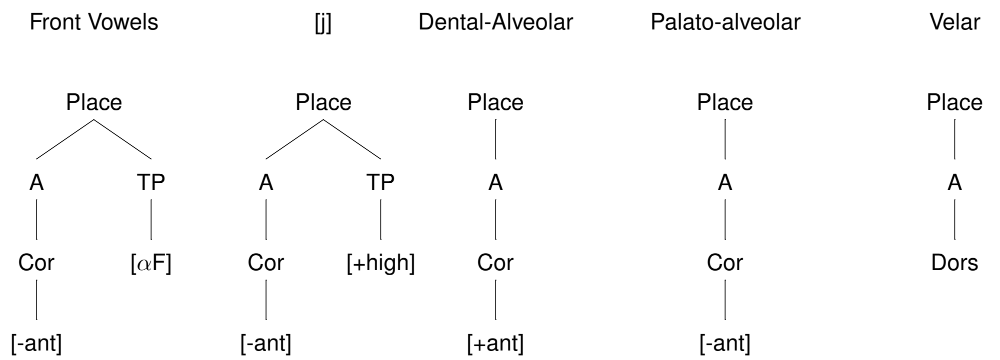


Figure 3.4.: Featural Representations according to Lahiri und Evers (1991, p. 90).

and a simultaneous delinking of the dorsal node results in an elegant way to represent velar fronting, as demonstrated in figure 3.5.

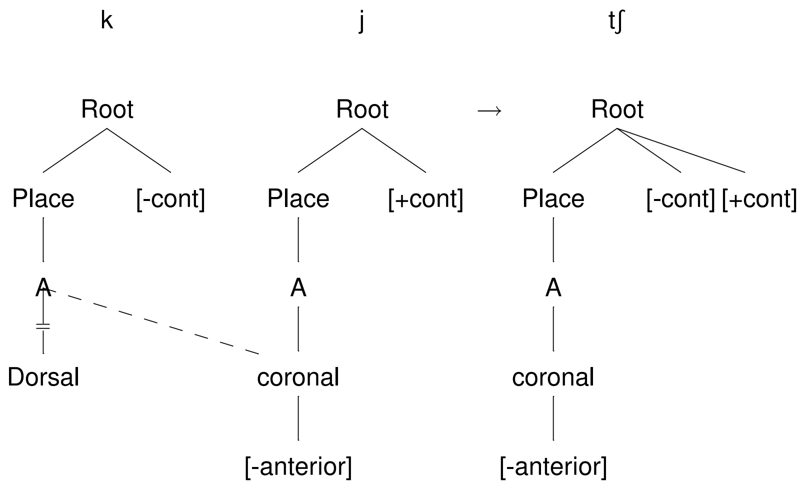


Figure 3.5.: Velar fronting according to Lahiri und Evers (1991).

They prefer this representation of velar fronting to the one offered by Clements und Hume (1995), which is given in figure 3.6.

Clement's representation models velar fronting in much the same way as labialization processes (e.g. [ku > kw k^w]): In the first step, the consonant acquires a secondary articulation from the V-Place of the following vowel (in the case shown in figure 3.6 from the palatal glide). At this stage it is represented as a complex segment. It is then transformed into the corono-velar by tier promotion, i.e. coronal is moved from the V-Place to the C-Place tier. Finally, to achieve the output tʃ, the complex segment is simplified - i.e. the dorsal node is deleted - and affrication added. Lahiri und Evers (1991, p.85) argue that such modelling gives a good account for the labialization process, it does not do so for the velar palatalization: (i) it is not clear what triggers the tier promotion, (ii) one would have to assume a [k^j], i.e a multiply articulated coronovelar as an intermediate output in palatalization and (iii) of particular relevance for palatalization, velar and dental/alveolar palatalization would have necessarily to go through a stage of secondary articulation like the palatalized velar ([k^j]) in the example in 3.6. This seems not to occur in all historical developments though. In essence, the problem arises from the modeling which requires that the assimilation must link V-Place coronality to the consonant first before spreading and tier promotion can occur. As mentioned, Lahiri und Evers (1991) provide an own account of the same data, which was already depicted in 3.4 and 3.5 for the same velar fronting data. Velar fronting was modeled by spreading the coronal node from the following vowel to the consonant, and at the same time delinking and deleting the dorsal node of the velar stop. Likewise already mentioned was the fact that this was achieved by a different account of the feature tree, which consists in sum in (i) a unitary representation for vowel and consonant features (ii) while

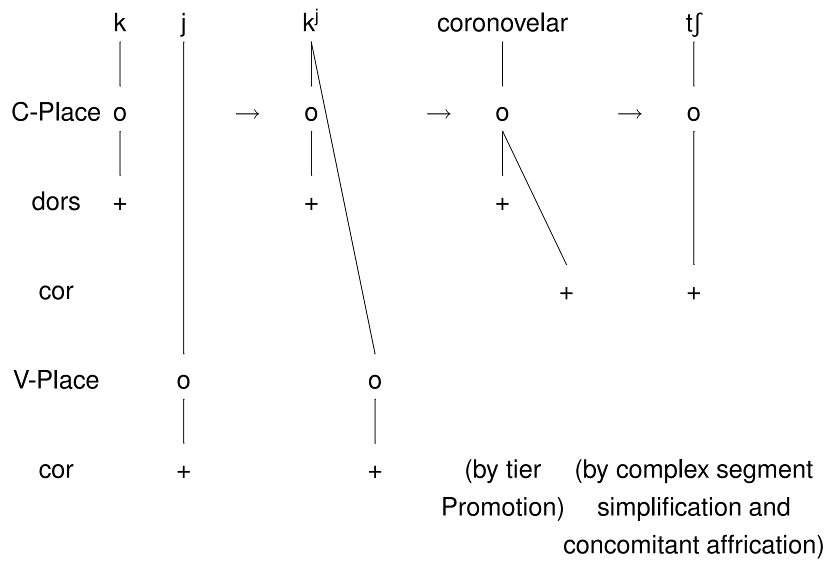


Figure 3.6.: Velar fronting according to Clements, after Lahiri und Evers (1991, p.84).

grouping coronal consonants, front vowels and the palatal glide under a single coronal articulator node.

3.3. Summary and outlook

This section provided a miniature of an introduction to feature geometry, displaying the most essential properties of a hierarchical structuring offering the possibility of a more efficient representation of phonological processes. The most important operations like spreading and delinking which can occur at various levels of the hierarchy were described. Admittedly, the approaches and processes described were highly selective, i.e. it is not concealed that there are numerous other approaches to feature hierarchies. The selection of Clements' view and the one advocated by Lahiri und Evers (1991) was primarily motivated by strategic reasons though: Although the modeling of velar fronting by Lahiri und Evers (1991) appears to be more elegant in the first place, it seems legitimate to speculate on the reasons why the conception of unitary features they propose apparently had so little impact on the development of feature systems. A partial answer - which is surely not the whole truth - again comes from the Phonetics-Phonology Interface: Articulatory Phonology (a preliminary sketch was given already in section 2.2) with its alternative and biologically more plausible timing conception had a large impact on the field, and one of its core aims was to deliver a representation that warrants a parallel of Autosegmental Representations - as delivered by Feature Geometry - with the "Functional Anatomy of the vocal tract" as elaborated by the gesturalists. This parallel was positively evaluated already early in the development of Gestural Phonology as evidenced by the following quote:

> " [...] there is much similarity and compatibility between feature geometry and the geometry of phonological gestures [...]. We argue that the gesture, and gestural scores, could usefully be incorporated into feature geometry." (Browman und Goldstein, 1989, p.221f)

Browman und Goldstein (1989) take an even stronger position and have argued that the convergence on a single geometry from the directions of Autosegmental Phonology and of phonetic function provide strong support for correctness of the geometry. In other words, researchers from phonology and phonetics would have found it difficult to agree on a geometry like the one proposed by Lahiri und Evers (1991), because the relationship between phonological and phonetic representations is severely attenuated. In order to arrive at a summary, this section apart from giving a very brief sketch of feature geometries constitutes to some extent the connective link to the derivation of particular hypotheses. It allows the construction of two further theoretical building blocks to be completed resuming discussions from the introductory part I. Now that the parallel of feature geometries and the functional vocal tract has been established, it becomes possible to describe the lexical part of Articulatory Phonology in necessary detail, which had been left blank in section 2.2 of the introductory part. Further, the introductory part mentioned the existence of an *interpolation-based* model of speech production as advocated - and partly extended in Keating (1990a). The window model of coarticulation (Keating, 1990b) is an instantiation close to featural models, and has been termed an "alphabetic model" in the introductory part. These two approaches will be reviewed in the following two sections.

4. Keating's approach: windows and underspecification

Recall the situation in the period after the publication of SPE: There is a clearcut separation between coarticulation and other context-dependent phenomena, such as assimilations: coarticulation deals with "transitions between vowels and adjacent consonants, the adjustments in the vocal tract shape made in anticipation of a subsequent motion" (SPE:295, cited from Farnetani und Recasens, 1999). In contrast, assimilations involve operations on phonological features, and are accounted for by phonological rules, which map lexical representations onto phonetic representations. Coarticulation and other properties of phonetic implementation are assumed to follow from universal principles of speech physiology.

The "window model of coarticulation" elaborated by Keating (1990b) blurs this clear-cut distinction between the grammar and the physics of speech.[14] Essential for her view is the concept of underspecification. While the concept of underspecification was located in the work of Keating's predecessors within phonology, the novelty consists in the expansion of underspecification into the phonetic domain. In Keating's view, the grammar has a phonological *and* a phonetic

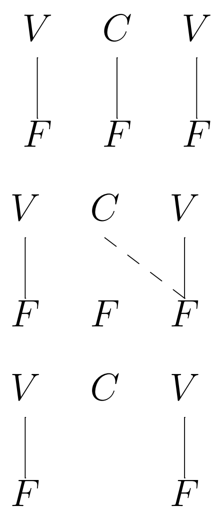
[14]It has to be mentioned though that this clear-cut distinction between the grammar and the physics of speech has been abandoned earlier by approaches stemming from within featural phonology, e.g. in approaches like feature spreading (Hammarberg, 1976), the look-ahead model, (Daniloff und Hammarberg, 1973), or the model of coarticulatory resistance (Bladon und Al-Bamerni, 1976).

component. This phonetic component of the grammar has the function to specify whether a feature is relevant in the paradigm in a given language. Keating exemplifies this point by mentioning language-specific implementation patterns of voicing in languages like English or German on the one hand and languages like Polish or Czech on the other. This kind of language-specific implementation has to be accounted for in the grammar. Altogether, there are three different ways to deal with underspecification: There is underspecification on the phonological level, which may persist in the phonetic domain. Then, there is phonetic underspecification which is conceptualized as a continuous notion. Additionally, unspecified features may be left unspecified or specified by rule. Now, if phonological assimilation rules assign a contextual feature to a segment, its associated window will be narrow before that context and the contour will have a plateau-like shape. If assimilation rules are not active, the key feature remains unspecified and the trajectories will be provided by "interpolation". Furthermore, inter-language differences in coarticulation can be of phonological or phonetic origin. If phonological assimilation rules operate in one language and not in the other, they are of phonological origin. They are of phonetic origin, if different languages interpret an unspecified feature differently. The window-model was already mentioned in the introductory part, and the problem to answer the question "phonetics or phonology" was already demonstrated on the basis of the preglottalization example comparing British and American English (see section 2.2). In the introductory part, the window model was classed as one instance of a "target-and-interpolation" approach. This is in principle true, but the slight change of emphasis in comparison to "proper" target-and-interpolation theories should be mentioned. It consists in a larger emphasis of systematic and random variation in comparison to targets and turning points over more pronounced target approaches. In target-and-interpolation approaches, targets were seen as invariant, and variation arises from the process of "connecting" these targets. Surface variation then arises from "constraints on speed of movement", i.e. the targets may be over- or undershot. The window model, in contrast, expresses targets not as static invariants, rather, as the name "window" already suggests, as target ranges. The windows then are best seen as a device to express "the observation that some segments vary more than others along various articulatory dimensions" (Keating, 1990b, p. 466). It therefore is to some extent comparable to previous work addressing the same issue: For example, Bladon und Al-Bamerni (1976) and Bladon und Nolan (1977) attempted to capture differences in contextual variation of different segments in *coarticulatory resistance* scales, which assigns a numerical coarticulatory resistance value to a segment, in some sense constituting a feature in its own right. The window model is similar in spirit: High coarticulatory resistance values indicate narrow windows, and low coarticulatory resistance values indicate wide windows. The notion of the window as a range of possible output values reminds of the work by Manuel und Krakow (1984): There, phonemes are not represented as invariant points, but rather as areas or regions, the size of which are controlled by "output constraints" which posit that no phoneme can intrude into another phoneme's area. The prototype here is the two-dimensional vowel graph, and therefore the size of the inventory determines the size of each single phoneme's space and therefore its possible contextual variability. In other words, output constraints are more a property of a whole class of segments - e.g. like the dispersion in vowels space - rather than of single segments. The window model slightly blends the idea of output ranges from output constraints à la Manuel und Krakow (1984) and the non system-specific but rather atomic specification of variability patterns into the window model. Window width is crucially related to featural specification, for example underspecification pertaining to the phonetic level relates to the generation of wide windows. This will be further elaborated in the next section about velar (under-)specification.

4.1. The status of velars

As has already become evident from the previous section, the work by Keating aims at a seamless integration of phonetic variation into the theoretical apparatus of a feature-based alphabetical model. Unlike e.g. the more phonological, morpheme-boundary centered papers described in the sections about the phonology of the German dorsal fricative like in example (3.3) by Hall, there is no reference to the morphology, rather Keating's (1988b, p.287) representations are confined to phonetic VCV sequences. With regard to a feature F three different output representations are possible in such a scenario, with distinct phonetic reflexes:

(3.11) (a) (b) (c)



In example 3.11(a) each of the segments is specified for a value of F, and there is no vowel-to-vowel effect in either direction, because they are blocked by the feature values. Predictions made from example 3.11(b), where the consonant acquires a value F from V2 by a feature spreading rule, would result in (i) no coarticulatory effect of V1 on V2 because the consonant's specification for F exerts a blocking effect as in 3.11(a), and (ii) effect of V2 on V1, because the (acquired) consonant feature value will affect V1. The transition between V1 and C will not be affected and be similar to 3.11(a), but since in the spreading representation 3.11(b) the consonantal specification is inherited from V2, variation in V1 will be dependent on both the following consonant and vowel. Finally, in example 3.11(c), vowel-to-vowel effects will occur in both directions, and the consonant will lack a phonetic quality of its own, and the quality of the transitions will be gradual.

Of course, these specifications are not intended to undermine consultation of the phonological patterning of a language: The scenario in example 3.11(a) only shows that a feature has a value output from the phonological "modules", and does not make any statement whether the feature has an underlying representation or is a later fill-in. Therefore, the phonology of a language must be consulted to discover the source of the feature value. The role Keating assigns to phonetics then is to provide evidence about which segments are underspecified for which features.

The treatment of dorsal fricatives in Russian

Generally, Keating assumes that if a feature participates in a phonological contrast, vowel-to-vowel interactions are blocked. On the contrary, when a segment does not contrast and has no feature value, such interactions are allowed. One of her examples comes from Russian palatalized versus nonpalatalized fricatives, which she lists as follows: [15]

(3.12)

f	s	ʃ	x
fʲ	sʲ	ʃʲ	

In contrast, the opposition between palatalized and plain is complete for stops (taken from Padgett und Zygis, 2003):

(3.13)

p	t	k
pʲ	tʲ	kʲ
b	d	g
bʲ	dʲ	gʲ

The asymmetry between stops and fricatives is quoted in Keating's own words: "The velar /x/ is described as allophonically variable. Even speakers who allow a contrast in the velar stops have none for the fricative" (Keating, 1988b, p. 288). She relates these phonological data to the classical study by Öhman (1966) who made phonetic recordings of plain and palatalized stops in Russian. His major finding was that the consonantal transitions (V1C and CV2) depend on the identity of the transconsonantal vowels But this coarticulatory variability was reduced to almost random fluctuation in the case of Russian. Öhman interprets these findings as follows: The tongue is considered a system of independently operating articulators driven by invariant articulatory commands. The apical articulator is involved in the formation of apical consonants, the dorsal articulator in the formation of palatal and velar consonants and the tongue body articulator in the formation of vowels. The reduced coarticulatory variability for the palatalized F2-transitions is seen as the result of conflicting vowel commands on the tongue body, i.e. an [i]-like palatalization commands exerting a blocking effect on the following vowel. This situation changes for the fricatives, Keating refers to data from Derkach et al. (1970) as well as to her own acoustic recordings. The major result is that for the velar fricative, as opposed to stops and fricatives at other places of articulation, coarticulation is *not* blocked. She observes a strong assimilation of the velar fricative to a following vowel, and because of the strength of this effect, the vowel preceding /x/ is also influenced. This effect is particularly strong in /axi/-sequences.

The treatment of dorsal stops

In another publication, Keating discusses the phenomena of velar fronting in terms of surface features for English. She states that velars "lack inherent specification for Back" (Keating,

[15] I am aware that this seems to be at odds with the system as described by other authors, who assume a contrast between palatalized and nonpalatalized velar fricatives (like e.g. Padgett und Zygis, 2003)

1993, p. 17). This implies the assumption of the underspecification scenario for contextual velar variation which was shown in example 3.11(a). Consider for example the VCV-sequence [aki]. The concrete within this underspecification scenario could be written as:

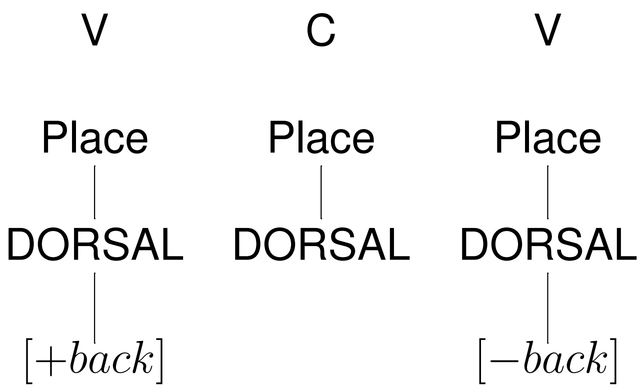


Figure 3.7.: Underspecification representation of contextual velar variation in the sense of Keating.

Further:

> "Stated another way, velar fronting is something that happens gradually over the course of the velar. Such temporal/spatial variation, or phonetic gradience, can be interpreted as a transparency effect on the velar with respect to backness." (Keating, 1991, p. 17)

One could ask here: What are the phonetic manifestations of transparency, how does "graduality" manifest itself "over the course of the velar"? Keating provides guidance explicitly relating to work by Houde (1968) and the loops. These loops have often been observed in V1-[Velar]-V2-sequences and describe the behaviour of the tongue dorsum during velar closure. A separate section has already been devoted to the loops issue in the introductory part (see section 2.2), therefore resuming this discussion is not necessary here. For the present purpose it is sufficient to note that the phonological description of the the velar as an underspecified segment at the same time leads Keating to expect large movement amplitude of the tongue during the interval of oral closure.[16]

This sketch seems to tacitly assume large "sliding" movements of the tongue during the closure interval, and it seems not to be too unfair to relate this large sliding movement to *movement amplitudes during oral closure* as to be derived from EMA recordings. This is a point which will be resumed later. But first, Keating's proposals for the featural specification of palatal segments will be described.

[16] This is also evidenced by a hand-drawn sketch in this publication (see Keating, 1993, p.18)

4.2. The status of palatals

In a whole series of papers, Keating (1988a, 1991, 1993) elaborated the status of palatals. Inspired by phonological anomalies and relative markedness in language inventories, Keating postulated a special status for palatals distinguishing them from “ordinary” segments. Keating (1988a) proposes that palatals should be considered complex segments on the basis of x-ray evidence, and Keating und Lahiri (1993) further argue that since both tongue blade and tongue body gestures are involved in the articulation of palatal segments, they should be considered phonologically complex in the sense of Sagey (1990). This representation is shown in figure 3.8:

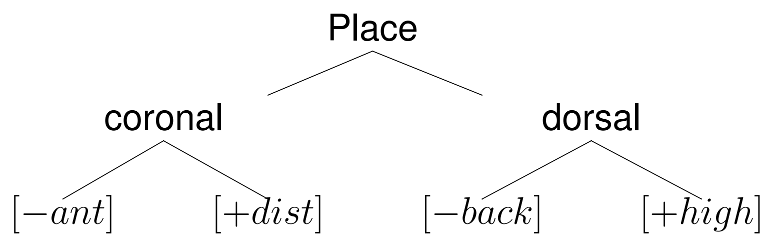


Figure 3.8.: Complex specification for the palatal noncontinuant as both a [-anterior] coronal and [-back] dorsal.

From an articulatory point of view, Keating (1991) describes a raised, fronted tongue body as indispensable for the palatal noncontinuants, noting the “extensive side-to-side and front-to-back lateral contact as for [j]” (Keating, 1991, p.37f.). She observes that in palatals the tongue is both higher and more fronted than for [i], the tongue blade and front forming a very long stricture. In other words, there is a clear distinction between the vocalic nature of /i/ and the consonantal stricture of consonants involving the raised tongue body. This derivation of the “true” palatal from the high front vowel and the demarcation from the alveolopalatal is achieved in (Keating, 1988a):

> “In contrast to the alveolopalatals, “true” palatals are being articulated with tongue surface further behind the blade area, forming a constriction extending further back on the palate. [...] This is because palatals are articulated like “consonantal” front vowels.” (Keating, 1988a, p.87)

As mentioned, the articulation of these palatals should involve both coronal and tongue body articulations, such that they would have a - complex - status like double-articulated labial-velars. The featural representation of the palatal then, as shown above, has place specifications on the dorsal as well as on the coronal node.[17] From a point of articulatory control, this implies that

[17]Keating (1991) mentions another possibility for the representation of palatals: ”However, another option in the representation of palatals is to treat them as simple coronals, and introduce at least one additional feature to distinguish them from the [-anterior] coronals. This is in fact what Halle (1988) does with his new feature Lower Incisors Contact. Actually, both options should be exercised for more descriptive coverage.” (Keating, 1991, p. 45)

adjacent parts of the tongue should be controlled separately. This idea is demonstrated in figure 3.9 which is an adaption of Keating's own hand-drawn sketches published in Keating (1993, p. 16):

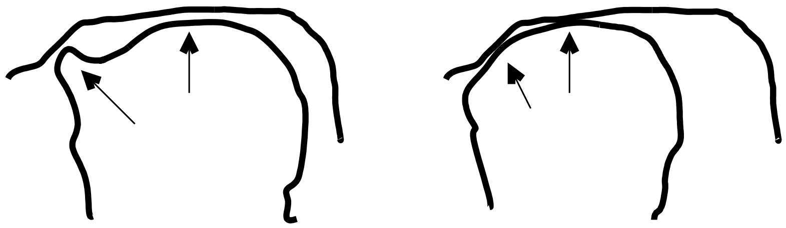


Figure 3.9.: Keating's simplification of the complex specification for palatal plosives into a blended articulation: Two separate articulations (left)get simplified, the outcome being an articulation with a single, long constriction (right).

Finally, Keating also provided an account of secondary articulations. For completeness, the featural specification of palatalized velars is also displayed (see figure 3.10).

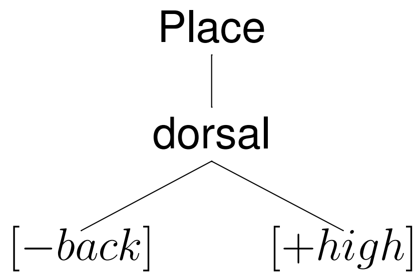


Figure 3.10.: Keating's representation of palatalized velars.

4.3. Summary

This section described the integration of featural description and articulatory variation as described in numerous papers by Keating. Keating's approach has been referred to as an *alphabetical model* in Beckman (1999) and also repeatedly in this work. From its characterizations of (i) dorsal obstruents as underspecified for backness and (ii) palatal stops characterized as complex segments, it becomes possible to develop at least gross operational hypotheses. One dependent variable for quantitative analysis was already mentioned: Movement amplitudes during closure intervals. But some further statements seem warranted within this framework: The complex representation of palatals and the underspecification representation of the velars appear to be two extremes of possible specifications for window widths, and it seems appropriate to assume extremely narrow windows for the palatals and the opposite for the velars. This amounts to little sensitivity to vowel-induced contextual coarticulation for palatals in contrast to velars. At the same time, given different featural descriptions, front velars should be distinguishable from palatals in their shape configurations, velars lacking the /i/-like "component". If the surface underspecification Keating adopts for English is also valid for Hungarian velars, then the velars should exhibit relatively large vowel-dependence in place of articulation. Furthermore, briefly reviewing her statements on the phonetic implementation on contextual velar fronting: "Stated another way, velar fronting is something that happens *gradually over the course of the velar*" implies that the *"transparency effect of the velar with respect to Backness."* (Keating, 1993, p. 17) is a specific of the velar stops. This in turn implies that this transparency effect should be absent *over the the course of the palatal*. Putting this together, according to Keating, velars and palatals should be distinguished by the amount of coarticulatorily induced variability *during* these stops.

In the introductory part I, the description of the non-alphabetic alternative AP was initiated, but not completed. The emphasis there was more on the alternative conception of time in comparison to the *alphabetical* theories described hitherto, and although there was a mention of the mechanism of gestural blending which is of crucial importance for velar consonants within AP, the discussion mostly was refined to single gestures, which the authors describe as the "atoms" of articulatory organization. The following section attempts to present this stage on the adjacent more aggregate level, the lexical part of AP, which deals with the "molecules" formed from these elementary gestures (see Browman und Goldstein, 1989, p. 201 for these notions).

5. Articulatory Phonology II

In chemistry, a molecule is defined as a sufficiently stable electrically neutral group of at least two atoms in a definite arrangement held together by strong chemical bonds. The concept of the gesture within Articulatory Phonology as the combinatoric atoms of phonological representations was already elaborated in the introductory part. A gesture is an elementary *action* unit which is defined by the task it has to perform. For example, for the objective of producing a bilabial stop, a gesture with the objective of forming a bilabial closure is necessary. In order to achieve this - linguistically relevant - goal, active movements of the articulators upper lips, lower lips and the jaw have to be coordinated in a temporally suitable fashion. A further building block of the definition of the gesture concept is that of *active control*, i.e. passive concomitant tissue movement cannot count as an action unit. The dynamical definition of gestures as mass-spring systems was already displayed in the introductory part. In the equation $m\ddot{x}+k(x-x_0)=0$ displayed there, there was no friction, and therefore the equation has to be extended in order to display a *a critically damped mass-spring* system:

$$m\ddot{x}+b\dot{x}+k(x-x_0)=0 \tag{3.14}$$

Now, changes in the dynamic parameters of such a system can generate different trajectories depending on the parameters altered: Within Articulatory Phonology, the mass of the spring is usually left unaltered and fixed to 1, at least in older versions of Articulatory Phonology. The damping *b* has to be specified such that the mass does not oscillate around its rest position, i.e. stable targets are attained. The equation of motion for the undamped system was displayed in the introductory part, where it resulted in sinusoidal oscillation as evidenced in equation 1.5. In a critically damped system, the target position of the spring is the rest position of the spring. The effect of the spring constant can also easily be inferred from 1.5, i.e. the undamped system: It is in the denominator of the formula defining the angular frequency ω, and therefore defines the frequency of oscillation of the system. Its role in the critically damped system changes though: It is *the* defining property of the characteristic of a particular gesture, because it determines its *temporal characteristics*: The higher the spring constant, the faster the gesture attains its target position and the shorter movement duration. At least older versions of Articulatory Phonology do not differentiate between the major sound classes, only between vowels and consonants with vowels being assigned lower values.

AP has some other interesting properties: The definition of the systems as just described is[18] a point attractor, i.e. under any starting conditions, it converges to a single point, the articulatory target. This allows a *context-independent* target attainment through settings of initial conditions alone, and no step-by-step interpolation and predefined targets are necessary like in target-and-interpolation approaches like Keating's. Further it offers an appealing model of motor equivalence, which means that through the adjustment of few settings whole sets of trajectories can be generated.

5.1. Tract variables and gestural scores

How are these atoms just described aggregated to molecular structures mentioned at the beginning of this section? The organization of articulatory movement defined in terms of particular tasks to be performed - in the example above lip closure - was already mentioned. The notion which is definitory for the task description is the *tract variable*. An overview of these tract variables is given in figure 3.11. Tract variables describe *location* (CL) and *degree* (CD) of a constriction. Constriction degree is always present, whereas constriction location is only relevant for oral gestures. Stiffness refers to the *k* values of the tract variables. Further, there were plans for a tract variable *constriction shape* which was not yet implemented in Browman und Goldstein (1989). CL and CD are furnished with sets of descriptors as follows:

CD closed, critical, narrow, mid, wide

CL protruded, labial, dental, alveolar, postalveolar, palatal, velar, uvular, pharyngeal

Articulatory movements are generated by these tract variables. It is important to note here that the tract variables are conceived as relatively abstract, discrete entities:

> "They are discrete in two senses: (1) the dynamic patterns of a gesture's control regime remain constant throughout the discrete interval of time during which the gesture is active, and (2) gestures in a language may differ from one another in discrete ways, as represented by different descriptor values." (Browman und Goldstein, 1989, p. 210).

This is a partly strategically motivated decision, because using this conception gestures can "function as units of contrast (and more generally capture aspects of phonological patterning)"

[18] An attractor is a set to which a dynamical system evolves after evolution of the system in time, even under disturbed circumstances. Attractors can be a point, a curve, a manifold, or a "strange attractor". Strange attractors point to the domain of *chaos theory*.

	tract variable	**articulator involved**
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, tongue body, jaw
TTCD	tongue tip constrict degree	tongue tip, tongue body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

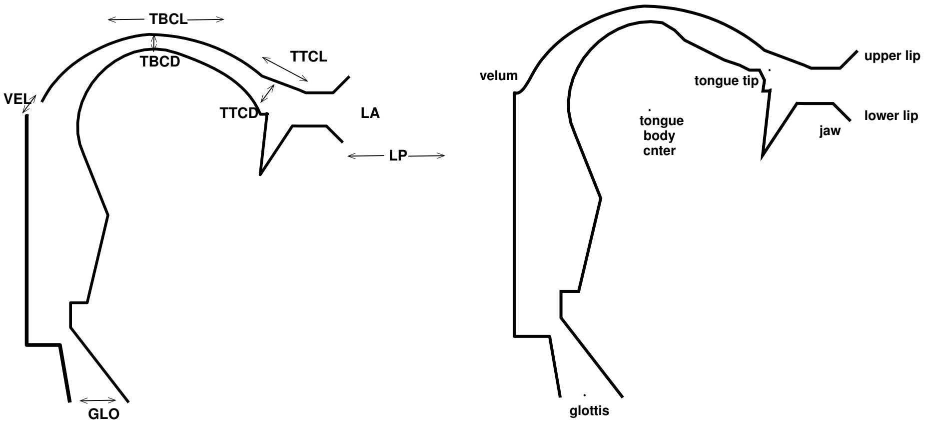


Figure 3.11.: Tract variables and their articulator associations in the framework of Articulatory Phonology, adapted from Browman und Goldstein (1990a, p. 301).

but still "characterize actual observed articulator movement (thus obviating the need for any additional implementation rules)." (Browman und Goldstein, 1989, p. 210)

The generation of articulatory movement by the means of these abstract gestures works by means of recruiting the necessary sets of - the more concrete - articulators, i.e. a gesture is defined as an articulatory "control regime".[19] Tract variables are organized into *gestural scores* - the gestural molecules - which define the necessary tract variables and their respective descriptors for CL, CD etc. and, what is essential for AP, the activation intervals defined by their respective beginnings and ends. As mentioned, activation is binary and therefore can only take

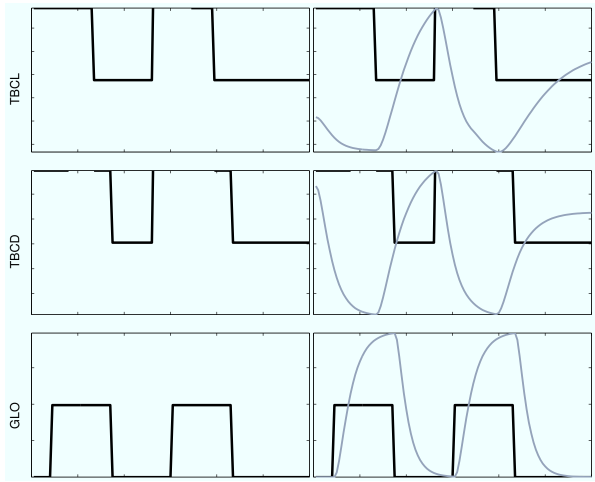


Figure 3.12.: Illustration of the functioning of gestural scores in pure *box notation* (left) and (right) with additional superimposed tract variable motion (adopted from Browman und Goldstein, 1989, p. 212).

"on" or "off" values. Further relevant information which can be taken from gestural scores is the relative phasing between gestures which corresponds to their relative overlap: In AP, gestures can overlap in time, and here the following scenarios have to be carefully kept apart: Overlapping gestures can be specified (i) in the domain of a common tract variable, which not only generates temporal but also spatial gestural overlap and (ii) in the domain of a *different* tract variables again generating contextual surface variability. The first possibility, gestural blending, was already displayed in the introductory part (see figure 1.7) and is of integral importance for this work dealing with dorsal articulation. The computational model accommodates the amount of blending with a separate quantitative variable specifying the amount of blending allowed. The second possibility is called *hiding*, because "with sufficient overlap, one gesture may completely obscure the other acoustically, rendering it inaudible" (Browman und Goldstein, 1989, p. 215).[20]

[19]Which lies in the domain of the computational task dynamics model. Together with the articulatory synthesizer by Rubin und Baer (1981) it builds a full-fledged articulatory synthesis system.

[20]Hiding of the tongue tip gesture for the alveolar stop in the utterance *perfect memory* belongs here.

A first example of gestural scores is given in figure 3.12. These scores come in three guises, (i) as pure *box notation* in which only gestural activation intervals are shown and (ii) showing gestural activation but with additional tract variable motion included, and (iii) in point notation, which is closest to traditional feature notation. Figure 3.12 demonstrates the notation of gestural scores for the utterance [əkakɐ]. The left panel shows trajectories of gestural activation that are roughly comparable to box notation - for for tongue body CL and CD and glottal aperture. In addition, the right panel superimposes the generated tract variable motion. The additional possibility of an illustration in point notation is omitted for brevity.

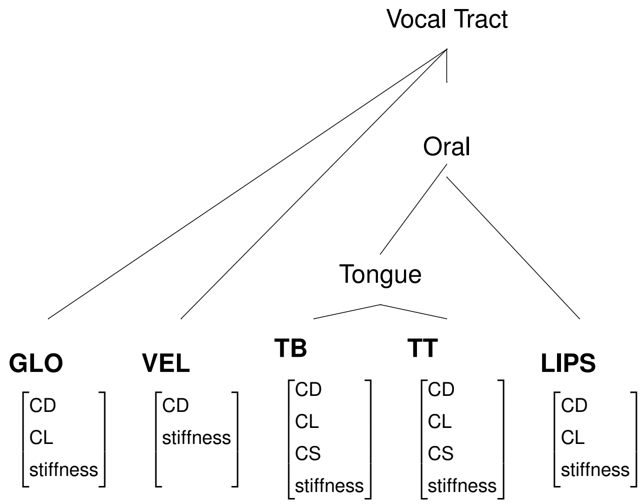


Figure 3.13.: Articulatory Feature Geometry tree, adopted from Browman und Goldstein, 1989, p. 210.

AP has built-in hierarchical structuring at different levels: First, tract variables are organized in different autonomous hierarchical tiers (i) the oral tier, (ii) the velic and (iii) the glottal tier. The oral tier is further differentiated into tongue and lips, the tongue in turn further subdivided into tongue body and tongue tip systems. This is actually the parallel with other Autosegmental Models, in particular feature geometry which was mentioned earlier. The display of the Articulatory Feature Geometry or the Functional Anatomy of the vocal tract shown in figure 3.13 does in fact make this point evident. A further essential point for AP is the functional division into two independent vocalic and consonantal tiers, again reminiscent of the standard feature-geometric topology already described. In accordance with Fowler (1980), consonantal gestures are superimposed on the vocalic cycle “in the background”. Further, the timing between vowels and consonants is established by phasing rules not described in greater detail here. It appears to be more important subsequent to this introductory sketch to have a closer look at the representation of the dorsal segments under consideration - palatal and velar obstruents.

5.2. The representation of dorsal obstruents

The conceptions of constriction location (CL) and constriction degree (CD) with their respective parameter values were already introduced in the preceding paragraph. While CD can roughly be identified as the gestural relative to manner of articulation specification, CL can likewise roughly be equated with the more traditional notion of place of articulation. It has to be added though that CL descriptors are conceived as multivalued, “since CL descriptors correspond to categorical ranges of the continuous dynamic parameters” (Browman und Goldstein, 1989, p. 228). Dynamic CL and CD parameters have to be specified for each of the constricting organs of the articulator set. Browman und Goldstein (1989, p. 227) suggest possible mappings between articulator sets and constriction locations for the purpose of speech synthesis within in the Task Dynamic Model. For the segments under consideration the relevant articulator sets are the tongue tip (TT) and the tongue body (TB). It seems noteworthy that the constriction location designated as "palatal" is nested under both the TT and the TB articulators. Still, it has to be added that AP views “CL as an independent cross-classifying descriptor dimension whose values cannot be hierarchically subsumed uniquely under particular articulator sets” (Browman und Goldstein, 1989, p. 228).

Browman und Goldstein (1989, p. 225) seem to be aware that “true” palatals are a kind of borderline articulation: “Palatal and palato-alveolar consonants are another type of articulation that falls between TT and TB articulations.” Anyway , Browman and Goldstein are critical against Keating’s formalization as complex segments, because “such a representation equates complex segments such as labio-velars, consisting of articulations of two separate articulators (lips and tongue), with palatals, arguably a single articulation of a single predorsal region of the tongue.” Given Browman und Goldstein’s multivalued conception of CL, they motivate palatal segments as follows:

> “For the categories that are further back (palatal and beyond), Wood (1982) hypothesizes that the distinct CLs emerge from the alignment of Stevens’ quantal considerations with the positioning possibilities allowed by tongue musculature. To the extent that this set of descriptor values is too limited, it can, again, be extended by combining descriptors, e.g. [palatal velar], or by using ’retracted’ and ’advanced’.” (Browman und Goldstein, 1989, p. 228)

Taken together, the representation of velars does not appear to be problematic. They are articulated by specifying TB tract variables, and available specifications of constriction locations are (i) uvular (ii) velar and (iii) possibly palatal. The finding of high velar contextual variability can be accommodated by an appropriate specification of blending. Concerning palatals, Browman und Goldstein (1989) reject Keating’s specification of palatals as complex segments due to the reasons described. This is warranted by separate devices for articulator sets and constriction locations which standard feature theory, - which lumps them into one set of place features - does not offer. Presumably, a differentiation between a fronted velar and a “true” palatal could be expressed if necessary at all, for example by furnishing a fronted velar as a TB articulation with a constriction location [velar palatal] and the “true” palatal as a TB articulation with the CL [palatal]. Further, the *constriction shape* variable could become important. Browman und

Goldstein (1989, p. 228) describe - at the time of this publication prospective - plans for an extension:

> "An additional TT tract variable (TTR) that specifies the orientation (angle) of the tongue tip in the sagittal place with respect to the CL and CD axes is currently incorporated into the task dynamic model." (Browman und Goldstein, 1989).

This variable is designed to produce apical/laminal differences, as well as to allow sublingual contact for retroflex articulation. A further extension proposed by the authors and relevant for the current work is the idea to use stiffness modulations to differentiate between palatal approximant and the high front vowel /i/: In the framework as described so far, both the palatal glide as well as the vowel would have to be described "a TB [narrow palatal] gesture" (Browman und Goldstein, 1989, p. 229). Browman und Goldstein (1989, p. 229) propose to use a higher value ([increased]) for stiffness k to differentiate between these articulations. It has to be seen whether these differentiations are strong enough to represent all possible palatal articulations.

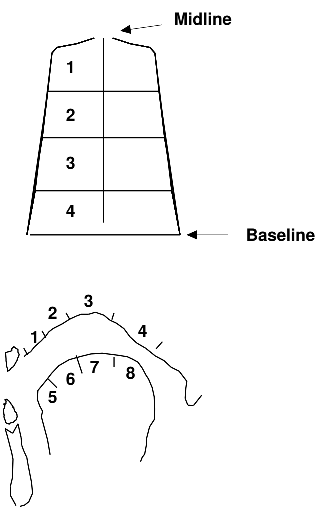
6. A more data-driven account

Already Keating und Lahiri (1993), while discussing the comparison between the different versions of English /k/s in e.g. *kin* vs. *kool* wondered whether the fronted palatal /k/ is to be phonologically represented differently with respect to place of articulation. They refer to SPE (Chomsky und Halle, 1968), according to which fronted velars are represented with all other stops produced at the hard palate by a single set of features.[21] In other words, it was not possible for SPE to distinguish between true palatals, fronted velars and palatalized velars, neither with respect to their underlying nor with respect to their surface form. I do not dare to answer the question whether AP at present can account for e.g. palatalized segments. More urgent in the current situation seems to get hold of a general picture of which palatal articulations are possible and observed in the world's languages and whether and how they differentiate on empirical grounds. The data analyzed in the phonological discourse has mainly utilized static X-Ray-images, linguo- and palatograms to determine a featural description of palatal articulation. In contrast, the research work which has studied palatal articulation from the more phonetically motivated viewpoint of lingual coarticulation has mainly relied on electropalatographic data, which has the merit of supplying time-varying data but only if there is a contact and with the drawback of limited spatial interpretability. Recasens and colleagues, in a large series of publications, have provided a taxonomy of palatal articulations while extending it towards a treatment with more up-to-date articulatory methods with better temporal resolution capable of measuring coarticulation phenomena. These coarticulation phenomena were quantified as a scale quantifying coarticulatory constraint (DAC). This work appears to be well-suited to complete the more linguistically shaped part of the introduction to the second experimental study and to crossover to the experimental work aimed at in this study.

[21] "Under this analysis, then, fronted velars, palatalised velars, and other palatals are all represented featurally as the same thing". (Keating und Lahiri, 1993, p. 74)

6.1. A phonetic taxonomy of palatal segments and their coarticulatory behaviour

Recasens (1990) proposed a revision of the phonetic characterization of palatal consonants with the main claim, that palatal consonants "involve a higher degree of articulatory precision than previously assumed" (Recasens, 1990, p. 267). Based on this claim Recasens (1990) elaborates a more data-driven taxonomy of palatal segments now involving alveolopalatals, front palatals, mid palatals and back palatals. This is at odds with the more traditional IPA conception including only a palatoalveolar zone for [ʃ,tʃ] and a palatal zone for sounds like [ɲ,c,ʎ,ç,j,ɟ]. The former are laminally produced at a postalveolar place of a articulation with some dorsopalatal constriction, the latter are assigned a dorsally produced place of articulation somewhere along the palatal region only. He provides a partitioning into palatal zones which is deemed necessary for the understanding of this taxonomy. It is reproduced in figure 3.14.



(1) alveolar zone
(2) prepalatal zone
(3) mediopalatal zone
(4) postpalatal zone
(5) laminal region
(6) predorsal region
(7) mediodorsal region
(8) postdorsal region

Figure 3.14.: Articulatory subdivisions for palatal articulations, after Recasens (1990, p.268)

Recasens (1990) provides classification for following segments within this framework.[22]

fricative [ʃ] and affricate [tʃ] are conceived as primarily lamino-postalveolar with some optional predorso-prepalatal contact. These sounds are articulated with the lamina, and Recasens (1990) refers to them as *lamino-postalveolars*.

[ɲ,ʎ,c] are of utmost interest in the present context: They are divided into front realizations and back realizations:

Front Realizations appear as primarily predorsally articulated with some involvement of the mediodorsum. Less frequently involvement of the lamina or the postdorsum may exist. Contact is made simultaneously at the postalveolar and prepalatal zones. Recasens (1990, p.271) stresses that there is one continuous constriction and thus

[22] The taxonomy has been extended and revised in several subsequent publications. For the present purpose the original Recasens (1990) formulations seem appropriate.

one place of articulation. The palatal nasal and the stop show very similar articulatory characteristics, although the oral stop shows more linguopalatal contact than the nasal stop. Recasens (1990, p.271) refers to them as *alveolopalatals*.

Back realizations Further, there are also back articulations, characterized as mediodorso-mediopalatal articulations. The major extension of this class are the fronted allophones of /k/ and some realizations of palatalized velars, and they usually involve additional contact or constriction along the prepalate and/or postpalate. These articulations are referred to as *mid palatals*.

Approximant [j] and fricative [ç] are mostly articulated with the pre- and mediodorsum at the pre- and mediopalate. They are referred to as *front palatals*.

On the background of this taxonomy, first results using aforementioned methods giving better temporal resolution were already mentioned in this publication (Recasens, 1990), but the database has been substantially broadened (e.g. Recasens und Espinosa, 2006; Recasens, 1997; Recasens; Farnetani; Fontdevila und Pallares, 1993) and it seems more suitable for the present purpose to review these publication together beginning with the degree of articulatory constraint (DAC)-scale.

6.2. The DAC

The DAC-scale is a development that arose from the taxonomy as just described in the previous section. Like AP, it refers to Fowler's version of coproduction: Fowler (1980) had argued against speech production theories in general which take phonological features as input. The features used as input for the speech production mechanism are timeless, abstract and static and have to be translated into articulatory movement. As Farnetani und Recasens (1999, p. 51) put it: "In this translation process, the speech plan supplies the spatial target and a central clock specifies when the articulators have to move." In contrast, Fowler's intention is to overcome this dichotomy and she suggests to modify the phonological units of the plan: The phonological units become dynamically specified phonetic gestures, with an *intrinsic temporal dimension*. The goal of the DAC (Degree of coarticulatory constraint)-scale attempts to characterize phonetic segments according to the types of articulatory constraints involved in their production. These values then can be used to predict the "coarticulatory resistance" of the segments. As in Fowler's theory, the DAC-model assumes that articulatory gestures associated with consecutive segments are coproduced and overlap to different degrees depending on their spatiotemporal properties, on prosodic factors and speech rate. According to DAC-scale,

> "consonants differ in DAC value according to the following order: dorsals (alveolopalatals, palatals, velars), lingual fricatives (/s/,/ʃ/), dark /l/, which can be assigned a maximum DAC value (DAC=3); dentals and alveolars such as /n/ and clear /l/ (DAC=2); and bilabials, with the lowest DAC value (DAC=1)[...] It is hypothesized that dorsal consonants are highly constrained based on the observation that their primary contact or constriction location stays relatively fixed in line with the large contact size involved and perhaps the sluggishness of the tongue dorsum. The

> same observation may even apply to velars provided that at least two targets in front and back vowel contexts are accounted for (Recasens, 2002, p. 2828)."

It should be added that vowels are assigned DAC values as well, with the palatal vowel /i/ assigned the highest value (DAC=3), an intermediate value for /a/ and /u/ (DAC=2) and the lowest value reserved for the schwa (DAC=1).

The motivation in terms of speech production is as follows: Recasens quotes Perkell's distinction between extrinsic muscles governing vowel gestures and intrinsic muscles governing consonantal constrictions. Recasens conjectures that segments with a high DAC - like palatals and velars - as well involve extrinsic tongue muscles. Consequently, the relation between vocalic and "extrinsic" and consonantal and "intrinsic" is not a perfect one and holds only for a subset of consonants. In other words, there is a slightly stronger emphasis on "articulators" and physiological parameters than in the gestural approach displayed in the previous section.

An elegant way to illustrate the functioning of this scale and its connection with the notion of tongue dorsum involvement could be a comparison of clear and dark /l/: For dark /l/, the tongue body has to be kept under stricter control to provide the secondary constriction. This should amount to a more context-independent articulation for dark /l/ in contrast to "clear" /l/. This finding was confirmed in a study comparing the realization of Catalan dark /l/ with the light /l/ as found in Standard German (Recasens; Fontdevila und Pallarès, 1994). For example, active predorsum lowering as required for the realization of dark /l/ prevented the coarticulation with the high front vowel. While the DAC-scale apparently is to be revised,[23] the basic observation of the tongue dorsum imposing severe restrictions on coarticulatory behaviour is retained. The core mechanisms is that the relatively clear assignment of extrinsic tongue musculature to vowel production and intrinsic tongue musculature to the production of consonants is blurred. This more physiological stance puts a slightly different emphasis on the notion of the *blending* concept than the one originally laid out in AP: AP offered a scalar value of the blending parameter in the context of the task-dynamic synthesis concept, the DAC uses a definition rooted in biomechanical coupling and recognizes (at least) two targets in front and back vowel context for the velars. But the question remains how "true" palatals which are not reasonably captured as alveopalatals, but rather have to be classified e.g. as "mid palatals", could be captured within the AP framework. Without promising an answer or even an attempt at a solution, it seems a good move to first have a look at more experimental evidence concerning the behaviour of palatals and velars.

6.3. More details on palatals and the palatal/velar relationship

Recasens und Espinosa (2006) seems to be a good entry point for such an undertaking. Especially some more explicit wording on blending and coarticulation notions make this paper of particular interest for this work. The first quote quickly characterizes the behaviour of alveolopalatals:

> "Alveolopalatals are highly resistant to vocalic effects at and behind the place of articulation. Moreover, the degree of coarticulation for those consonants varies in-

[23] Personal communication, D. Recasens. Presumably stimulated much by more recent work on prosodic factors and clusters.

> versely with dorsopalatal contact such that, e.g., /ɲ/ shows more dorsopalatal contact and less coarticulation than alveolopalatal /ʎ/." (Recasens und Espinosa, 2006, p. 299).

This behaviour is contrasted with the behaviour of for example the velar stops:

> "On the other hand, dorsovelars blend with, i.e., adapt or accommodate closure location to, the following vowel thus becoming postpalatal or medio-postpalatal before front vowels while staying velar before back vowels. The final stop outcome of this blending process may exhibit the same or a close place of articulation to that of the following vowel but also a large closure resulting from the addition of the closure and constriction areas for the two consecutive phonetic segments [...]. This blending mechanism is clearly in contrast with contextual coarticulatory variations which would not affect substantially the place of articulation for the consonant and could also occur in front or behind it." (Recasens und Espinosa, 2006, p. 299)

Recasens und Espinosa (2006, p. 299) motivate the decision between these two scenarios - i.e. the "blending scenario" and the "coarticulation scenario" for palatal stops (and nasals) - as central research question by referring to the Ibibio data from (Connell, 1991). A similar question is also relevant for own experimental work to be undertaken in the current work, so should also be mentioned: Data for pure palatal nasals in Ibibio indicate that palatals show large differences in vowel-dependent closure fronting, i.e. they are (i) produced with dorsal contact all over the palatal zone in high vowel (/i,u/) sequences and (ii) just the postpalate zone in low vowel sequence. This would be an instance of the blending scenario, and Recasens und Espinosa (2006) contrast it with the alternative coarticulation scenario as:

> "The alternative prediction is that large degrees of dorsopalatal contact for palatal consonants ought to cause them to become highly resistant to vowel coarticulation independently of their place of articulation (Recasens und Espinosa, 2006, p. 299)."

In other words, for velar stops, Recasens und Espinosa (2006) assume the "classical" blending scenario for the velar stops, whereas the decision for the palatals has to be made on empirical grounds, as Connell's data suggest a second, "mild" blending scenario for Ibibio palatal nasals.

Another interesting point discussed in Recasens und Espinosa (2006) is the phonemic status. Recasens und Espinosa (2006) investigate the behaviour of palatals in Majorcan Catalan. In this dialect the palatal [c] is the fronted allophone of the velar stop. In languages in which the palatal stop is phonemic it does not seem to require the "front" alveolopalatal closure like in Czech, but can also be implemented with the pure palatal closure as in Hungarian (Keating und Lahiri, 1993). For this latter class of languages, Recasens und Espinosa (2006) mention Icelandic and Ngwo apart from Hungarian. For these languages, Recasens und Espinosa (2006) pose one suitable research question targeting the palatal:

> "It deserves to be seen whether the need to differentiate two phonemes, i.e., /c/ and /k/, causes [c] to exhibit less variability in place of articulation in these languages than in those where [c] is an allophone of /k/. (Recasens und Espinosa, 2006, p. 297)"

Recasens und Espinosa (2006) reflect the relationship between place of articulation concerning the palatal stop and nasal in a similar fashion: “A more anterior articulation for [ɲ] than for [c] cannot possibly be attributed to the need to avoid confusion between [ɲ] and [ŋ] since the velar nasal is allophonic in these languages and dialects” (Recasens und Espinosa, 2006, p. 297).

Similar considerations reflecting the phonemic status and (co-)articulatory behaviour were highly relevant for the design of own experimental work to be presented in the current work, but are put off for the actual derivation of hypotheses. Instead, the next section reviews the very few articulatory studies on the German palatal fricatives known to me.

6.4. The German palatal fricative: Articulatory studies

The only articulatory studies I am aware of concerning German dorsal fricatives are Ambrazaitis und John (2004) and Pompino-Marschall und Mooshammer (1997). This section will quickly recapitulate their results. Pompino-Marschall und Mooshammer (1997) acquired EMA and EPG data on these sounds.[24] The authors interpret the surfacing of the palatal fricative after front vowels as “due to a mandatory phonological process of an all-or-nothing nature” (Pompino-Marschall und Mooshammer, 1997, p. 378). In contrast, the variation in the realizations of the dorsal fricatives is seen as “due to optional and gradual coarticulatory phonetic processes”. They also support the pattern of the fronting of the velar fricative before front vowels, this process is seen as a “regressive assimilatory force [...] being due to optional and gradual coarticulatory processes.” (Pompino-Marschall und Mooshammer, 1997, p. 378)

Ambrazaitis und John (2004) also performed an EPG study similar in spirit to the one just introduced on the German dorsal fricatives, likewise dwelling on the asymmetry between the German fricative allophony and the phonemics of the (voiceless) velar stop. They quote Wiese:

> “It may be the case that the assimilation of the dorsal fricative is categorical, while the dorsal stops (i.e. /g/ and /k/) assimilate in a more gradual and variable manner. If this difference could be verified phonetically, it would provide good evidence for treating DFA [= dorsal fricative assimilation] as a phonological, and the assimilation of /k/ as a phonetic process.” (Wiese, 1996, after Ambrazaitis und John, 2004, p. 2)

The distinction between the gradual process for the stops and the allophonic variation for the fricatives is hypothesized to be “manifested in production through a larger articulatory distance between the most anterior and the most posterior allophone for /x/ than for /k/”(Ambrazaitis und John, 2004, p.1). The main analysis strategies applied by Ambrazaitis und John (2004) consists in an operationalization of such a distance measure by transforming the Centre of Gravity measures into Mahalanobis distances which in turn allows a treatment of these distances by analysis of variance methods. The most striking result is that the existence of the different allophonic categories [ç,x,χ] as advocated by Kohler could not be replicated, most of the speakers merged the velar with the uvular realizations. This result might have been through the sole reliance on the use of palatography, which does not record any uvular contact, and even the resolution in the velar region might be problematic. This point is weakened to some extent, because the only South German speaker in the study by Ambrazaitis und John (2004) realized uvular variants

[24] The data mainly are analyzed with respect to the EPG centres of gravity and amount of contact

identifiable by means of EPG, although not fully consistent with the predictions as made by Kohler (1995).

Of more relevance for my own work, and at the same time casting the discussion in direction of the topic of the next section, Ambrazaitis und John (2004, p.3f) raise concerns against applicability of the notion of assimilation as applied by Wiese (1996) above, because “no words exist that can contain either, e.g. a palatal, or a velar realization of /x/, depending on assimilation factors such as speech rate or style.” Instead, they argue that the question of a difference between the allophonically conditioned behaviour of the fricative and the single velar stop phoneme should not be cast in terms of assimilation at all. They rather propose a division in terms of articulatory control (“Steuerung”) versus coarticulation adopted from the early work by Menzerath und de Lacerda (1933). The association intended is obviously that the presence of several allophones has to be represented “more centrally” while the variable realization of the velar stops is attributed to the phonetic periphery. A discussion of this topic in moderate depth is the goal of the next section.

6.5. Coarticulation and control

In the post-Chomsky era, the mainstream opinion consisted in conceptualizing coarticulation phenomena as reflexes of neuromuscular movement production schemes: This implied the distinction between assimilations and the - unavoidable - presence of transitions between linguistic units like first conceptualized by Menzerath und de Lacerda (1933). This claim was first challenged the by demonstration of transconsonantal effects (e.g. Öhman, 1966). Some more recent work (e.g. Abbs und Connor, 1991) reasons closer to the neural substrate of motor planning, arguing that kinematic patterns reflect underlying linguistic representations, but are subject to sensorimotor adjustments transforming these higher level representations and making them difficult to determine from the peripheral speech output:

As the common denominator for distinguishing between mental (or cognitive) on the one hand and “peripheral” presumably is the initiation in the cerebral cortex.[25] Formerly it was believed that the cortex was exclusively devoted to mental or cognitive processes. Motor processes in contrast were conceived as one-way down to the subcortical implementing system. This highly modular and rigid division these days is not upheld any more, rather most researchers recognize that portions of the frontal and parietal lobes interact with subcortical structures in motor production. Still, cortical areas for mental/cognitive processes connected to speech planning and assigning slots in temporal organization lie upstream from cortical areas which are responsible for motor production. Broca’s area is relevant here, although Broca’s area is known to be multifunctional, with the syntagmatic synthesis of utterances and speech articulator programming being the crucial aspects in this place, but only one responsibility of this area.

One level further downstream, where contributions of different articulators would expectedly to be merged, the Supplementary Motor Area (SMA), a part of the frontal cortex, programs movement complexes together with the cerebellum. The subcortical motor systems are not known to advance or delay the initiation of speech movements. The SMA triggers the precentral motor cortex via the basal ganglia which are not involved in feedback-control of on-

[25] This paragraph to a large extent adopted from Wood (1996).

going movement and apparently cannot change the movement program. The cerebellum has a modifying function and is responsive to proprioception of movement progression. Its function though appears to be refined to assure the reaching of motor targets rather than actually changing them. Finally, the motor plan reaches the peripheral articulatory biomechanics which was already briefly described in the introductory part of this work.

The sketching of motor pathways in the last paragraph was intended to remind that the production of speech is a complex and highly interactive multilevel process. Given the complexity of the embodiment of linguistic structure, will it be meaningfull or at all be possible to decide between levels of implementation in the sense of Menzerath und de Lacerda (1933)? Or will it be necessary to devise a different rationale for the division between phonology and phonetics? Put differently, and closer to the purpose of this work, is it useful to distinguish between phonetic coarticulation and phonological assimilation in a strict sense? Here it seems worthwhile to quote from the theoretical karate Fowler (1983) applies against the criticism Hammarberg (1982) raised against her theoretical perspective. Fowler (1983) actually aims at maintaining the division between phonetics and phonology. The reason to keep coarticulation apart from phonological assimilation is justified by recurring to vowel harmony: If one would collapse coarticulation and assimilation, critical differences would be obscured: Vowel harmony cannot be rooted in natural phonetic processes like coarticulation, because it cannot offer an answer to the question why it occurs for example in Hungarian but not in English. Here, she uses a line of argumentation inherited by Anderson (1981) who argues that phonological processes are "just what is arbitrary about systematic processes involving phonological segments. That is, phonological processes are just what cannot be exhaustively explained by invoking, for example, articulatory dispositions." (Fowler, 1983, p. 314) The same statement, repeated in Anderson's own wording closes the theoretical sections, and illustrates the strategies pursued in the following experimental studies:

> "On this view, it is still very much part of the business of phonologists to look for "phonetic explanations" of phonological phenomena, but not in order to justify the traditional hope that all phenomena of interest can be exhaustively reduced this way. Rather, [...] the reason is to determine what sorts of facts the linguistic system proper is *not* responsible for: to isolate the core of features whose arbitrariness from other points of view makes them a secure basis for assessing the properties of a language faculty itself." (Anderson, 1981, p. 497, after Fowler,1983, p. 314)

7. Wrapping up: The derivation of hypotheses

As should have already become evident from the theoretical introduction, the experimental studies aim at a crosslinguistic comparison of dorsal obstruent articulation in German and Hungarian. Both languages have velar stops in the inventory, which function as the common denominator for the experimental design in some sense. There are three logically possible contrasts derivable from such a layout: The intra-language comparisons for (i) German and (ii) Hungarian, and, additionally, (iii) comparisons between velar stops for both languages. The following sections will summarize experimental predictions for all three of these contrasts, in the order listed.

7.1. Intralinguistic: Hungarian

According to Keating, palatals are complex segments with both a dorsal and a coronal specification. In contrast velars - at least in her examples showing English /VCV/ sequences - are handled by underspecification for backness. In line with her window approach, this would predict relatively little sensitivity to vowel-induced contextual coarticulation for palatals in contrast to velars. At the same time, front velars should be distinct from palatals in their shape configurations, velars lacking the /i/-like "component". If the surface underspecification Keating adopts for English is also valid for Hungarian velars, then the velars should exhibit relatively large vowel-dependence in place of articulation. On the other hand, a scenario on which the underspecification of backness would be abandoned for the velars would also be possible. This should lead to reduced coarticulatory variability of the velars in this language. Furthermore, repeating her statements on the phonetic implementation on contextual velar fronting (see already section 4) seems in place: "Stated another way, velar fronting is something that happens *gradually over the course of the velar*". As mentioned, Keating seems to view this as a *"transparency effect of the velar with respect to Backness"* (Keating, 1993, p. 17). It should be added that this implies that this transparency effect should be absent *over the the course of the palatal*. Putting this together, according to Keating, velars and palatals should be distinguished by the amount of coarticulatorily induced variability and the articulatory behavior *during* these stops. An operational definition of coarticulatorily induced variability will be devised in the method section (see section 8).

Recasens elaborated a more fine-grained description of palatal articulations which distinguishes at least between *alveolopalatals* and *mid palatals* for the stops. In a previous section, his stronger data-driven taxonomy of palatal segments was introduced. According to this, the first task will be to arrive at a judgment on the palatal stops of Hungarian. In its first version (Recasens, 1990), it distinguished between *alveolopalatals* and *mid palatals* for the stops. So, the first major point of interest is to arrive at a data-based description of these segments. On the basis of observations made by Keating und Lahiri (1993), it does not seem unreasonable to expect a classification different from the alveolopalatal one for the Hungarian stops. Depending on these results, coarticulatory behavior subject to different vocalic environments will be focused. More explicitly, the question will be whether the coarticulation scenario for alveolopalatals or the blending scenario for velars will be more appropriate for the description of these sounds. Recall that data for pure palatal nasals in Ibibio indicate that palatals exhibit large differences in vowel-dependent closure fronting, i.e. a "mild blending" scenario. Note the methodological differences though: Most studies on palatal articulations relied on palatography, which might lead to difficulties in comparing the results of the current study with those found in the literature. Further it will be interesting to see whether the fronting of velars in the context of the palatal vowel /i/ leads to a convergence with the tongue shape of palatal stops.

7.2. Intralinguistic: German

Concerning the German stop versus fricative, the - operational - hypothesis as put forward by Ambrazaitis und John (2004) will be adopted, according to which the articulatory distance between the most anterior phonemes and the most posterior phonemes should be larger for the fricative allophones than for velar stop coarticulation. These authors already mentioned that the use of EPG is presumably suboptimal for providing answers to their question. In this sense, the work here which will be based on articulatory movement data acquired by EMA constitutes a validation of their data by a second method.

In my view - given positive results - this does not allow to infer different articulatory strategies from these data, e.g. in the spirit of Menzerath und de Lacerda's *Koartikulation und Steuerung*, but in the first place provide descriptive evidence only. Therefore I would content myself with the tenet often found in the literature that phonemic and allophonic variation should be distinguishable on the phonetic surface (e.g. Keating und Lahiri, 1993) - and not attempt to make inferences about particular motor control regime even on the basis of observations conforming with such a view. The reason for this is not a naive agnostic view on articulatory behaviour, rather alternative explanations exist which are straightforwardly conceivable, for example in terms of "articulatory economy" it could be that the velar fricative is produced "more comfortably" in back vowel contexts at the soft palate than the velar stop: it could be easier to make a constriction at the soft palate than a completely sealed closure. This could account for why [x] may become uvular while the back stop allophone of /k/ never does.[26] Presumably such a hypothesis can also not be tested with movement data as those acquired for the current work. However, any viable alternative explanation in my view invalidates inferences about central nervous processes to be attributed to phonemic or allophonic phonological patterning a priori.

7.3. Crosslinguistic comparison: Hungarian vs. German

This is different for the crosslinguistic hypotheses concerning the behaviour of velars in Hungarian and German. The leading question is whether the existence of a palatal phoneme in the inventory - like in Hungarian - influences the variational patterns of the velar stops. In Keating's terms this surely could be captured by a deviation from the velar underspecification as described for English velars. More interesting though are genuine phonetic questions: Consider a less palatal realization of the velar in front vowel context for Hungarian: This would indicate that the speaker is not committed to contextual velar fronting due to biomechanical reasons alone and therefore indeed provides evidence on possible motor control regimes in the sense that a purely peripheral explanation for velar contextual variation can safely be ruled out. The following method section will display the way chosen to answer this set of questions.

[26] Actually a suggestion by D. Recasens, personal communication.

8. Method

The first issue is to display the basic experimental design. This involves a priori decisions about vocalic environments and solutions how to deal with positional restrictions, in particular on the German dorsal fricatives:

For Hungarian, it seemed reasonable to acquire both palatal and velar stops in initial position, while for the German dorsal fricatives, this was not attempted, because the probability of such a procedure leading to inconsistent realization was deemed too high: German dorsal fricatives occur word-initially only in loanwords, and it was unclear whether (and how) different speakers - with different dialects - would have assimilated them. Therefore, it was found acceptable to acquire the German fricatives in medial position only. Therefore initial German fricatives were replaced by velar stops in order to keep coarticulatory influences exerted by initial consonant identity constant.

A second point of concern is the question how to equate German and Hungarian vowel systems in order to establish the crosslinguistic comparison. According to the IPA-handbook (1999), the short Hungarian vowel corresponding to long /aː/ is said to be "higher and backer" than its long counterpart. It was decided to equate it with the vocalized /r/ of German in post-tonic positions. This seems justifiable because word-level stress is nondistinctive in Hungarian and fixed on the morpheme-initial syllable.

	C1	V1	C2	V2
Hungarian	[c]	[iau] x length	[c]	[ɑ]
Hungarian	[k]	[iau] x length	[k]	[ɑ]
German	[k]	[iau] x tenseness	[ç,x,χ]	[ɐ]
German	[k]	[iau] x tenseness	[k]	[ɐ]

Table 3.6.: Asymmetries of the experimental design: German palatal fricatives have no regular occurrence word-initially, and was replaced with the velar stop, short Hungarian /ɑ/ is equated with with German vocalized /r/.

This amounts to the a priori decisions about the speech material to be acquired which is displayed in tabular form in table 3.6. Note that this listing only displays voiceless obstruents. Further voiced segments were recorded as well, but skipped in the display, because at the moment only the unavoidable asymmetries of the crosslinguistic approach are intended to be highlighted. Starting from this basic design, the next section will introduce further details about experimental data acquisition.

8.1. Articulatory data acquisition

Hungarian

Tongue, jaw and lower lip movements of two female and two male speakers of Hungarian were recorded by means of the magnetometer. Four speakers were acquired by means of Electromagnetic Midsagittal Articulography (EMMA, AG100, Carstens). Four sensors were attached to the tongue, one as far back as possible (TB), one approximately 1 cm behind the tongue tip (TT). The two middle sensors, tongue dorsum (TD) and tongue mid (TM) were located at equidistant points between them. Additional sensors were glued on the vermilion border of the lower lip (LLIP) and on the lower incisors (JAW). For one of the speakers, there was no additional sensor on the lower lip, only on the jaw. Two sensors on the nasion and on the upper incisors served as reference for compensation of head movements relative to the helmet and definition of an intermediate coordinate system. The final coordinate system was defined by recordings of two sensors on a T-bar acquired in order to rotate the data to the occlusion plane for each speaker individually (Hoole, 1996). After the speech material was recorded, one of the tongue sensors - usually the tongue tip - was removed from the tongue. This sensor was then used to record the contour of the palate in a separate session by moving the sensor along the palate in back-to-front direction. These tracings were later aligned by visual inspection with the tongue (applying vertical and horizontal translations and rotations). Original sample frequencies were 400 Hz for EMMA data and 48 kHz for the audio signal. For the analysis, the EMMA signals were low-pass filtered and downsampled to 200 Hz. The material consisted of /CVCa/ nonsense words with either velar or palatal voiceless and voiced stops as consonantal context and one of the long or short corner vowels /i,a,u/. Two realizations of the test words were embedded in the carrier sentence "Most a ... es a ... volt" ("This was a ... and a ... now") and repeated between six and ten times by the four Hungarian study participants.

German

Likewise, two female and two male speakers of German were recorded. Data acquisition and processing was essentially equivalent to the Hungarian corpus. Essential differences in the materials concerning consonantal material were already mentioned at the beginning of the method section, i.e. the initial consonant was restricted to be the velar stop. The material consisted of /C_1VC_2ɐ/ nonsense material. As mentioned, the place of articulation of C_1 was always velar and comprised both voiceless and voiced stops. If initial C_1 was the voiced velar stop, medial C_2 was also voiced. If initial C_1 was the voiceless velar stop, then medial C_2 could be the voiceless stop but also the fricative. As an illustration, consider possible testwords in /i/-contexts: [gigɐ,kikɐ,kiçɐ]. Vocalic context was fixed to one of the tense or lax corner vowels /i,a,u/. Two realizations of the test words were embedded in the carrier sentence "Ich habe ... ohne ... erwähnt" ("I mentioned ... without ... "). Each item was repeated 10 times.

8.2. Segmentation criteria

Segmentation criteria were kept alike for both languages, therefore their description will not be split by language: After automatically generating phonemic transcriptions of the speech signals by means of the Munich Automatic Segmentation System MAUS (Kipp et al., 1996; Schiel, 1999), additional manual segmentation and labelling of the speech data of all speakers was carried out on the basis of waveform, spectrogram and auditory impression with the software package PRAAT (Boersma und Weenink, 1992–2007). Apart from the offset of the trailing vowels of the carrier phrase, the following temporal landmarks were extracted (see figure 3.15):

a) the burst (or frication) onset of the initial consonant

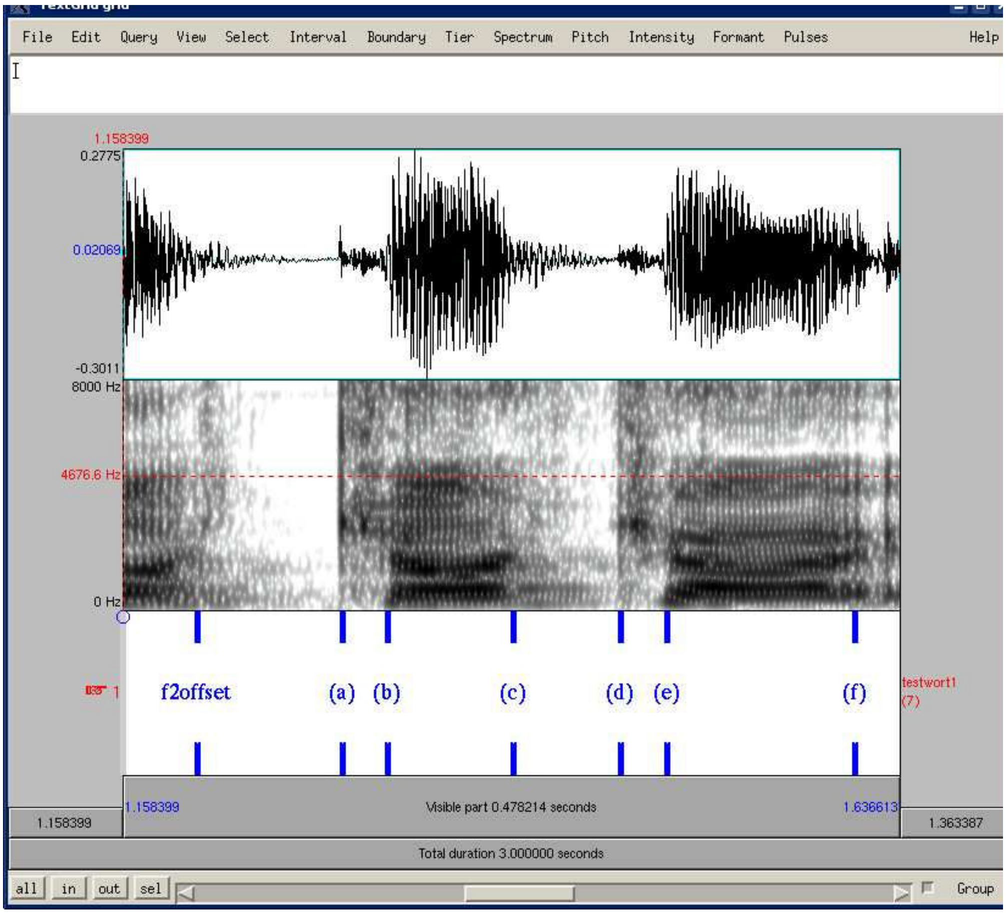


Figure 3.15.: Illustration of segmentation criteria. For a more detailed description see text. The example sound is a voiceless Hungarian palatal in the context of /iː/.

b) the onset of the second formant of the first vowel

c) the offset of the second formant of the first vowel

d) the burst (or frication) onset of the the medial consonant

e) the onset of the second formant of the second vowel and

f) the offset of the second formant of the second vowel

8.3. Analyses

This section describes the methods used in the subsequent analyses. Again, the division between the languages Hungarian and German and also between the different research question is collapsed. Occasionally, analysis methods reported refer to a particular research question. This will be indicated.

Acoustic analyses: formant tracks

After manual segmentation and labelling of the speech signals as described in section 8.2, subsequent formant analyzes were performed by means of the EMU system for speech database analysis (Cassidy und Harrington, 2001) which in turn incorporates the formant estimation tool "forest". Forest obtains raw resonance frequency and bandwidth values by root-solving of the Linear Prediction polynomial from the autocorrelation method and the Split-Levinson algorithm (SLA). It then classifies resonances as formants using the Pisarenko frequencies (by-product of the SLA) and a formant frequency range table derived from the nominal F1 frequency. The algorithm was specified to use the default nominal F1 value of 500 Hz for male speakers and a value of 600Hz for the female speakers.

Vowel formants estimates were determined at the temporal midpoint of the vowel using the default Blackman window. In cases of implausible formant estimates (which most often was the case for /u/ and /u:/) the algorithm was either reparametrized to use a different (lower) nominal first formant or the number of coefficients of the Linear Prediction polynomial was increased by one. In cases where no reasonable formant estimate could be derived, the value was discarded from the analysis presented. Overall, 1.8% of the cases had to be excluded.

Sensor distances traveled

The analysis of distances certain dorsal sensors travel during different temporally defined intervals of the production of the (C)VCV-sequences is intended to be informative about the velar transparency effect in Keating's terms about sliding movements of the tongue during oral closure, i.e. loop trajectories.

For the analyzes of the distances reported below, four different time intervals were defined. The analyzes target the medial consonants of the /(C)VCV/ sequences:

(i) the movement from the midpoint of the first vowel to the onset of closure, where the midpoint of the vowel is defined as the central sample of the temporal landmarks of (b) and (c) as defined in the previous section (8.2),

(ii) the interval during closure defined as the time between (c) and (d),

(iii) the interval between the burst and voicing onset as the time between (d) and (e), and

(iv) the interval during the second vowel, where the vowel mid of the second vowel is defined in analogy to the interval of the first vowel.

These intervals serve as the basis for the *articulatory* analyzes reported below. In my opinion, the use of acoustic landmarks is appropriate, because the commonly applied definition of articulatory landmarks from the speed signal was not consistently possible in all vowel environments.

The distances traveled during the intervals defined above can be calculated by summing the tangential velocities which in turn is calculated as $v = \sqrt{(\dot{x}^2 + \dot{y}^2)}$, i.e. the square root of of the sum of the squared velocities of the x- and y-components of the movement.

Tongue postures

Postural displays at various instances in time are informative for answering the research questions under consideration. They will be supplemented by standard 2-σ-dispersion plots in most instances. Further, articulatory postures at different phases of the articulatory trajectories will be subjected to speaker-independent factor analyzes which hopefully will provide interesting cross-linguistic projections. These are “projective” methods for “putting someone’s tongue in someone else’s mouth”, and particularly suited - respectively the only solution that I am aware of - for the cross-linguistic comparison of tongue shapes. These methods are non-standard and therefore will be quickly reviewed in the results section on the crosslinguistic aspects of the current work (section 11). This move is not ad hoc but rather motivated by the need for a special type of display particularly suited for the crosslinguistic analyzes aimed at.

9. Results I: Hungarian

9.1. Preliminary qualitative evaluation on palatal stop realizations in Hungarian

As already discussed, there has been a long-standing debate whether the palatal obstruent in Hungarian is better classed as a stop or an affricate. Without making detailed analyzes or raise this issue to a central topic, realizations of the four speakers presented here exhibit the following pattern: During the stop interval, no full silence was achieved but the whole interval was accompanied by frication. Additionally there was a portion of the signal resembling a (residual) release burst. This was followed by a second frication portion with a change in the spectral energy distribution. One further speaker has been recorded without the magnetometer, whose data will not be analyzed. Preliminary inspection gave evidence of a clear palatal stop realization. Since for the former speakers frication during the closure phase only occurred for the palatal but not for the velar stop it is assumed that it is not an artifact of the recording procedure but rather a speaker-dependent allophonic variation. More thorough and detailed spectral analyzes would be needed before a conclusive categorization of the observed patterns would be possible.

9.2. Vowel formants

The following section aims to discover the consonantal effects of palatal and velar stops in the /CVCa/ sequences on the vowel nuclei. Unlike in other analyzes to follow, the analyzes reported here pertain to both voiced and voiceless consonants. Before starting with the evaluation of the observed patterns, it seems in place to repeat some basic properties of the vowel system of Hungarian. The IPA-handbook (1999) describes the vowel system of Hungarian as a quantitative system, with seven vowels occurring in distinctively long and short quantities. One of the peculiarities resisting this general pattern is the short vowel corresponding to long /a:/, which is said to be higher and backer than its long counterpart. Therefore, it is described as [ɑ] in the handbook, although it is also hypothesized that is has some additional rounding making a transcription as [ɒ] a reasonable alternative to this decision. A second peculiarity not considered in further detail here is that mid front unrounded long [e:] is considerably higher and more peripheral than its low mid front unrounded shorter counterpart [ɛ]. All other vowels are analyzed as possessing pure quantity opposition by the IPA. Most importantly, /i,i:/ and /u,u:/ belong to this class with only a quantitative opposition. Vowel plots are displayed in figures 3.16 to 3.19 using 95% ellipse plots.

Both ordinates and abscissae in these plots contain the formant frequencies in Hz. Axis ordering is reversed according to convention in order to facilitate interpretation. Furthermore, axis limits are chosen for each speaker separately to accomodate to individual vowel space sizes. The top panels in each of these plots show the first two formant frequencies of vowel in voiceless (left) and voiced velar stops contexts, the bottom panels equivalently for the palatals. The observed patterns are astonishingly consistent over speakers and show no obvious quality differences between long and short [i,iː] and [u,uː] for velar consonants. Likewise the reported "higher" and "backer" pattern for the short low vowel - as described above referring to the IPA description - is evident. This pattern changes quite drastically for the palatals: While the high front vowel is still well characterized by the absence of a quality difference between long and short vowels, such a quality difference is introduced in the high rounded back vowels, with short [u] acquiring an almost central quality. A further point is the observation that the short [u] in palatal context is highly variable, a pattern which is as well consistent over the four speakers. This effect of the palatal context shifting the vowel quality of short vowels forward effectively raising F2 is also observable for the the /aː,a/ opposition. This effect can, in the extreme case, almost completely neutralize the quality difference between the long and the short versions of the low vowels, as after the voiced palatal stop for speaker km1 in the right bottom panel of figure 3.17.

In order to take a closer look at these patterns, mean F2 values and corresponding standard deviations are also shown supplementing the plots of the formant ellipses just displayed. These are shown in figure 3.20. The increased variability of the second formant for short [u] becomes evident, the effect of raising of the second formants for [a] is also visible, but the patterns for the high front vowels are not easily interpretable by these graphs. In order to achieve a more formal evaluation, repeated measurement ANOVAS with the second formant frequencies as the dependent variable and the factors place of articulation and vowel length were carried out separately for each of the basic vowel qualities [i,a,u]. Analyzes were carried out using the software package R (Cribari-Neto und Zarkos, 1999). For the context of the high front vowel, only the length

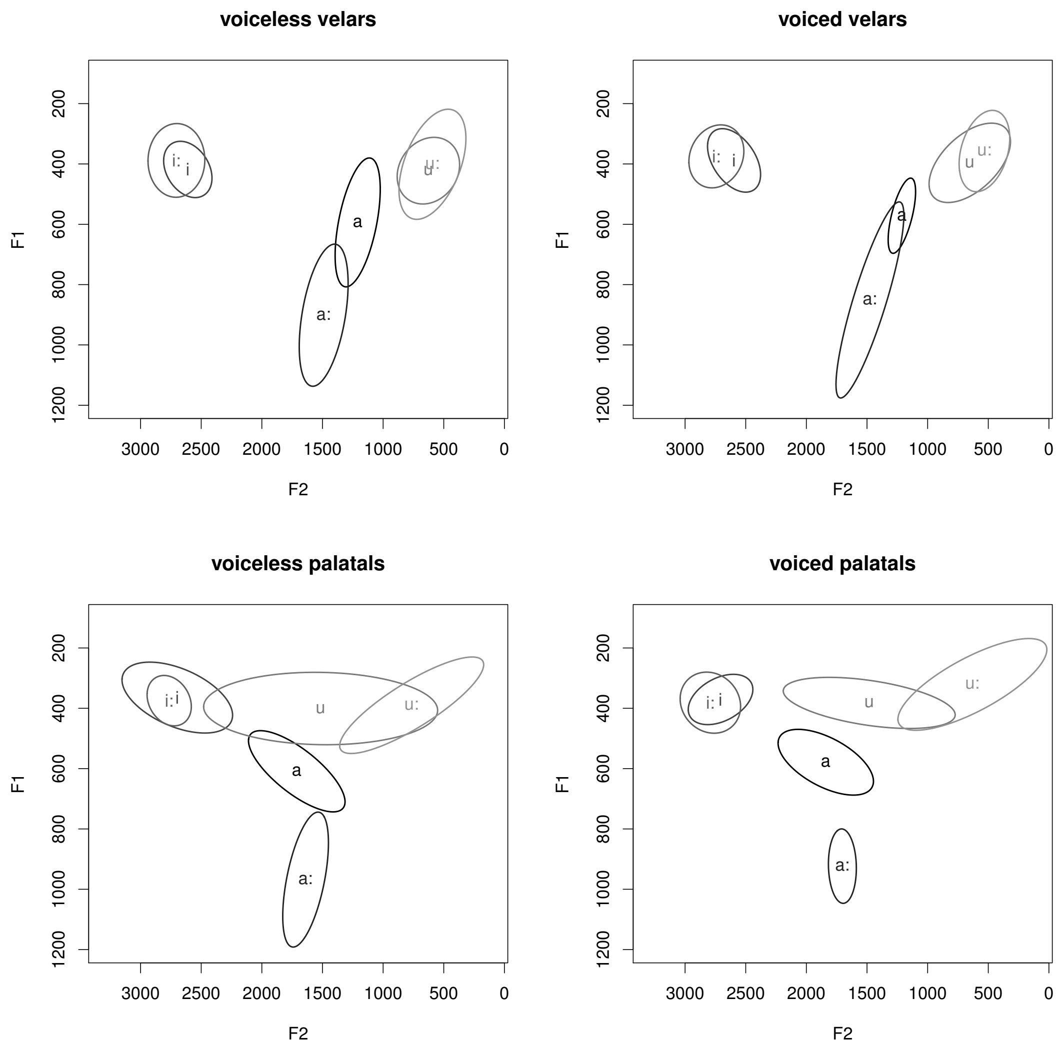


Figure 3.16.: The first two formants at the temporal midpoint of the corner vowels for the Hungarian data set as 95% ellipses: This graph shows the data for speaker ap.

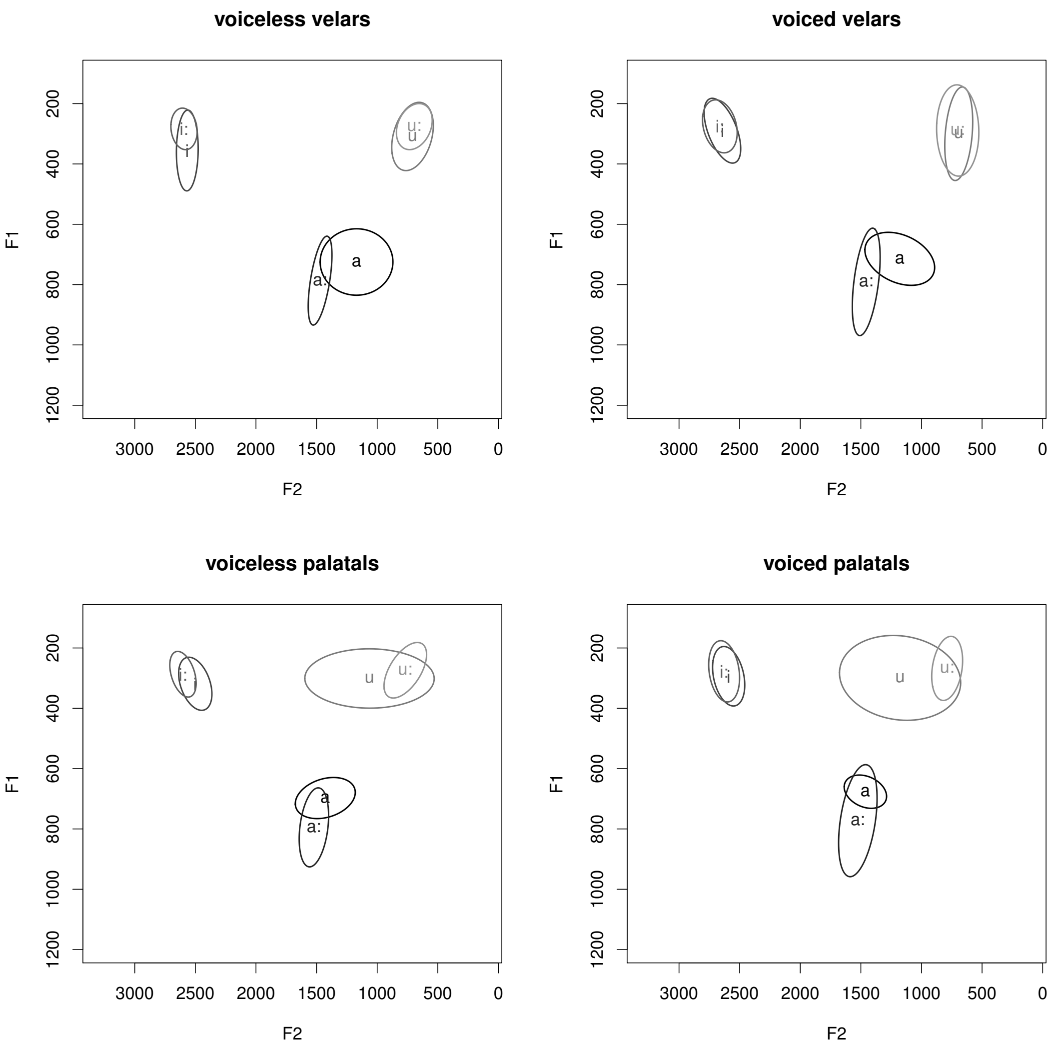


Figure 3.17.: The first two formants at the temporal midpoint of the corner vowels for the Hungarian data set as 95% ellipses: This graph shows the data for speaker km1.

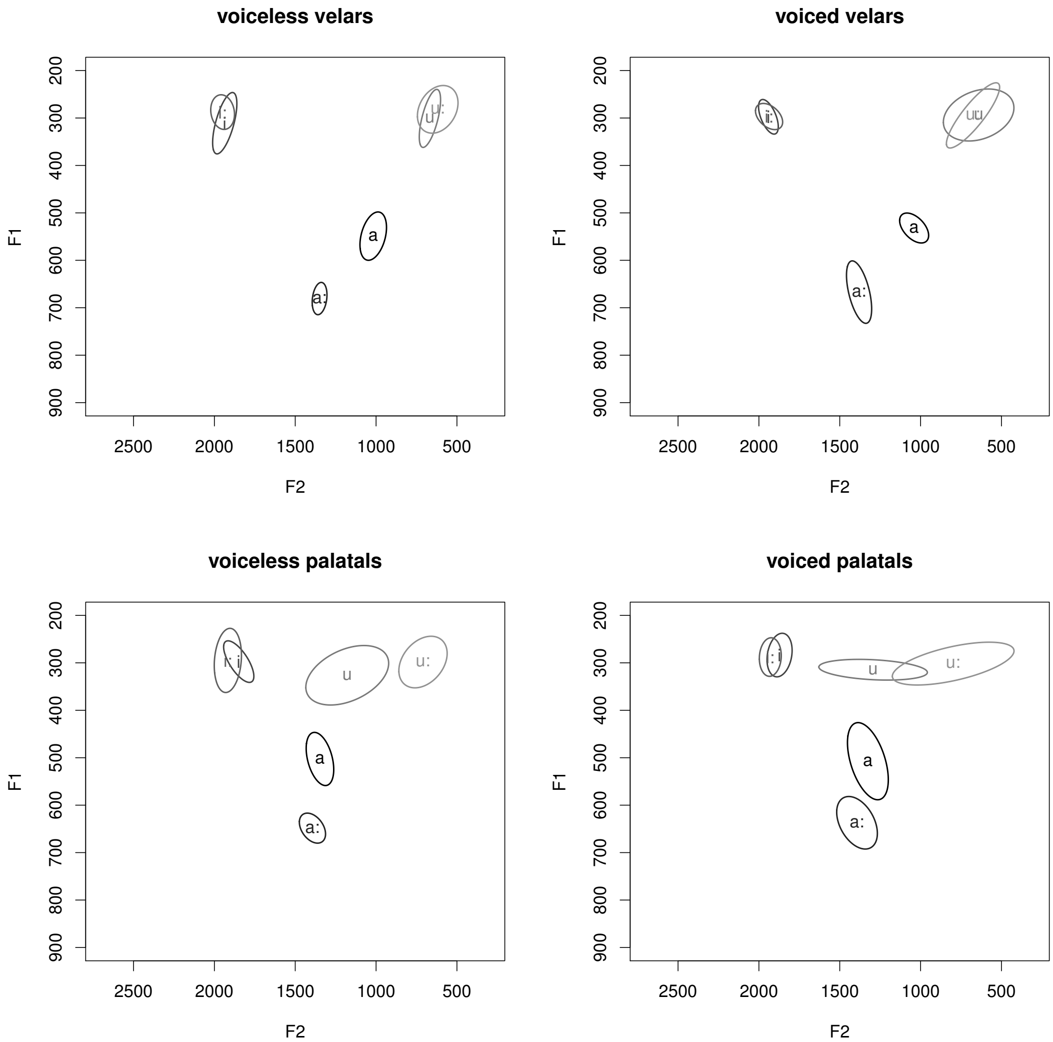


Figure 3.18.: The first two formants at the temporal midpoint of the corner vowels for the Hungarian data set as 95% ellipses: This graph shows the data for speaker lt.

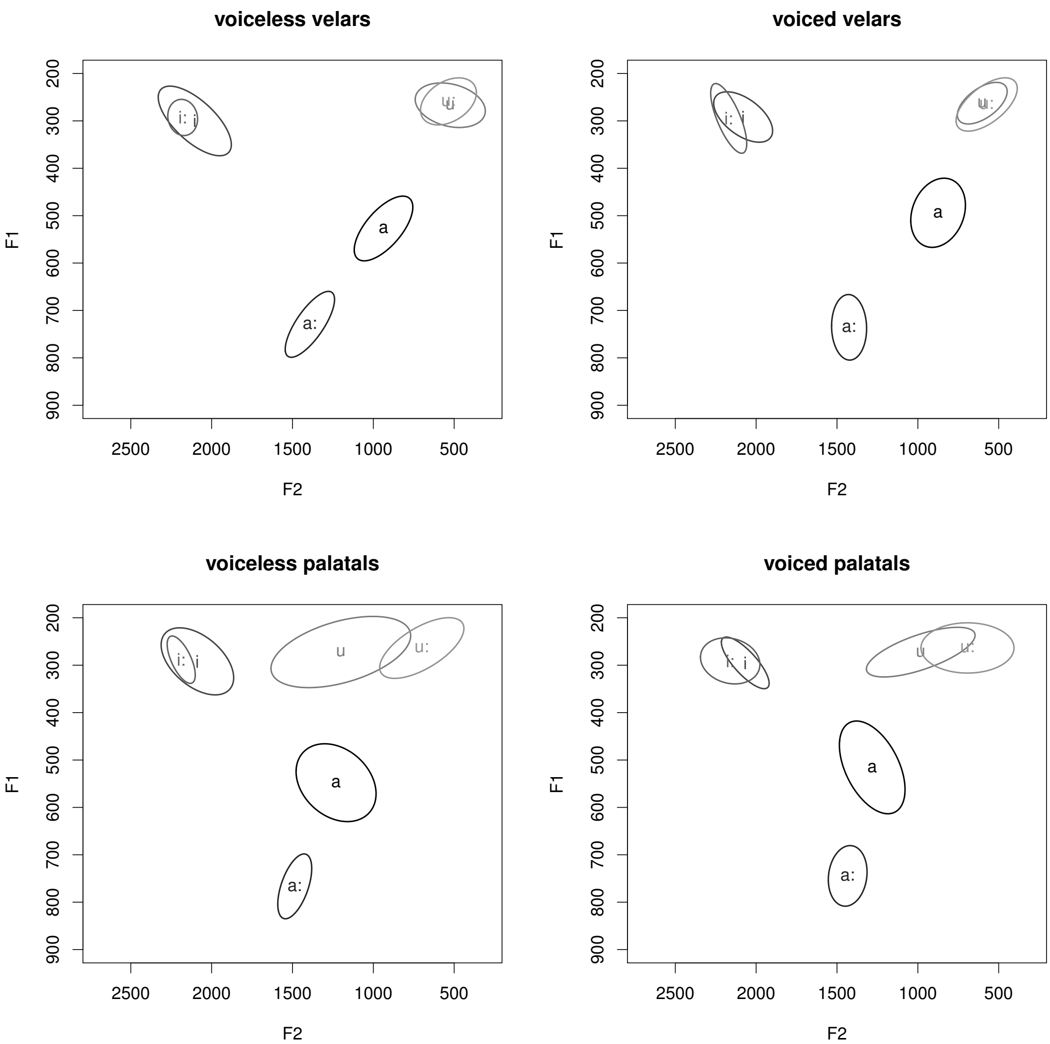


Figure 3.19.: The first two formants at the temporal midpoint of the corner vowels for the Hungarian data set as 95% ellipses: This graph shows the data for speaker rn.

variable is significant (p<0.05), as shown in table 3.7. In fact, for some speakers, the short version exhibits a centralization tendency. These are speaker-specificities though, e.g. km1 rather shows a higher first formant for short [i] rather than any change in the second formant. Therefore this pattern apparently does not warrant any consistent information and is rather displayed for the completeness of the formant analysis. For the [u,uː] contexts, all three effects are significant, the place main effect (p<0.01), the length main effect (p<0.05) as well as the the interaction (p<0.05). This is consistent with a raising effect of the palatal context on the second formant, with short [u] being affected selectively strong. The situation is similar for the low vowel, again with the effect of place (p<0.05), length (p<0.05) as well as the interaction (p<0.01) reaching the level of significance. The effects differ in their relative strengths comparing [a] and [u] contexts, with the interaction being more dominant in [a] context. Taken together, the scenario conforms to the palatal context behaving as a kind of "magnet", attracting the formants of vowels other than high front [i] in its direction.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	3	1685627.37	561875.79		
place	1	60.96	60.96	0.02	0.8989
Residuals1	3	9584.25	3194.75		
length	1	17873.08	17873.08	18.07	0.0239*
Residuals2	3	2967.42	989.14		
place:length	1	322.92	322.92	0.63	0.4859
Residuals	3	1542.10	514.03		

Table 3.7.: ANOVA of the second formant frequencies of the data plotted in the ellipse plots, with F2 in /i/-vowel contexts serving as dependent variable, and place of articulation (palatal versus velar) and vowel length (long versus short) as factors.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	3	34261.56	11420.52		
place	1	512761.21	512761.21	48.74	0.0060**
Residuals1	3	31558.66	10519.55		
length	1	280346.54	280346.54	19.75	0.0212*
Residuals2	3	42586.27	14195.42		
place:length	1	233130.66	233130.66	32.01	0.0109*
Residuals	3	21849.20	7283.07		

Table 3.8.: ANOVA on the second formant frequencies of the data plotted in the ellipse plots, with F2 in /u/-vowel contexts serving as dependent variable, and place of articulation (palatal versus velar) and vowel length (long versus short) as factors.

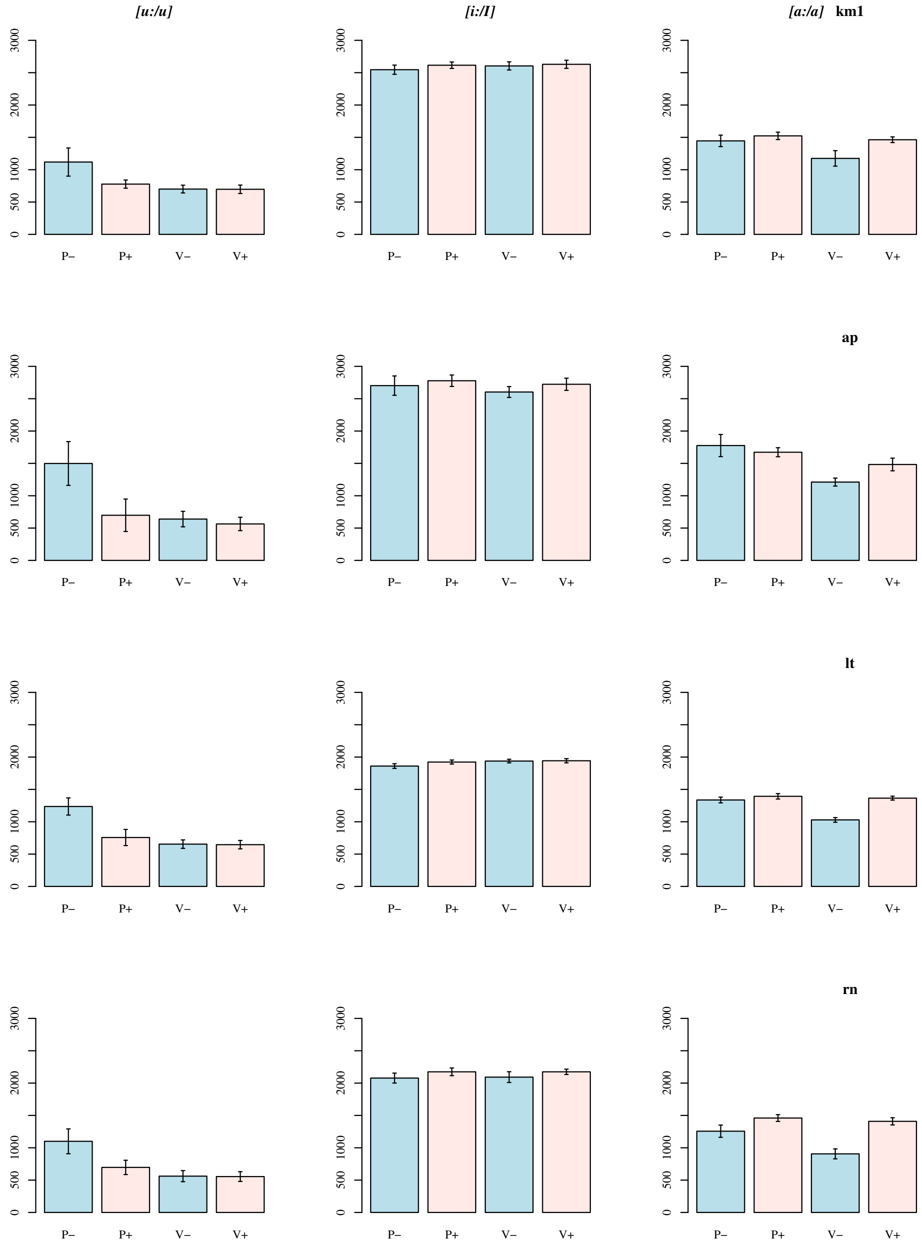


Figure 3.20.: Bar plots of the second formant frequencies in the different vowel contexts. Vowels contexts (i) in the left panels: [u/uː], (ii) in the middle panels: [i,iː], and (iii) in the right panels [a,aː]. speakers from top to bottom: km1,ap,lt,rn. P: palatal consonant before target vowel, V: velar consonant. Vowel length is indicated by ± signs.

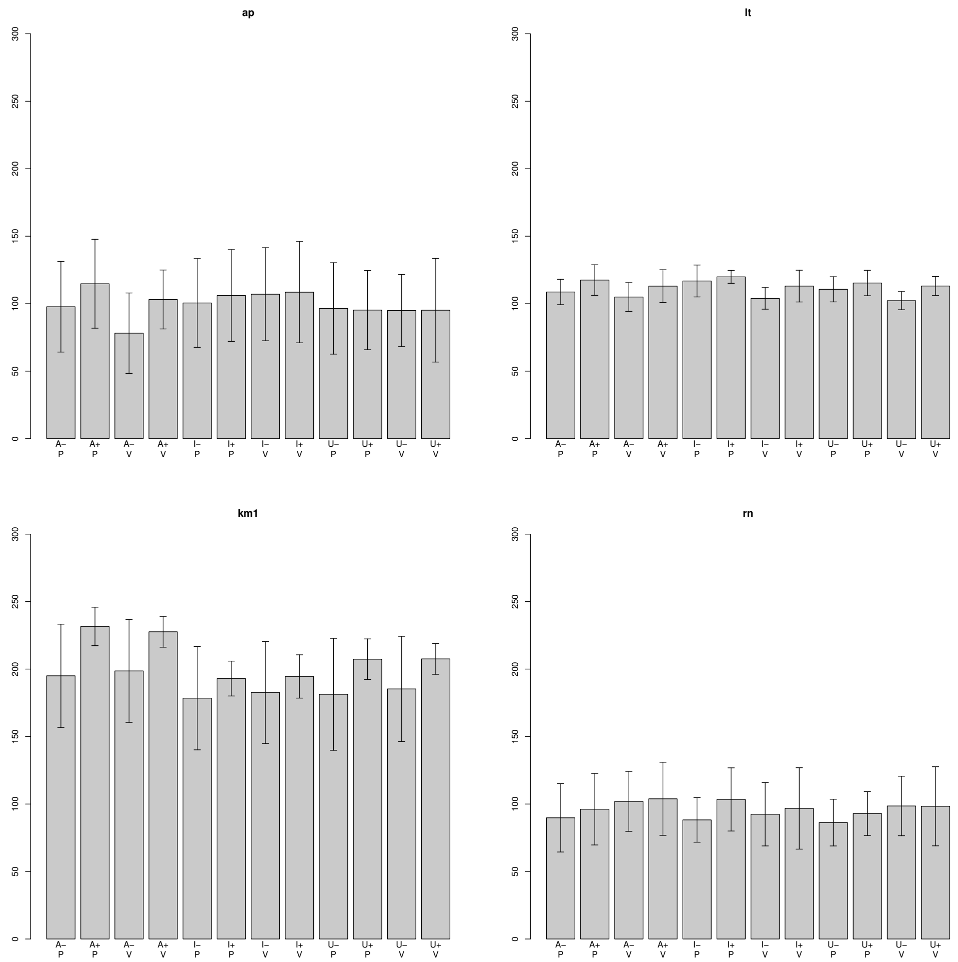


Figure 3.21.: Mean closure durations in velar and palatal contexts for the four speakers of the Hungarian EMA study. Speaker rn is plotted in the bottom right corner of the plot. Bars indicate 1 standard error.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	3	196010.96	65336.99		
place	1	207729.31	207729.31	20.43	0.0202*
Residuals1	3	30500.67	10166.89		
length	1	166964.99	166964.99	13.49	0.0349*
Residuals2	3	37130.14	12376.71		
place:length	1	84835.00	84835.00	74.04	0.0033**
Residuals	3	3437.27	1145.76		

Table 3.9.: ANOVA on the second formant frequencies of the data plotted in the ellipse plots, with F2 in /a/-vowel contexts serving as dependent variable, and place of articulation (palatal versus velar) and vowel length (long versus short) as factors.

9.3. Postures

What are the productional correlates of these differences? To gain first basic insights about the spatial organization of palatal versus velar stops in Hungarian as depending on vowel context, this section will display and discuss tongue, lip and jaw configurations by drawing dispersion ellipses at various temporal instances. Before turning to the organization of these plots, some methodological issues have to be mentioned: After inspection of the first versions of these graphs, problems for two speakers arose: First, it turned out that for speaker lt, the first 115 trials had to be eliminated because head position had altered within the helmet at around this time of the recording session. Therefore, following articulatory analyzes of the data of this speaker are not based on the full amount of trials like in the acoustic analyzes, rather about 25% of the data had to be removed. A more substantial problem arose for speaker rn: Although there was no apparent experimental artifact present in his data, the dispersion ellipses for this speaker pointed to a generally much higher amount of within-category variation. This effect will be evident throughout all analyzes but ignored for the posture analyzes. Partly, this pattern might be explained by the patterns of acoustic closure durations (see fig. 3.21). While closure durations for speaker km1 are partly more than twice the closure durations for rn, speakers ap and lt show intermediate durations. Durations for lt are still visibly longer than for speaker rn whereas for speaker ap this pattern is not evident. A further attempt was made though: Hypothesizing that the acoustic segmentation criteria as described above are not efficient for this speaker, an alternative articulatory segmentation rationale for determining the burst was applied: The minimum velocity of the tongue back sensor for oral closure. This segmentation strategy did not prove more efficient and was abandoned again for the sake of consistency with the procedure for the other speakers. Still it prooves that the large amount of variability for this speaker are not due to poor experimental design like the lack of control of speaking rate, i.e. while presumably a partial reason for the increased variabilities observed in rn's data, they cannot be fully attributed to particularly fast speech.

As mentioned, until so far most of the articulatory studies on palatal articulations have used palatography. Therefore, the logical first step consists in qualitatively describing the observed postural patterns (where appropriate, with reference to the formant measurements described in the last section). These qualitative observations refer to some selected graphical displays embedded in the text, but in order to also give a complete account of these data, they are additionally rendered fully in Appendix B.

Qualitative evaluation of palatal place of articulation

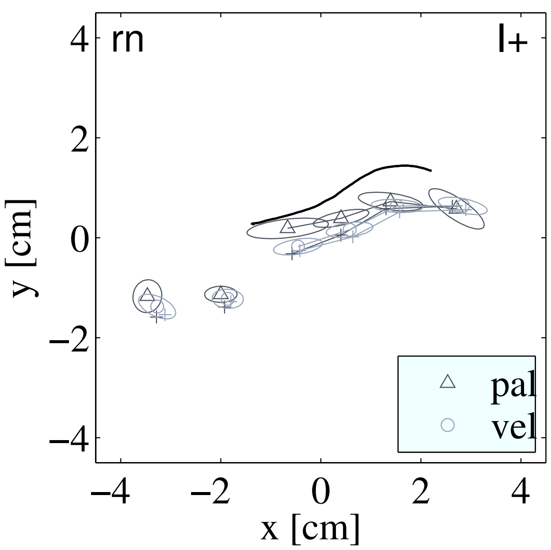
The issue here is to elaborate valid descriptions of the place of articulation for the palatal articulations of the four speakers. These descriptions will mainly rely on the initial positions, often also evidence from medial positions will be added. Although, as already mentioned, the fleshpoint data acquired in the present study are not fully compatible, a description in terms of the taxonomy by Recasens (1990) - which is mainly EPG-based - will be attempted where possible. The figures show 2-σ ellipse plots for palatal and velar consonants, measured at burst time. Black triangles indicate mean configurations for palatals, black circles for velars. Further, mean positions of the following vowels are displayed in darkgrey for palatal and in lightgrey for velar contexts. They are indicated by + signs, their dispersion ellipses are not shown. Although the focus is on the palatal consonants, the additional display of the velar consonants and the the vocalic midpoints is hoped to provide additional orientation.

The most serious problems for arriving at a classification presented speaker **rn**. First, as mentioned already, his articulatory behaviour is characterized by strong horizontal movements. Perhaps this also has to do with his palate morphology which can be described by (i) a small but domed palatal vault and (ii) a markedly long and flat "prepalate" - according to the classification scheme shown in figure 3.14.[27] My interpretation for this speaker is that for most environments the two frontmost sensors TT and TM are most substantially involved in forming the closure in initial position, and that the constriction is predominantly at the alveolars and the prepalate for most configurations like in figure 3.22a. In some situation, especially in medial positions, the third (TD) sensor appears to become also relevant though, like in figure 3.22b, which might involve parts of the mediodorsum and cause the constriction to extend also to parts of the mediopalate.

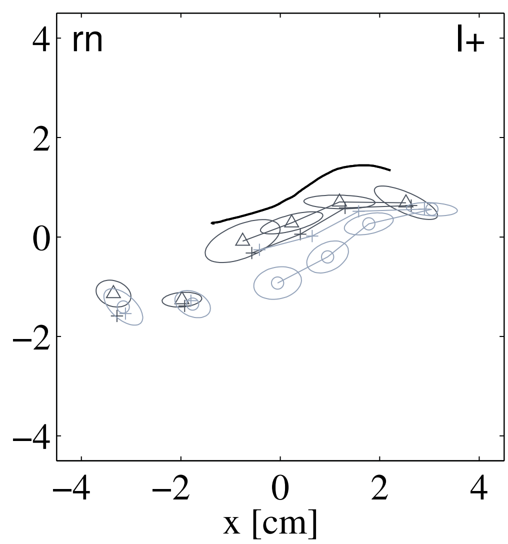
In contrast to the difficulties in describing the patterns for rn, the case for speaker **km1** is relatively clear. The morphology points to a relatively flat palate shape, and the articulations are relatively consistent given visual inspection. The sample in figure 3.23 - this time displaying short [i]-contexts - confirms this pattern: The tongue tip (TT) is down, the constriction is achieved with the two mid sensors TM and TD, also the TB sensor is not directed towards the palate. The active raising of the tongue towards the palate, or better, the critical sensor affiliation differentiating the palatal closure from the tongue shape of the palatal vowel is the TM sensor. Taken together, predorsum and mediodorsum are involved in forming the closure at a pre- and mediopalatal location.

The pattern displayed by speaker **ap** is still another one. This time, due to reasons to be discussed, the figures in 3.24 displays the configurations in testwords with initial [aː] and [uː].

[27] Recall that the palatal outline tracings as shown in for example in figure 3.22a were adapted to the tongue configurations by visual inspection rather than by a more scientific criterion.

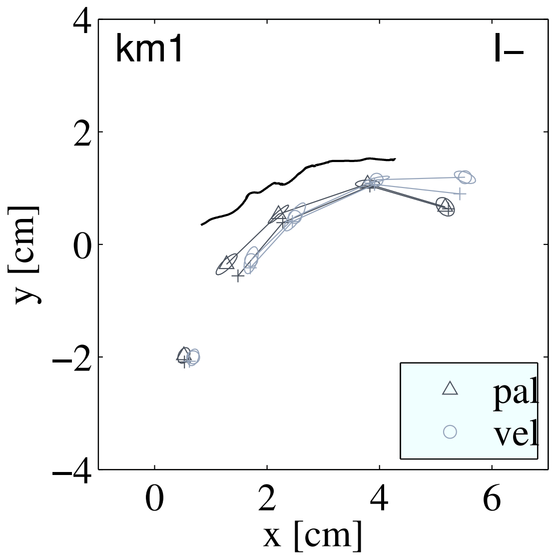


(a) Pos: initial, vowel context: [iː]

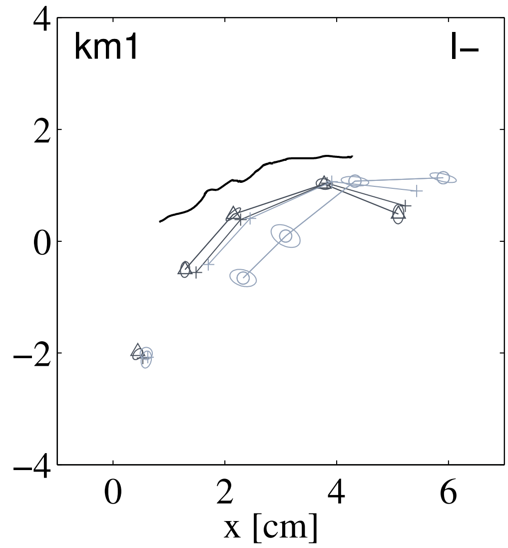


(b) Pos: medial, vowel context: [iː]

Figure 3.22.: Configurations (mean and 2-σ dispersion ellipses) of the initial and medial consonants in Hungarian [CiːCɑ] sequences. Left panel (a): initial consonants, right panel (b): medial consonants. Triangles indicate palatal, circles velar consonants. Vowel configurations are also displayed using + signs, but without ellipses. Corresponding vocalic configurations are also indicated by the same color scheme as the consonants: Vowels in palatal context are shown in darkgrey, velar context in lightgrey. Speaker: rn.

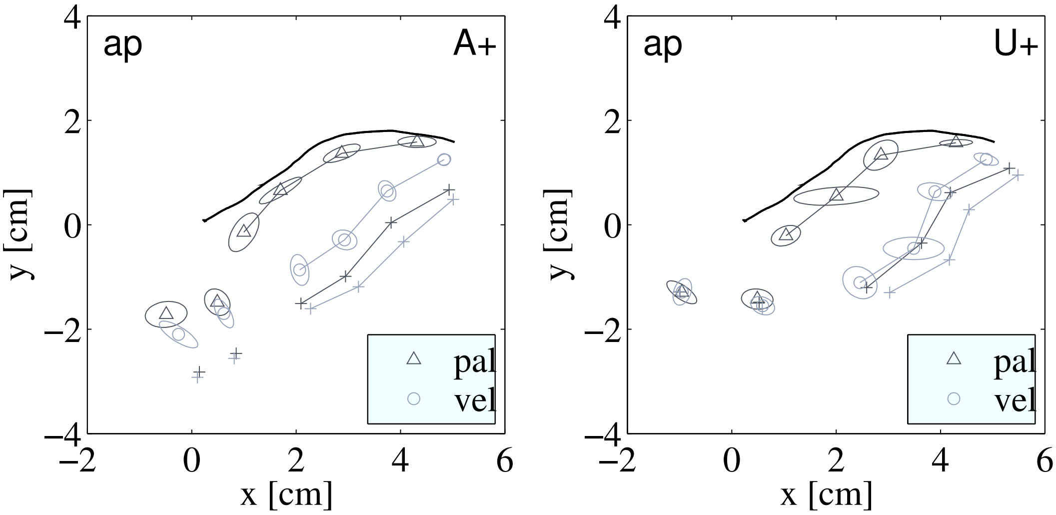


(a) Pos: initial, vowel context: [i]



(b) Pos: medial, vowel context: [i]

Figure 3.23.: Configurations (mean and 2-σ dispersion ellipses) of the initial and medial consonants in Hungarian [CiːCɑ] sequences. Left panel (a): initial consonants, right panel (b): medial consonants. Triangles indicate palatal, circles velar consonants. Vowel configurations are also displayed using + signs, but without ellipses. Corresponding vocalic configurations are also indicated by the same color scheme as the consonants: Vowels in palatal context are shown in darkgrey, velar context in lightgrey. Speaker: km1.



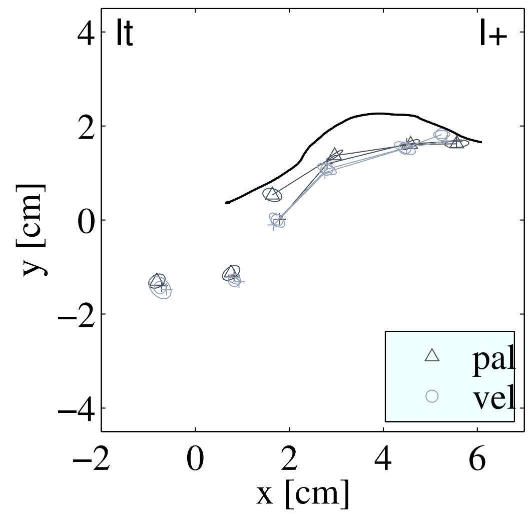
(a) Pos: initial, vowel context: [aː]

(b) Pos: initial, vowel context: [uː]

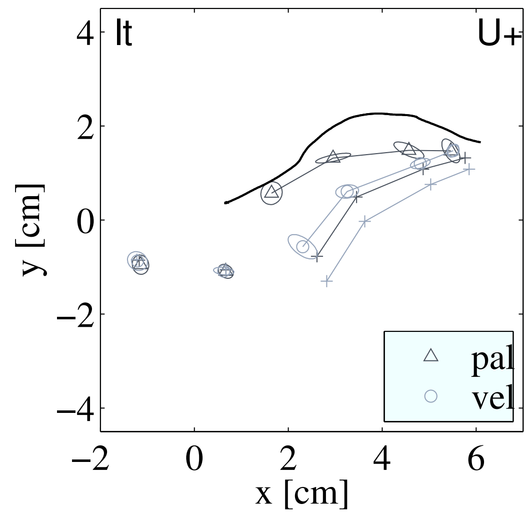
Figure 3.24.: Configurations (mean and 2-σ dispersion ellipses) of the initial consonants in Hungarian [CaːCɑ] (left) and [CuːCɑ] sequences. Triangles indicate palatal, circles velar consonants. Vowel configurations are also displayed using + signs, but without ellipses. Vowels in palatal context are shown in darkgrey, velar context in lightgrey. speaker: ap.

For the [aː] context shown in figure 3.24a, an interpretation of a long constriction with the three rearmost sensors TB, TD and TM seems to be consistent with the data. In contrast, the [uː] in figure 3.24b seems to suggest that the TM sensor is not involved in all realizations of the palatal closure before [uː], i.e. the large variability of TM suggests that sometimes only TB and TD are actively involved in closure formation. Taken together, this speaker's realizations point to a pre- and medio-dorsal realization with eventually even postdorsal parts of the tongue involved. Similarly place of constriction formation is relatively far back, not only involving pre- and mediopalatal zones, but also the postpalatal region is likely to be involved.

For speaker **lt**, a mention of the palate morphology seems to be in place before turning to the interpretation of tongue shapes: This speaker has a quite pronounced and sagitally domed palatal vault. Figure 3.25 shows initial consonantal profiles - before long [iː] and [uː]. An interesting feature of this speaker's profiles for the stop in the context of [iː] is that the dorsal sensor is more retracted for the palatal than for the velar plosive. A further point is that lt's initial short [i] shows some retraction of the tongue blade, i.e. the first two tongue sensors TT and TM. Taken together, a predominantly laminal and predorsal articulation of the palatal stop seems to be a convenient interpretation in the light of articulatory facts. More difficult is the judgment of the place of articulation. Apart from the pre- and mediopalate articulation, the sometimes quite strongly retracted rearmost sensor possibly could be hypothesized to form a second, si-



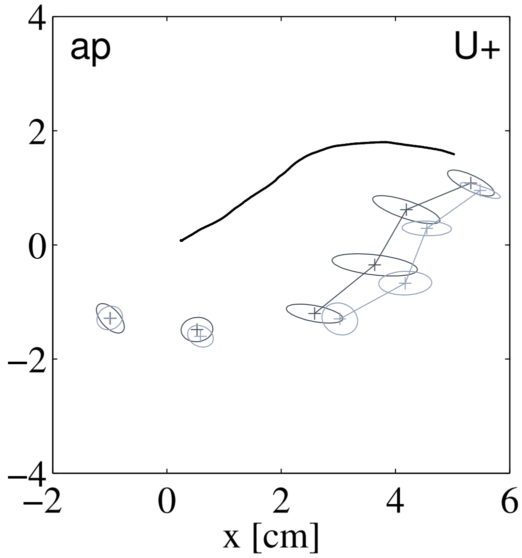
(a) Pos: initial, vowel context: [iː]



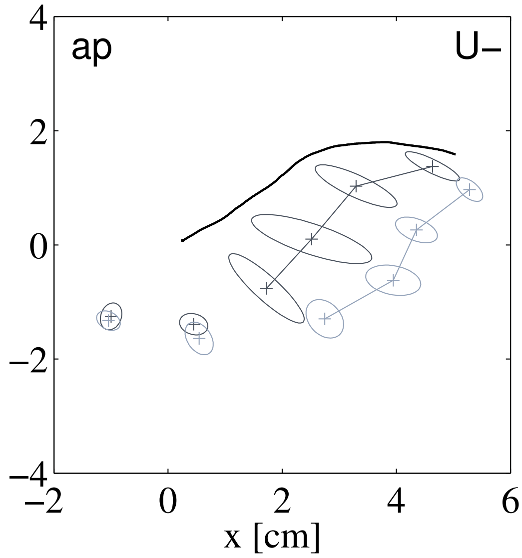
(b) Pos: initial, vowel context: [uː]

Figure 3.25.: Configurations (mean and 2-σ dispersion ellipses) of the initial consonants in Hungarian [CiːCɑ] (left) and [CuːCɑ] sequences. Triangles indicate palatal, circles velar consonants. Vowel configurations are also displayed using + signs, but without ellipses. Vowels in palatal context are shown in darkgrey, velar context in lightgrey. Speaker lt.

multaneous postdorsal closure at the postpalate. In order to summarize this section, arriving at a classification of the Hungarian palatal stop, these data confirm Recasens' observation that palatal obstruents are not a single homogeneous class. Hungarian palatal stops definitively are not alveolopalatals, a classification as predominantly pre- and mediodorsal articulation at pre- and mediopalatal place of articulation appears to be more plausible in the light of the data just presented. Nevertheless, some departures from this classification occurred in some vocalic contexts, which could be related to some conspicuous morphological features of the speakers (lt and rn). In some respect bridging the gap to the next section, and simultaneously substantiating the reliability of the formant analyzes in the preceding one, the patterns of the palatals of speaker ap in the context of the long and short back rounded vowel are presented in figure 3.26. These articulatory vowel configurations correspond to the formant pattern in figure 3.16: The high variability in the formant measurements of short [u] correlates with the large articulatory dispersion ellipses as shown in figure 3.26 and the large error bars in the barplot of the second formants shown in figure 3.20 in comparison to their long counterpart. The observation that this large variability is a function of the following vowel-context for ap points to a discussion of coarticulatory phenomena in the next section: blending and coarticulation.



(a) vowel configuration of [uː]



(b) vowel configuration of [u]

Figure 3.26.: Configuration of (a) [uː] and (b) [u] for speaker ap. Articulatory configurations correspond to 95% formant ellipses parallel to figure 3.16.The color scheme is as before: Vowels in palatal consonant context are shown in darkgrey, and vowels in velar context are shown in lightgrey.

9.4. Blending and coarticulation patterns

As already mentioned while assembling the leading questions of the present work (see section 7.1), it is one of the goals of this work to decide whether the palatals obey a blending or a coarticulation scenario. In AP terminology, blending indicates the conflict between gestural demands acting on the same gestural tier leading to the vast place of articulation changes for velars. A first basic illustration for the behaviour of velars and palatals is plotted in figure 3.27. It shows 1-σ-ellipses of palatal and velar consonants calculated by the means over all vowel contexts at burst time for initial and medial consonants. Small ellipses indicate the palatal configuration and large ellipses show the velar configuration which already gives the answer for the velars: they blend. Still, this picture offers no conclusive interpretation for the palatals: In section 6.3, the existence of a "mild blending" scenario was mentioned, and following analyzes will attempt to provide an answer whether this scenario is viable in order to arrive at an interpretation.

One way of quantifying the place changes the palatal stop undergoes when varying vowel context, vowel length and position within the word is to devise factor-analytic projections into "vowel space".[28] The analyzes reported here provide an extension to similar analyzes as fur-

[28]Typically, factor analytic projections have been used for research dealing with vowels. In the present context, talking about "consonant spaces" or "consonant/vowel spaces" would in principle be more accurate.

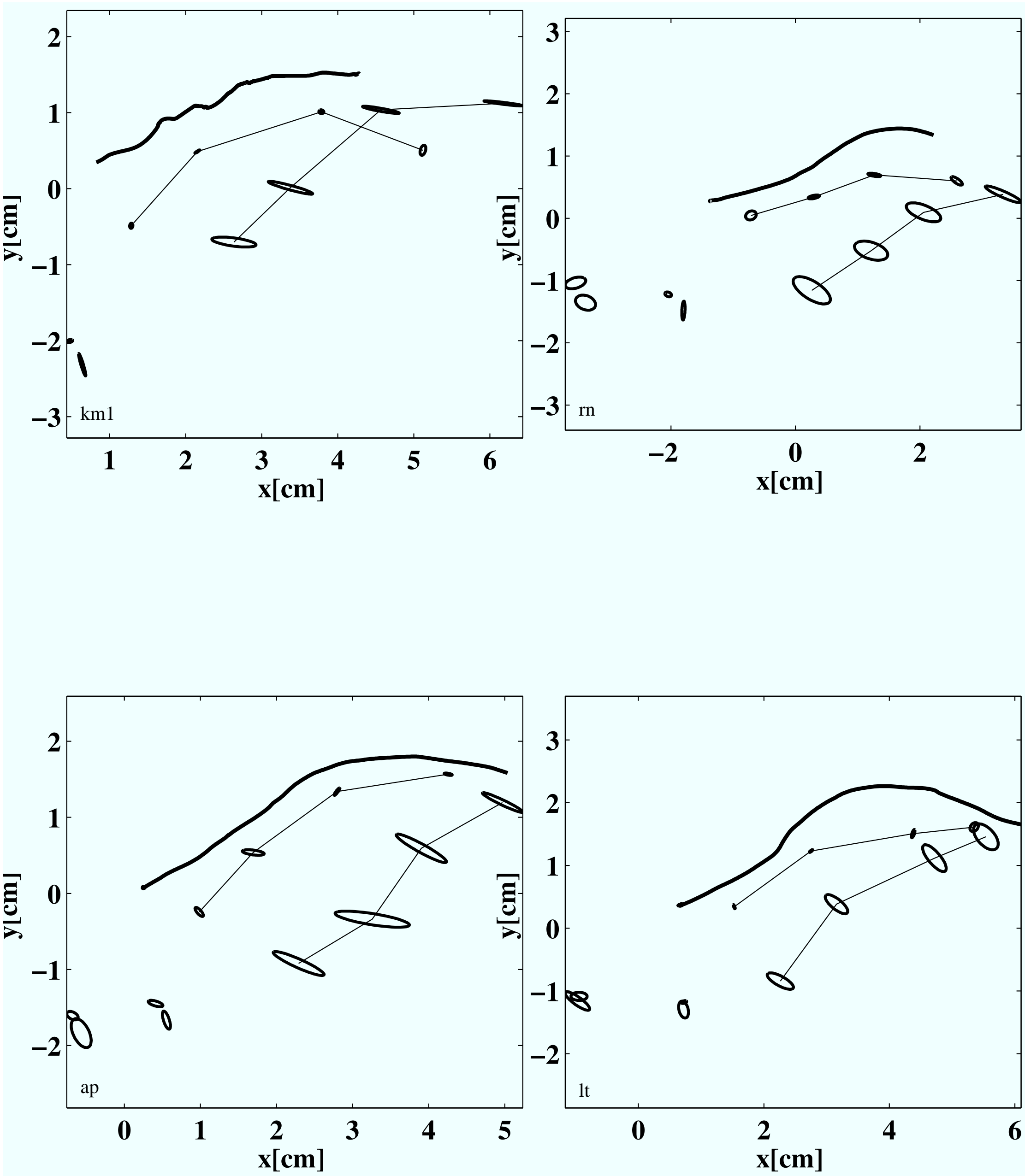


Figure 3.27.: 1-σ-ellipses as indication for the variation of palatal and velar consonants as calculated by the means over all vowel contexts at burst time for initial and medial consonants. Small ellipses: Palatal configuration; large ellipses: velar configuration.

nished in Geng und Mooshammer (2004), with some substantial differences though: While the focus in this paper was on general organizational properties of postures for the sounds under consideration in articulatory space, the question here is shifted closer to the particular blending question under consideration. This will be manifested in the shape of the input data: While the data in Geng und Mooshammer (2004) included also both jaw and lip sensor data, the focus is shifted towards mere tongue shapes in the - formally - similar analyzes reported here. In figure 3.28, results of Principal Component analyzes of the covariance matrices of the averaged articulatory configurations excluding the lips during (a) the initial burst, (b) the medial burst, (c) the vowel configurations at the midpoint of V1 are shown and (d), the vowel configurations at the midpoint of V2(=a) are shown. Principal component analysis (PCA) involves a decomposition of a larger number of usually correlated variables into a (usually smaller) number of not directly observable uncorrelated variables. The aim of both methods is to reveal meaningful underlying variables, in our case, the articulatory configurations containing separate x- and y-positions for four tongue sensors can be decomposed into a two-dimensional representation still representing the gross topology of the articulatory space analyzed. The first principal component accounts for as much of the variability in the data as possible, and so does each succeeding component. For these Principal Component Analyzes, the first factor explains between 80.4 and 89.9% of the variance and the second factor between 8.1 and 16.4%. The total amount of variance varies between 93.8 and 98%. This means that the most substantial portion of the variance is accounted for in the PCA representations. The grey "triangles" shown in figure 3.28 are the vowel configurations at the midpoint of V1, with the darkgrey triangle for the palatal context and the lightgrey triangle for velar context. The most salient pattern of all these plots seems that the velars show much more inclination to blend with their vocalic environment in comparison with the palatals, further substantiating the strong blending scenario for the velars. This correlates with the findings as shown above in figure 3.27, i.e. large place variation for the velar stop and the impression of a stable palatal configuration very close to the first factor score of /i/. The velars substantially show more variation for all speakers.[29]

Further, only the second factor on the y-axis distinguishes between the palatal postures. The first factor is almost constant for the palatals and in line with the classic interpretation as "front-raising" to an /i/-like shape (Harshman et al., 1977). The most consistent pattern for the velars is that the shape of the velar in /uː/-context is close to the /uː/-corner. Of particular interest are the projections of the /iː/ in velar context and initial position: Here, the patterns do not seem to be fully conclusive: For km1, the /k/ is separated from the main palatal cluster, while for speaker lt, the initial /k/ in /iː/-context is almost completely "dissolved" in the palatal cluster. The same holds for speaker rn and ap. Still, the initial velar in /i/-context is the closest to the palatals of all velars for all speakers. With respect to answering the question of blending, the mild blending scenario to some extent can be confirmed: The palatals are linearly separable by the second factor, i.e. the one on the y-axis. Within initial palatals, the results are quite consistent: For three out of four speakers, - the exception is speaker ap - the palatal in /uː/ context is sandwiched between /iː/ and /aː/ contexts. Further, the medial contexts - always followed by [a] - are usually

[29]One could raise the objection that the place of maximal constriction is not reliably measured even by the EMA tongue back sensors, but the parallel orientation of the ellipses for the rearmost sensor indicates that the constriction is mostly caught.

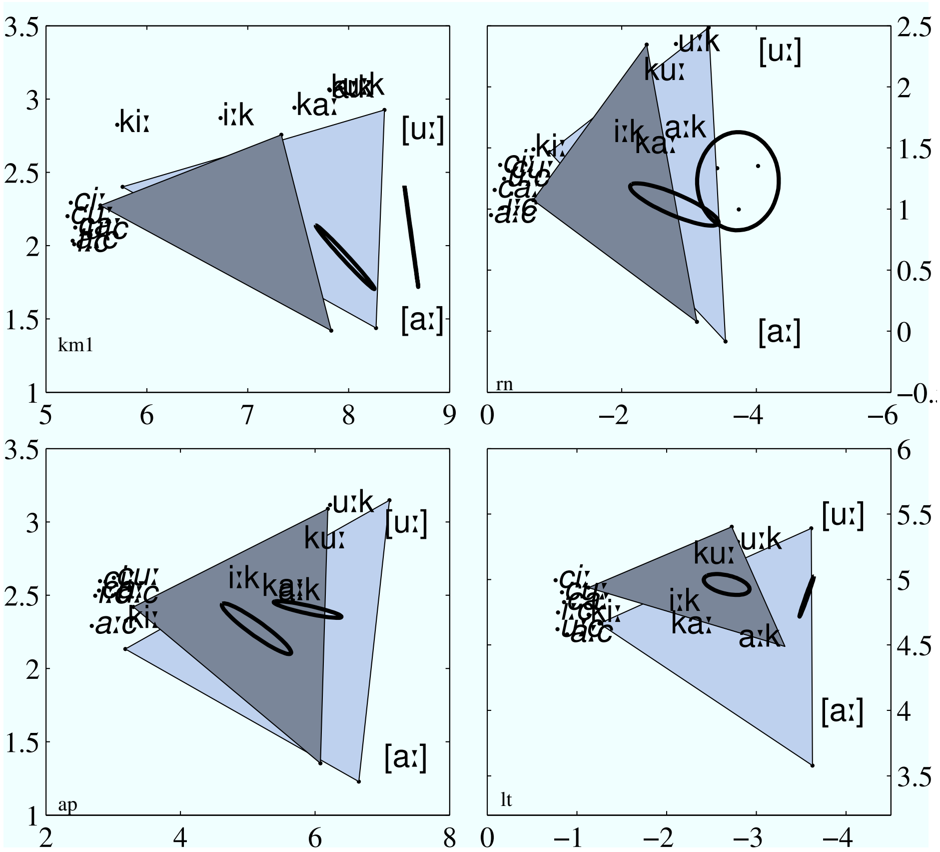


Figure 3.28.: Speaker-dependent Principal Component analyzes of the mean tongue configurations at different temporal landmarks: the initial release, the medial release. Further, the light gray triangles are the midpoints of the long corner vowels in the context of velar consonants, the dark gray triangles the corresponding projections of the long corner vowels in the palatal contexts. The dispersion ellipses capture posttonic V2 variations. The palatal ellipses are always those closer to high front configurations like [i].

closer to the initial /caː/ context than are the palatals in the other vocalic contexts. In other words, there is some basic topological information preserved in the factors extracted. Moreover, another finding concerning the vowel triangles is remarkable: The sizes of the triangles are reduced in the palatal context, indicating, that the palatal consonants themselves exert a stronger coarticulatory influence on the vowels than the velars: in comparison with the velars, this results in a fronting of [u] and a rising of the [a]. This “palatal magnet” pattern also prevails in the posttonic [a]. The dispersion ellipses in figure 3.28 for these vowels consistently travel in palatal direction for all speakers selectively in palatal contexts.

9.5. Kinematic characteristics

Figure 3.29 shows the distances traveled during the four intervals which were defined earlier - see also in the caption - for the tongue dorsum sensor. Similar results were obtained for the tongue back sensor. The most salient aspect of these plots is that the total distances traveled by the TD sensor during the /VCa/-sequences are larger in the palatal contexts for /a/ and /u/-contexts; the reverse holds for the /i/-contexts. Concerning the velar contexts, we found some relatively surprising patterns in comparison with the data we analyzed earlier. In particular, in Geng et al. (2003) we found consistently larger **total amplitudes** in the /aː/-contexts. We interpreted this finding in agreement with Munhall et al. (1991), who observed a reduction in movement complexity after algorithmic removal of the jaw influence. This tendency is weaker in this corpus and even reversed for speaker ap, who exhibits longer total movement amplitudes in the context of the high back vowel. So if the patterns for the /a/-contexts could at least partly be explained through a contribution of the jaw for *both* consonants, this explanation is not justified for the large total amplitudes for the palatals observed in the context of /uː/. If we cannot attribute these large movements to an influence of the jaw, then this pattern must be attributed to a strong movement component by the tongue itself. We will return to this point later.

Another quite general observation in these plots is that the movement **amplitudes during the stop**, - i.e. the black parts of the bars - are usually larger and more evenly distributed between vocalic contexts for the velar consonants. This pattern is not fully valid for the /aː/ and /uː/-contexts, which show large movements during the stops for palatals consonants as well. This pattern is most probably due to the large total amplitudes aforementioned. Similar observations can be made for the distances traveled between the stop release and the onset of the second vowel /a/.

In order to devise a crude method for quantifying the *direction* the tongue paths travel during the closure interval, a “direction coefficient” was calculated which weighted the distances the sensors traced in the closure interval by a “direction coefficient”. This direction coefficient was determined as the sign function of the difference between the x-coordinates of the first sample of the closure interval and the last sample of the closure interval, i.e., negative values indicate the tendency to make a movement in backward direction during the closure. The error bars in figure 3.30 indicate the standard deviations of this composite for the tongue dorsum sensor. Again, similar results were obtained for the tongue back sensor. Note that this is a very gross measure, in particular, a mainly vertical movement during closure would have the consequence of making the sign function, which is only based on the horizontal movement in this interval, relatively arbitrary. So note that the partially large standard deviations in these plots underlyingly might

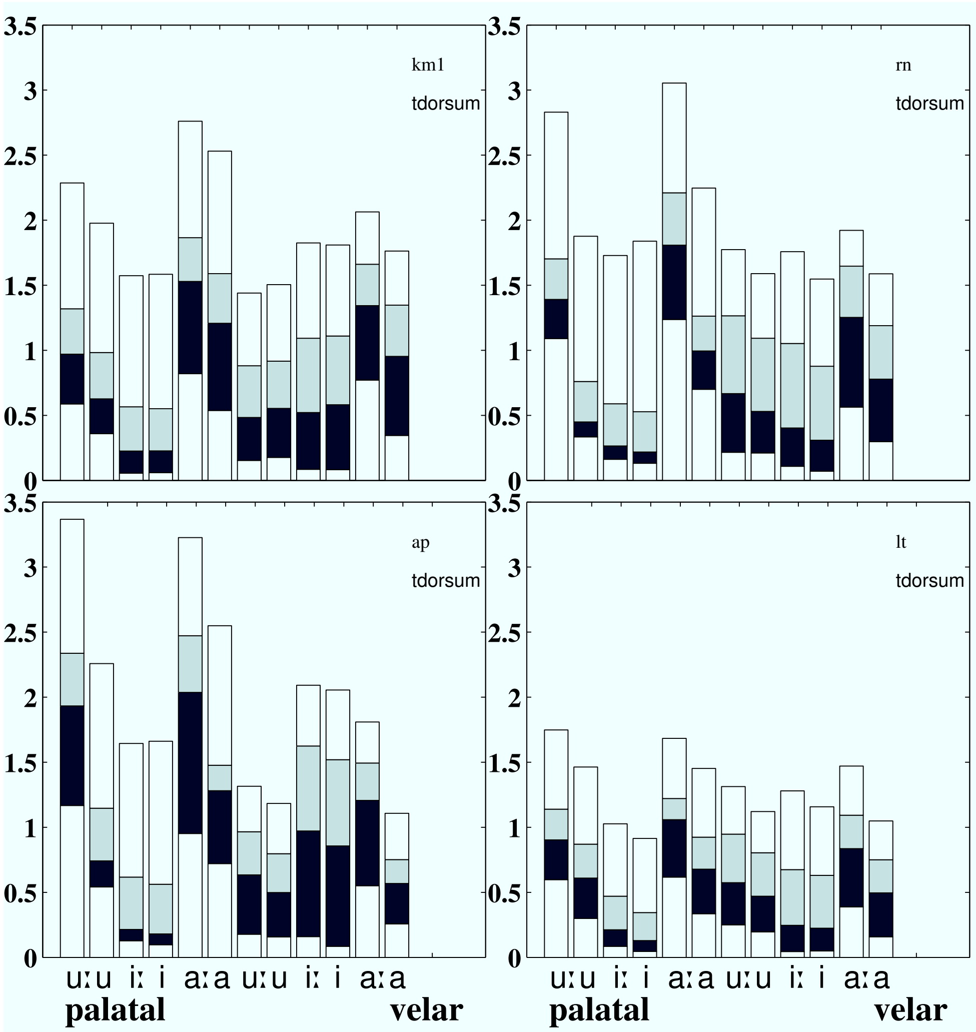


Figure 3.29.: Distances the tongue dorsum sensor traveled during the four different intervals of the medial sequences described. The stack bars indicate from bottom to top: first stack, white, distance traveled during the first vowel; second stack, black, distance traveled during oral closure; third stack, gray, distance traveled between stop release and the onset of the second vowel; fourth stack, white, distance traveled during the second vowel. Vowel context is displayed underneath, palatals are shown in the left part of the graphic, velars in the right. Vocalic context is displayed underneath.

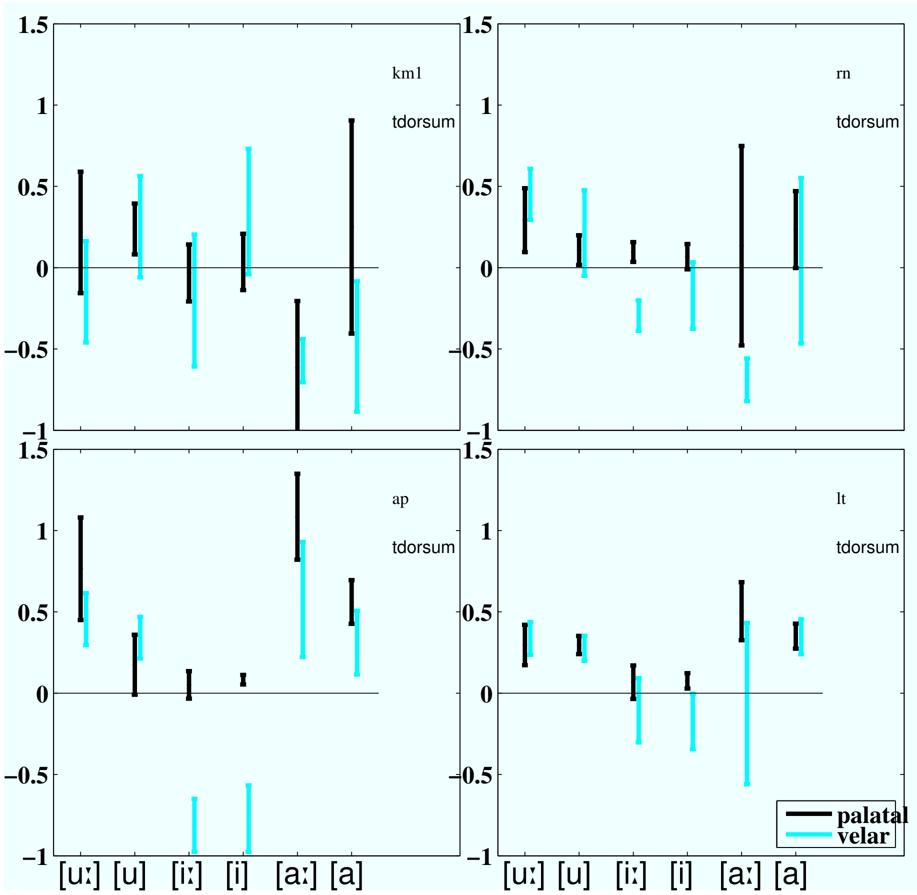


Figure 3.30.: Distances the tongue dorsum sensor traveled during the the closure interval weighted by a direction coefficient. The direction coefficient was determined as the sign function of the difference between the first sample of the closure interval and the last sample of the closure interval, i.e., a negative value indicates the tendency to make a movement in backward direction during the closure. Black: Palatals; grey: velars

represent (a) heterogeneity in the true movement as an expression of token-to-token-variability or (b) the predominance of a vertical component causing a "noisy" sign function.

So the description of results will be limited to the data which do not have this uncertainty and exclude the patterns which have large standard deviations and values above and below zero. For example, this concerns the results for speaker lt in the context of /aː/: As due to the influence of the jaw, a vertical movement dominates, which results in a noisy sign function and therefore in large standard deviations.

Following the front vowel /i/ (long and short) and short /u/, very little movement was found during the palatal closure. A higher degree of forward movement occurred during the palatal following /uː/ and /aː/, both with the exception of speaker km1. After /i/, three speakers showed little movement during /k/ which is consistent with the data on German (Geng et al., 2003). ap in contrast showed a quite pronounced backward movement during the velar stop for both long and short [i]. Movement directions following long /a/ varied inter- and intraindividually (e.g. large standard deviations for speaker lt and km1). These results suggest that the palatal stop is produced close to the constriction location for /i/, therefore no movement is required during closure.

Figure 3.31 shows the correlations between the positions in the mid of the first vowel and the distances traveled during the stops for the tongue dorsum sensor. Again, similar tendencies were observed for the tongue back sensor. The left panel shows the correlation of distance and x-position and the right panel of distance and y-position. These results will be interpreted on the background of comparable data on velar productions for German (from Geng et al., 2003). In this paper we had found the following pattern, consistently for all speakers:

(i) the more anterior the sensor location during the following vowel, the larger the distance traveled during the stop closure: This implies a *positive correlation*

(ii) the higher the position of the sensor during the preceding vowel, the smaller the distance traveled during the preceding stop: This implies a *negative correlation*.

Of particular interest are of course the velars and whether their behaviour is equivalent to that of the German dataset. The data are incommensurable though: For the correlation of anteriority and traveled distances, three out of four speakers exhibited the opposite pattern of a negative correlation, only the behaviour of speaker lt conformed to the predecessor study on German. For two speaker, km1 and ap, the correlation furthermore was substantial, whereas the correlation for rn was marginal. Comparing the correlations between vertical position in the preceding vowel and the sensor distances traveled during oral closure, two speakers, lt and rn conform to the patterns observed for German, i.e. they show a substantial negative correlation between those parameters. For km1, the correlation is spurious, and the opposite pattern was again found for speaker ap.[30]

9.6. Discussion

The main results of the results on the Hungarian data can be summarized as follows:

[30] Note that the vowel environment was richer in Geng et al. (2003).

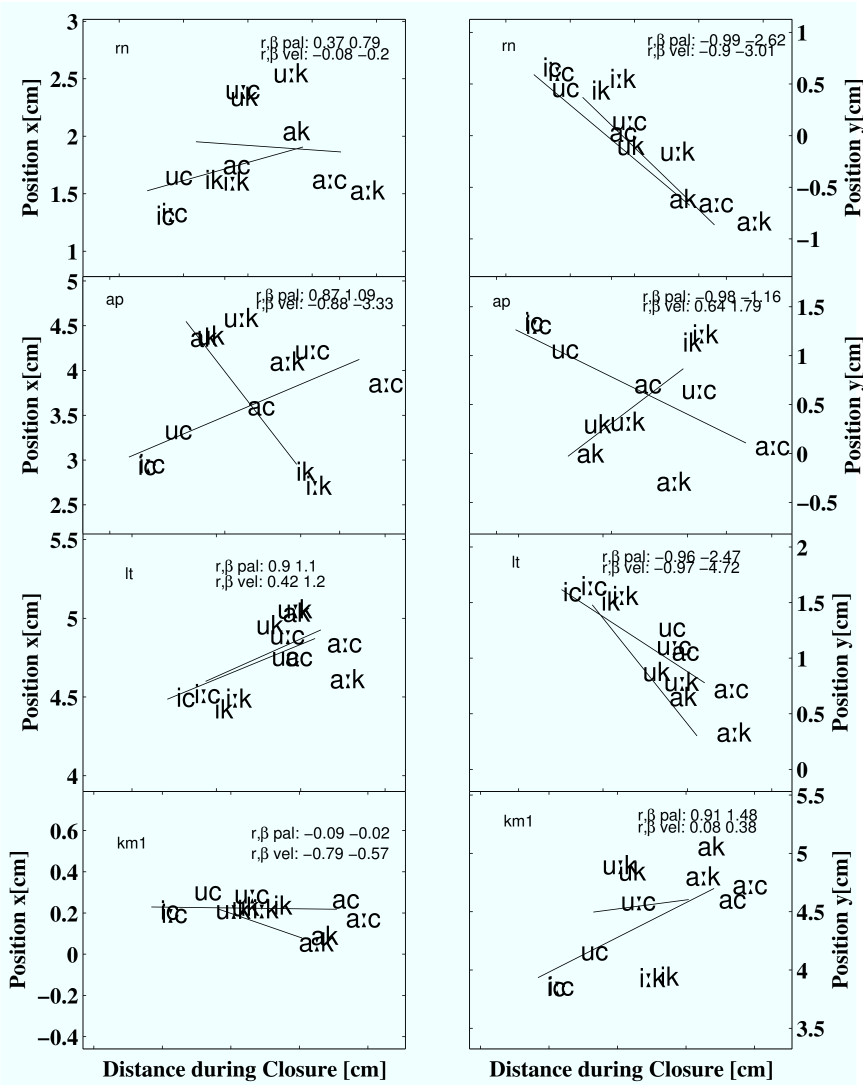


Figure 3.31.: Correlations between the positions in the mid of the first vowel and the distances traveled during the stops, both for the tongue dorsum sensor. The left panel shows the correlation of distance and x-position and the right panel of distance and y-position.

- A classification of the articulation of the palatal: The palatal stop in Hungarian is a dorsal sound, primarily formed at the pre- and mediopalate. Still, the data might point to the possibility of a second postdorsal constriction formed at the postpalate for one speaker (lt) or a more anterior constriction for yet another speaker (rn). These articulations have been traced to the effects of tract, ore more precisely, palate morphology, i.e. the dome-shaped palatal vault for lt and the comparatively long prepalate for rn might play a crucial role for these speaker-specific strategies. It has to be highlighted again that the midsagittal tracings of palate outlines do not offer the possibility of extensive morphological comparisons, the importance of which has long been recognized (e.g. Stone, 1991).

- Coarticulatory resistance of the palatal stop: The palatal stop is not very prone to vowel-like influences in the sense that (i) the palatal itself exhibits a very stable configuration in comparison to the velar. This stable configuration is similar to an /i/-like shape and (ii) the palatal itself exerts a strong influence on the neighbouring vowels inasmuch as the size of the vowel space is shrunk in comparison to vowel space in the context of the velar stop.

- Factor-analytic projections illustrate that for three our of four speakers the tongue shape of the velar in the context of the palatal vowel converges to a palatal shape.

- There is systematic place-of-articulation variability in palatal stop articulation depending on the vocalic environment as evidenced by factor analyses.

- Another important finding involves the extremely large distances the palatal has to travel in the context of the long /uː/ over the whole VCV-sequence.

- The palatal can exhibit considerable movement amplitudes during oral closure. The movement amplitude of the palatal in /uː/-context *during* closure sometimes by far exceeds the one of the velar in the same context. This piece of evidence runs counter the phonetic classification of the Hungarian palatal as an affricate which might be considered as plausible on the basis of its overall positional stability.

These results can be interpreted in the theoretical framework as follows: Keating's approach creates its prediction from underspecification at different phonological or phonetic levels. This amounts to relatively little sensitivity to vowel-induced contextual coarticulation for palatals in contrast to velars if the derivations used are the intended ones for the Hungarian language. On the macro-level, Keating's approach makes quite good descriptions. Here, the necessity arises to summarize the influences Keating sees at work in shaping coarticulatory patterns. These are the following factors: (a) production constraints, operating both within and across languages, (b) constraints deriving from language-specific phonological structure and (c) language-particular constraints, unrelated to production or phonology and therefore unpredictable. The cross-linguistic inconsistencies concerning velar production - as demonstrated in comparison to the data on the German velars in Geng et al. (2003) - could lead to question the validity of the "transparency effect with respect to backness". An obvious solution would be not to leave unspecified backness to the phonetic level, but such a suggestion would require further phonetic substantiation. Another aspect with regard to Keating's work is that velar contextual variation is something that happens "over the course of the velar" alluding to the phenomenon of velar loops. The fact that palatals can - in certain vocalic environments - also generate large movement amplitudes during oral closure makes this suggestion quite unlikely though. The data suggest that an interpretation of the patterns for the palatals requires recourse on biomechanics like advocated by Perrier et al. (2003), but a the large overall movement amplitudes correlated with the large movement components also during oral closure demand recourse on general movement economy and motor control principles like advocated by e.g. Löfqvist und Gracco (2002).

Of more interest is the classification of the palatals though: The results of this study point to a specification of the palatals which confirm the "true palatal" obstruent classification for /c/ as described in Recasens (1990). Still, speaker-dependent strategies which potentially are strongly conditioned by tract and palate morphology call for more experimental research, surely not solely relying on EMA or EPG data, but rather a broader, a multimethodological developmental approach would be desirable in order to capture the complex transactions generating speaker-specific strategies developed on the basis of organismic factors. In the same spirit: The effect of the palatal context to attract the formant patterns in the direction of the high front vowels for contexts other than [i] would deserve further exploration: Although the effect has been demonstrated repeatedly in the literature (Recasens, 1985; Stevens und House, 1963), its exact articulatory implementation would deserve further exploration. It could be shown that there are sometimes drastic (see figure 3.26) correlations with supralaryngeal behaviour, the interplay with pharyngeal regions would be worth investigating using methods like Dynamic MRI or ultrasound. Of interest is as well the differential strength of this "palatal magnet" on short and long /a/ and /u/, with the short cognates being affected selectively strong.

A further question as laid out in the section describing the hypotheses is whether the "mild blending" secenario is viable in describing the patterns observed for the palatals. Given the topology-preserving structure in the second factors of the PCA analyses, this question can be answered positively, with some caveats. Recall that the second factor vowel loadings of the stop configurations in initial position were ordered [ciː > cuː > caː], with the medial palatals always alligning with initial [a]-context.

Of the utmost strategical relevance for the present work is the result of the convergence of the fronted velar with the palatal stop articulation for three out of four speakers as predicted by Recasens (1990, p. 276). It nurtures the hope that the presence of the palatal stop in the inventory might have a reactive effect on the velar stop and that the cross-linguistic comparison of velars might reveal informative patterns. But first, the analysis of the German dorsal data is indispensable.

10. Results II: German

In principle, the aim for the analyses of the German fricatives is similar in spirit to the study by Ambrazaitis und John (2004) and was already mentioned in section 7.2: The confirmation of the pattern of more contextually induced variability for fricatives as found by Ambrazaitis und John (2004). This implies that one could go directly in medias res, but some comments on the participants are in place, because these partly overlap not only with the current study, but also with the study by Pompino-Marschall und Mooshammer (1997): In particular, speaker jd in the current study is identical with speaker jdr in Pompino-Marschall und Mooshammer (1997) and

Ambrazaitis und John (2004), and speaker cg in the current experiment is identical with speaker cge in Ambrazaitis und John (2004). These aspects might become informative when starting - equivalent to the Hungarian data - with descriptions of the place of articulation alternations for the German data, this time with special emphasis on the fricatives.

10.1. Qualitative evaluation of place of articulation

The presentation of dispersion ellipses follows the scheme devised for Hungarian, although the selection of the figures displayed will focus more strongly on medial positions, in line with the working hypothesis borrowed from Ambrazaitis und John (2004). Again, a full record of the data is rendered in a separate appendix (see appendix C) which is owed to the selectivity of data presentation. Further recall the asymmetries in the experimental design: In German, dorsal fricatives have no regular occurrence in initial position and no voiced counterpart, therefore the initial context was equated to the velar stop. Before starting with the qualitative descriptions, recall from Ambrazaitis und John (2004) that the observed results for these speakers - although obtained by means of palatography - were the following: Speaker jd was characterized by the complete absence of uvular realizations of dorsal fricatives in his speech, whereas cg realized uvular fricatives, but in environments not completely consistent with the predictions made by Kohler (1995). Therefore it is tempting to discuss these two speakers in greater depth than the other two speakers.

Figure 3.32 shows ellipse plots of medial configurations for stops and fricatives for the two speakers jd(left) and cg(right). Stops are coded by ○ and fricatives by △. Contours of the preceding vowels are also displayed - without ellipses - and marked by + for both consonantal contexts. The manner context of these vocalic configurations is indicated by colors with darkgrey for the vowel in fricative context and lightgrey for the vowel in stop context.

The first striking observation is the different implementation of the stop and the fricative after [iː] (top panels). Speaker jd exhibits not much of a difference between these palatal configurations, i.e. the tongue shape displays a configuration which is in essence adapted to the shape of the hard palate. In contrast, for speaker cg, the stop configuration before [iː] is different in terms of the tongue tip and blade which are down for the stop configuration, i.e. the closure is produced with the backdorsum, or, in other terms, is less palatal. Recall that the following context is always [ɐ], therefore this could be due to anticipatory coarticulation. In contrast, for the fricative cg realizes a palatal [i]-like shape. The observation of the dorsal-only gesture for the stop for cg is also in line with its quite low jaw position.[31] After [uː] and [aː], jd also differentiates between stops and fricatives, with the constriction for the fricative more posterior. This pattern resembles that of speaker cg. A major difference between jd and cg in these non-front contexts is though that for jd, - and especially for the stop configurations - the rearmost (TB) and probably also the second rearmost sensor (TD) are likely to be associated to forming the constriction. This pattern contrasts with the pattern for cg, where none of the sensors apparently is in the vicinity of the actual constriction which therefore can be assumed to be back velar for speaker cg and velar for speaker jd. The analysis of the lax vowel contexts revealed no further insights, they are therefore not displayed here (but see appendix C). Rather, further qualitative

[31] For example, Keating et al. (1994) found jaw positions for [k] which were even lower than for some of the sonorants.

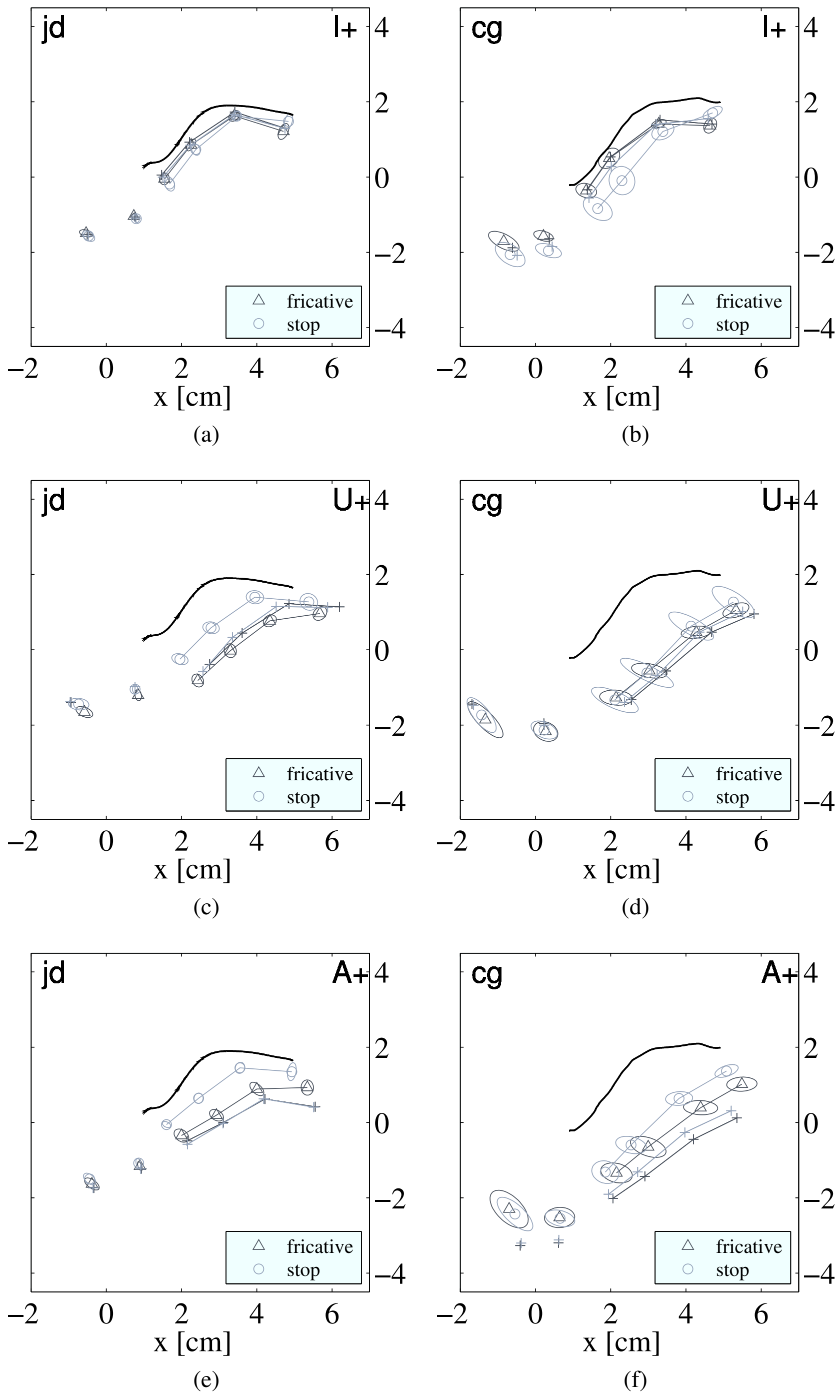
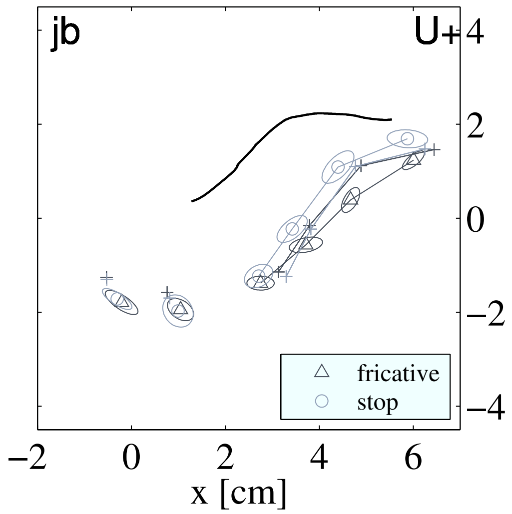
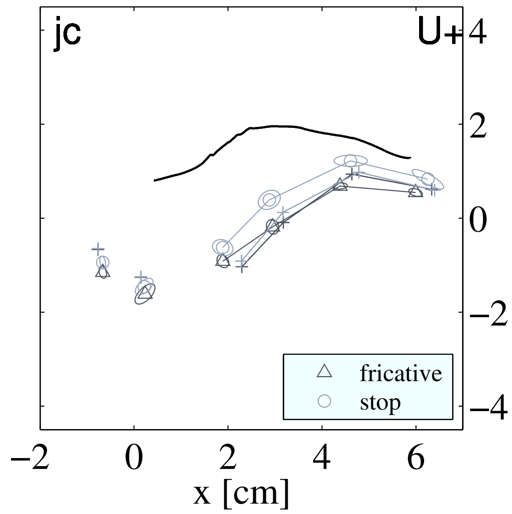


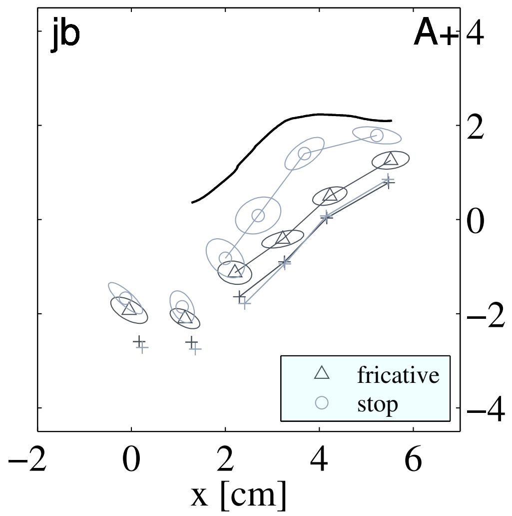
Figure 3.32.: Medial configurations for speakers jd (left) and cg (right). For further explanations see text.



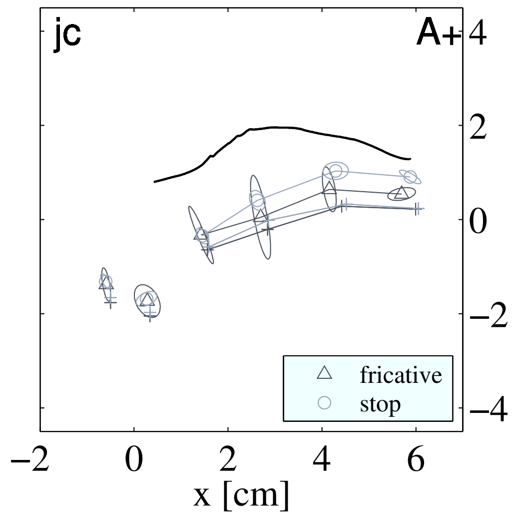
(a)



(b)



(c)



(d)

Figure 3.33.: Medial tongue configurations for the female speakers jb (left) and jc (right). For further explanations see text.

observations regarding sensor affiliations are in place for the female speakers jb and jc. They are displayed in figure 3.33. Regarding the stop configurations, there might be an interaction for tongue shape - and therefore presumably involvement in constriction formation - between speaker and vocalic environment. While for jb the TD sensor appears to be higher in the context of [aː] than in the context of [uː], the opposite seems to be true for speaker jc. Without putting too much emphasis on possible constriction affiliations - qualitative descriptions like these are always to some amount speculative - it still has to be noted that it might be wise to operationalize the hypothesis of more allphonically induced variation in several different ways.

10.2. The amount of allophonic variation

The different operationalizations for quantifying the amount of contextual allophonic variation for the stop versus fricative contrasts will quantify distances (i) between tongue shape descriptors expressed in Principal Coordinates ("factor loadings") and (ii) between coordinates in Euclidean space without prior decorrelation. Before arriving at a discussion, these analyzes will be conducted in separation.

These analyses were influenced by Ambrazaitis und John (2004) concerning the usage of Mahalanobis distances in order to determine similarities between front and back allophones. The Mahalanobis distance is a multivariate distance measure indicating the similarity of a data set from a single measurement. Its most widespread use is as the criterion minimized in linear discriminant analysis. The analysis in the present work differs from Ambrazaitis und John (2004) with respect to the usage of the dependent variable: While their study used EPG-indices (centres of gravity), the analysis presented here utilized x- and y-coordinates of the TB EMA-sensors to measure distances between back and front allophones in a comparable fashion. The results are shown in figure 3.34. While the contextual variation is consistently larger in fricative contexts in comparison to stops in the context of [uː], this finding is not consistently preserved in the context of the low vowel: While jb and jd indicate this pattern, speakers cg and jc do not. This is evidenced as well by paired t-tests: While it becomes significant in the context of [uː] ($t = 3.31$, $df = 3$, $p<0.05$), it does not reach the level of significance in [aː]-context ($t=1.23, p=0.30$). In principle the same kind of analysis, but with the dependent variable consisting in Euclidean distances between the different stop and fricative variants as derived from Principal Component Analyses is displayed in figure 3.35. This is a shape-based analysis, it is again displayed selectively for distances of tense [uː] and [aː] contexts from the high front vowel context. Restricting the analyses to tense vowel contexts is justified by expected larger distances between front and back variants for long vowel contexts in contrast to their short counterparts. The analyses this time are fully consistent, more allophonic than mere contextual variation is confirmed for both [uː] and [aː] contexts. Both t-tests are highly significant as shown in the upper right corners of figure 3.35 (contrast [aː]: $t=7.89$, $p<0.01$, [uː]: $t=8.49$, $p<0.01$).

The results displayed can best be seen as examples of the general trend obtained in these calculations: Although the tendency for the fricative allophones to display a larger amount of allophonic variation is a quite robust finding, the exact pattern to a substantial degree depends on the precise operationalization chosen: The "pure distance" could - and was in fact - extended to include several sensors, and the shape based methods in turn also can - and were also - limited to include only a subset of sensors. The patterns found point to the finding that the claim of more

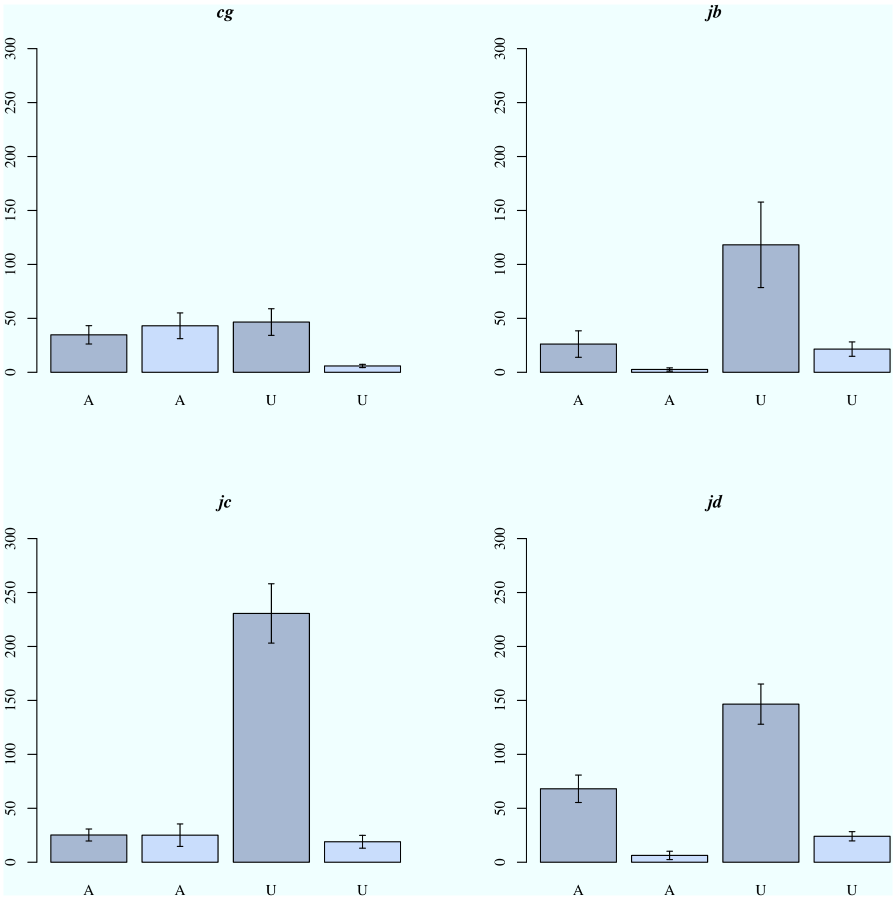
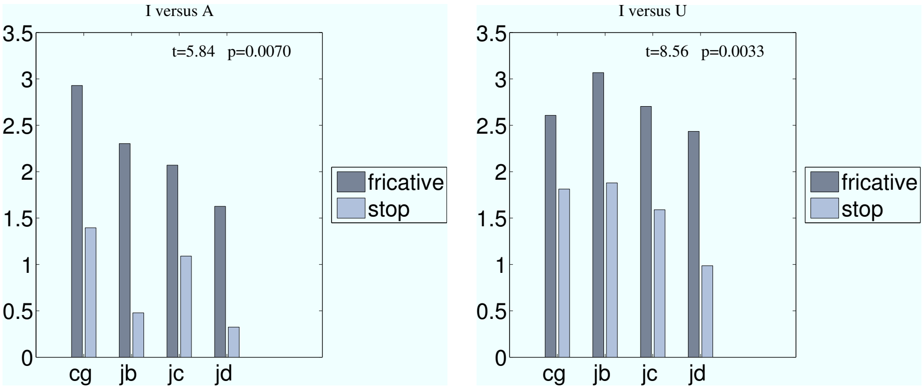


Figure 3.34.: Bar plot of Mahalanobis distances between postures of obstruents in back vowel contexts (tense vowels only: [uː and aː]) and articulatory postures of corresponding front vowel contexts ([iː] serving as reference in both [uː] and [aː] contexts). Postures are extacted at burst time for the stops and in the mid of the frication interval for fricatives. Darkgrey bars indicate fricative ([ç] and allophones), lightgrey bars velar (stop) contexts. The labels of the abscissa indicate vowel contexts.



(a) Contrast [aː] vs. [iː]

(b) Contrast [uː] vs. [iː]

Figure 3.35.: Bar plot of Euclidean distances in factor space, calculated as distances from articulatory postures of obstruents in back vowel contexts to their counterparts in front vowel (constantly [iː]) contexts. Postures are measured at burst time for stops and at the temporal midpoint of the frication interval for their continuant counterparts. Data in the left panel contrast [aː] vs. [iː] contexts, data in the right panel contrast [uː] vs. [iː] contexts.

contextual variation of the fricative allophones in comparison to the velar stop is substantial and valid, but that a fully consistent operationalization is not easily achieved.

10.3. Blending characteristics

The question about potential differences is already mentioned by asserting more contextual variation for the German dorsal fricatives than for the stops, which implies a similar, but even more extreme blending scenario for the fricative allophones than for the stops. Further substantiation, which parallels the organization of the presentation of results for the Hungarian dataset can be obtained by showing format ellipses of V_1 in the [CV_1Cɐ]-sequences. Note that the results are not directly comparable to the results obtained from the Hungarian sample data because the identity of C_1 is not matched: Whereas the identity for the initial consonant was varied for the Hungarian dataset - and therefore includes initial palatals - this was not possible for the German dataset due to distributional properties mentioned earlier. Still, the example given in figure 3.36 shows that there is no interpretable effect of the identity of the *following* consonant on the formant patterns of V1. For completeness, formant patterns are given in more complete form in yet another appendix (see appendix D).

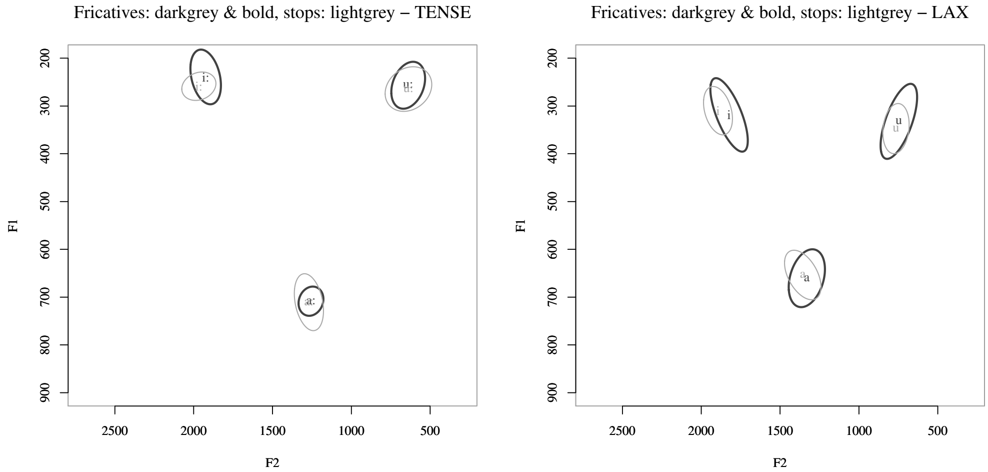


Figure 3.36.: The first two formants at the temporal midpoint of the corner vowels for the German data set. This figure displays the data for speaker cg. Bold, darkgrey lines indicate that the following sound is a palatal fricative /ç/ or one of its allophones (/x,χ/). Normal linetype in lightgrey refers to the contextual variants of the voiceless velar stop /k/.

10.4. Kinematic characteristics

Of more interest is the discussion of kinematic characteristics of the medial consonants. In section 9.5, correlations between the positions in the mid of the first vowel and distances traveled during oral closure for the tongue dorsum sensor were reported (see figure 3.31) for the Hungarian data. It was wondered whether the lack of convergence between results for velar stops as reported in Geng; Fuchs; Mooshammer und Pompino-Marschall (2003) and results for the Hungarian velars could be attributed to some substantial property of the implementation of the Hungarian velar. The new German data acquisition of the present study allows further testing and validation: Again, two of the three speakers in Geng et al. (2003) - jd and cg - are identical with the speakers of the current study. One might claim that this move might make some results redundant, but then again, explanations in terms of artefacts caused by the experimental situation like e.g. sensor placement are ruled out. Figure 3.37 shows the correlations between the positions in the mid of the first vowel V_1 and the distances traveled during the stops for tongue dorsum sensor. Again, similar tendencies were observed for the tongue back sensor, therefore only data on the tongue dorsum sensor are reported. Like in figure 3.31, the left panel shows the correlation of distance traveled and horizontal x-position and the right panel of distance and y-position. Recall the expected *positive* correlation between anteriority and the *negative* correlation between sensor height and the distances traveled during oral closure. Of particular interest are the velar stops again. The results for the "duplicate" speakers, jd and cg, almost exactly mimic the results as obtained in Geng et al. (2003) with respect to the signs of the correlations. This replicability rules out an explanation in terms of mere sensor placement. Also speaker jb perfectly conforms to this scenario. Only speaker jc shows only a moderate correlation between the horizontal position in the preceding vowel and the traveled distance (i.e. the left panel of figure 3.37). But still, no sign change of the correlation is observed for this speaker, only an attenuation. Further, this attenuation could be explained in line with data from a study by Munhall; Ostry und Flanagan (1991), who observed a reduced complexity in movement paths after algorithmic removal of the jaw. The patterns for the fricatives will not be discussed in depth, rather only minor remarks seem to be in place: Observed movement amplitudes during the fricative constriction are fairly small in comparison to the movement amplitudes during oral closure for the stops. This property also differentiates the German fricatives from the Hungarian palatal stops for which substantial movement amplitudes during oral closure were observed.[32]

10.5. Discussion

A more thorough discussion is not strictly necessary: The main results were already given in section 10.2 discussing the amount of allophonic variation. The difficulty for these analyses obviously consists in the derivation of a coherent operational definition for "more contextual variation", which can reasonably be defined in terms of shape descriptors like extracted by factor analysis or "blunt" distance measures like Mahalanobis distances. Still, the effect of more allophonic variation for the fricatives as compared to their plosive counterpart is fairly robust, but not fully consistent. Further recall that I am not willing to make major inferences from the overt speech behavior as observed in the current study to relationships concerning the "phoneme and the brain". This point was already highlighted in section 8.1 and is not repeated here. Starting as an a-priori assertion, the validity of the "full-fledged" blending scenario for the German dorsal fricatives was again illustrated by the formant patterns of the flanking vowels. These analyses were not discussed in full detail, because results were fairly unambiguous and further discussion therefore is unnecessary. The section on kinematic characteristics further substantiated results from Geng et al. (2003) on a - partly overlapping - sample and facilitates the transition to the presentation of cross-linguistic results: These results at least hint at a different implementation of velar stops in Hungarian, consider the tendency to exhibit the consistent and beforehand unattested tendency for backwards movements during the closure for at least one speaker in Hungarian and the severely disturbed patterns concerning correlations of movement amplitudes during the stop and horizontal positioning of the tongue in the preceding vowels - see figures 3.31 and 3.37 for Hungarian and German respectively. Until so far, it only has been shown that the question of differential motor implementation of velar stops in Hungarian and German is indeed promising: The next section will attempt to provide an answer.

11. Results III: Crosslinguistic analyses

Low- dimensional and speaker-independent linear vocal tract parameterizations can be obtained using the 3-mode PARAFAC factor analysis procedure first introduced by Harshman; Ladefoged

[32]Recall that the definition of the intervals was different for the German dorsal fricatives than for the Hungarian palatal stops. A methdologically cleaner analysis could be established in terms of velocity and accelerations signals.

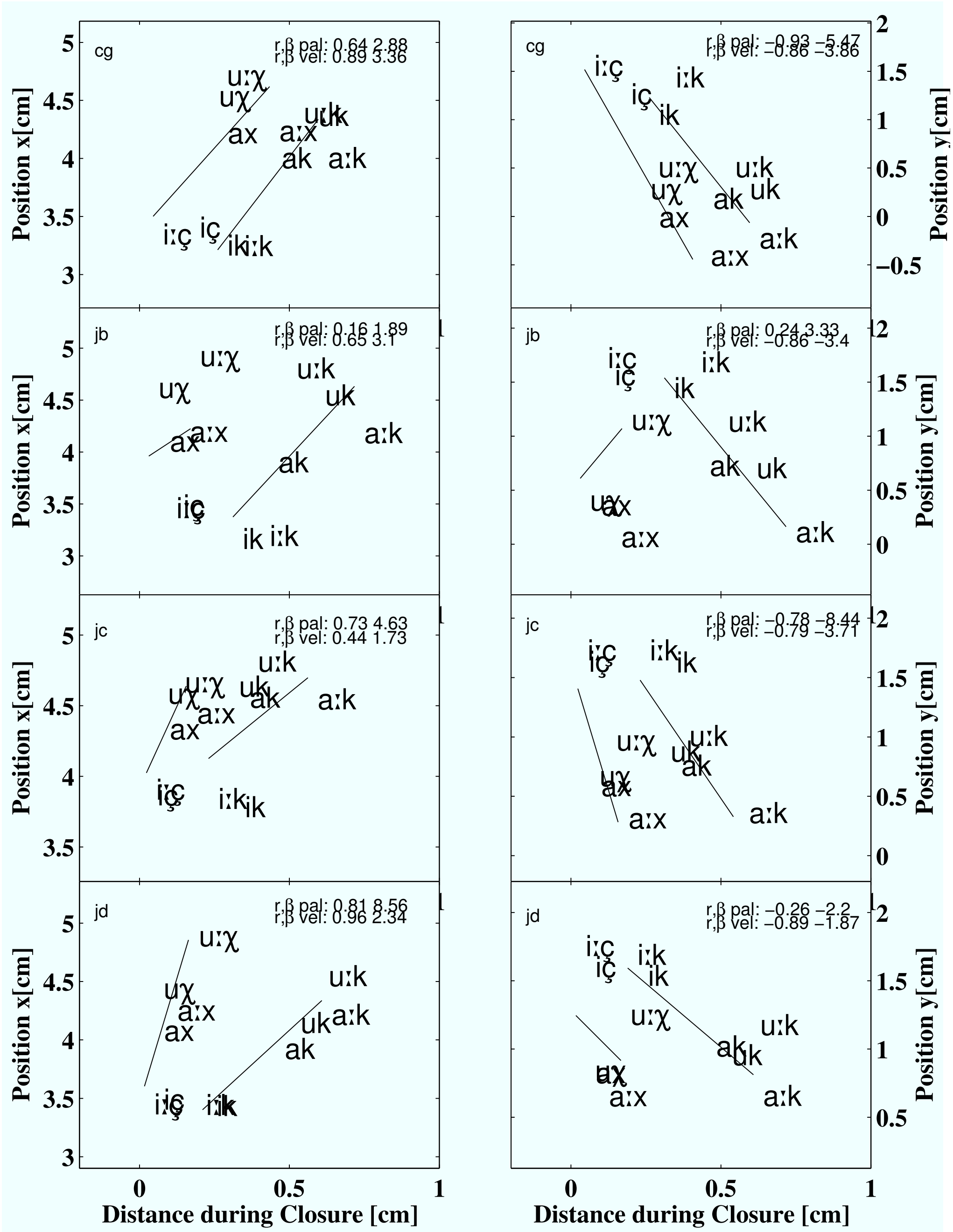


Figure 3.37.: Correlations between the positions in the mid of the first vowel and the distances traveled during the velar stops and the palatal fricatives, both for the tongue dorsum sensor. The left panel shows the correlation of distance and x-position and the right panel of distance and y - position.

und Goldstein (1977) and discussed in a series of subsequent papers in the Journal of the Acoustical Society of America (Jackson, 1988; Nix; Papcun; Hodgen und Zlokarnik, 1996; Hoole, 1999; Zheng; Hasegawa-Johnson und Pizza, 2003).

PARAFAC is a type of multi-mode analysis procedure and therefore contrasting with Principal Component Analysis (PCA) or factor analysis, which are two mode representations. PARAFAC requires an at least three-dimensional data structure with the third dimension usually being represented by different speakers, i.e. if all speaker weights are fixed to be one, then PARAFAC reduces to PCA. The advantage of PARAFAC is that there is no rotational indeterminacy as in PCA, in other words, PARAFAC gives unique results. The PARAFAC (in accordance with literature from now on called PARAFAC1) model can be written as (following Kiers et al., 1999, alternative notations are given in Harshman et al., 1977 or Nix et al., 1996):

$$X_k = AS_kV^T \tag{3.15}$$

where X_k is the kth "slab" of the input data matrix, with k the number of speakers, A is the matrix of articulator loadings, S is the diagonalized matrix of speaker loadings for speaker k and V the loading matrix for vowels. The matrix of articulator weights is held constant for each slab of the data cube, i.e. for all k speakers. This addresses Cattell's notion of parallel proportional profiles:

> "The basic assumption is that, if a factor corresponds to some real organic unity, then from one study to another it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study." (Cattell & Cattell, 1955, after Harshman und Lundy, 1984, p.151)

Another way to put the same idea:

> "Thus if speaker A uses more of factor 1 than does speaker B for a particular vowel, then speaker A must use more of factor 1 than speaker B in all other vowels. The ratio of any two speakers' usage of a given factor must be the same for all vowels." (Harshman et al., 1977, p. 609)

Fitting the PARAFAC1 to the data in the least squares sense amounts to minimizing

$$\sigma_1(A, V, S_1, \ldots, S_k) = \sum_{k=1}^{k} ||AS_kV^T||^2 \tag{3.16}$$

There is a unique solution minimizing (3.16) up to scaling and permutation. Cattell und Cattell's (1955) notion of "parallel proportional profiles" does not always have to be a valid assumption, it rather can turn out to be too restrictive in some cases. A less restricted algorithm is also at hand as shown elsewhere (Geng und Mooshammer, 2000). This less restricted algorithm, PARAFAC2, offers an attractive alternative, especially in difficult situations. PARAFAC2 can be expressed as

$$X_k = A_kS_kV^T. \tag{3.17}$$

Within PARAFAC2, each loading matrix for the articulators, A_k, is expressed as $A_k = P_kA$. P_k is an $I * R$ matrix, where R denotes the number of factors and I the number of measurements in the articulator domain. A is constant over all these individual profiles and of size $R*R$. The rotational freedom provided by the PARAFAC2 model is adequate for approximating certain deviations from the strict linearity required in PARAFAC1. PARAFAC2 incorporates an invariance constraint on the factor scores as a milder version of factorial invariance: The cross-product matrix $A_k^T A_k$ is constrained to be constant over k speakers. The model structure is determined by the choice of the structure of A_k. Bro (1998) compares PARAFAC2's flexibility in this respect to Procrustes analysis. In Geng und Mooshammer (2000) we have shown that the strict assumptions required in the classical PARAFAC1 model were too strong to capture stress-specific variation in full detail. In contrast, PARAFAC2 allowed to account for systematic variation produced by word stress by imposing this weaker structure on the data. In particular, PARAFAC2 modeled the physical properties of the vocal tract shape in a more realistic and plausible way with respect to the description of mean factor shapes. These results might prove useful especially when cross-linguistic comparisons are required: The speaker-specific articulatory profiles consist in an additional device where additional cross-linguistic variation can be "stored" which has prooved too detailed to be captured by the more rigid PARAFAC I model.

Recall that the research question concerning crosslinguistic differences was put as the question of a "less palatal realization of the velar in front vowel context for Hungarian" in section 7.3. On a more descriptive level, the question of interest is whether postures of the German velars show differences in comparison to the Hungarian velars at all, as mentioned with special emphasis on the front vowel contexts. The input data to PARAFAC will be the mean values of the four tongue coils aggregated over consonantal category, vocalic context and position in the CVCV sequence. Not all data were selected for fitting the models though, due to different reasons: Palatal stops for Hungarian and the German dorsal fricatives were omitted throughout the first analysis as well as all short vowels. For the consonants, the design is not balanced concerning the identity of the initial consonant; as mentioned, the German corpus contains no regular initial fricatives and these were replaced by the corresponding velar stops. Furthermore, dorsal German fricatives exhibit different blending properties than Hungarian palatals and therefore should not be compared in vowel contexts other than the front vowels. Further, the omission of the short vowels is motivated by the fact that the vowel systems are not directly commensurable, for example the Hungarian short low vowel has a pronouncedly different quality than its German short lax counterpart. Therefore the dataset for the question about potential crosslinguistic differences between the velar stops contained (i) the long corner vowels and (ii) the initial and medial velar stops of both languages in the context of the high front vowels. Inputting vowels for an analysis targeting consonants might seem awkward in the first place, but is motivated by the aim to mimic the antagonistic functioning of the extrinsic tongue musculature in the data as already shown in the introductory part while discussing statistical approaches to model tongue physiology (see section 2.2). Summarizing, input data for this particular analysis consist in five different phonetic entities (three vowels in position V_1 and two velar consonants in initial and medial positions) acquired from eight different speakers, speaking two different languages. These were measured with respect to positions of the four tongue sensors with an x- and a y-dimension each. The data from the different target languages were treated as different speakers, i.e. language identity was not seen by the algorithm and the data were formatted as a 5 x 8 x 8 (4 Hungarian + 4 German) matrix. Before the data are seen by such algorithms, they usually are preprocessed by subtracting mean articulatory shapes of each speaker from the data. This operation was also performed here.

The first step consisted in conducting a PARAFAC1 analysis with two extracted factors. It turned out that this attempt failed to converge to a robust model.[33] Therefore an orthogonality-constraint PARAFAC2 model was fitted to the data which amounted in a decent fit, and two factors were extracted. These two factors accounted for 97.8% of the variance, with the first factor accounting for 19.4% and the second factor accounting for 78.5% of the variance. This high explained variance is of importance per se if compared to the results obtained by the just mentioned invalid PARAFAC1 solution (73%):[34] This high gain of fit by simply altering the choice of the algorithm is quite unusual held against the background of earlier comparisons (Geng und Hoole, 2005; Geng und Mooshammer, 2000). This finding could imply that the added speaker-specific profiles of PARAFAC2 possibly contain information with regard to crosslinguistic differences.

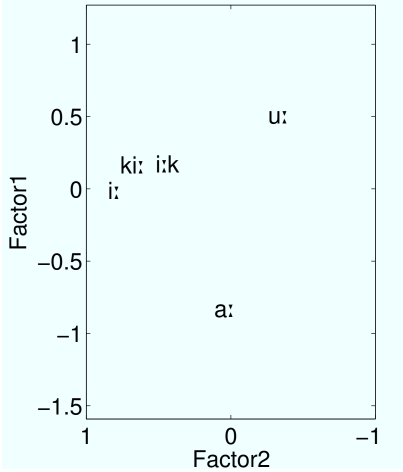
Within phonetics, the fitting of such models usually results in the display of (i) linguistic units in factor space, usually called the "vowel space", surely a "consonant-vowel space" is the more appropriate notion in the present context, (ii) the corresponding display of the "speaker weights" and (iii) the display of the effect of the factors on the articulatory coordinates, calculated by imposing the shape effects extracted by the factors on the mean configuration. This presentation is also chosen in figure 3.38. The combined consonant-vowel space in figure 3.38a shows the expected pattern: The initial velar (from [kiː]) displays a configuration closer to the high front vowel than does the medial fricative which more substantially coarticulates with the following low vowel. Of more interest is the inspection of the speaker weights: Neither Factor 1 on the y-axis nor Factor 2 on the x-axis in figure 3.38b linearly separates between the Hungarian and the German subset of speakers (which are enclosed by darkgrey circles and lightgrey squares respectively). This factor structure corresponds to the tongue projections in figure 3.38c: Factor 2[35] resembles the classical "front-raising" as first obtained by Harshman et al. (1977). The more relevant benchmark for the present purpose is probably Hoole (1999) because of the use of the same EMA fleshpoint technique which also was applied in this study. And indeed, Factor 2 as obtained here is quite close to Factor 1 as obtained by Hoole (1999). Larger differences in comparison to Hoole (1999) emerge from the patterning of the first factor obtained here. It still is close to "back raising" in a superficial sense but the movement component captures less the backing and more the raising component of this construct in comparison to e.g. Hoole (1999). This likely is due to the presence of the velar consonants present in the current analysis.

Of course, the mere structural description of the solution obtained is per se not of any relevance for the question of cross-linguistic differences in the implementation of the velar stops postures. As mentioned above, the PARAFAC1 approach resulted in degeneracies. Furthermore, the only device in PARAFAC1 is the matrix of speaker weights which could account for the cross-linguistic posture differences between German and Hungarian velars as shown by $X_k = AS_kV^T$ in equation 3.15. Compare this formula to the notation PARAFAC1 with

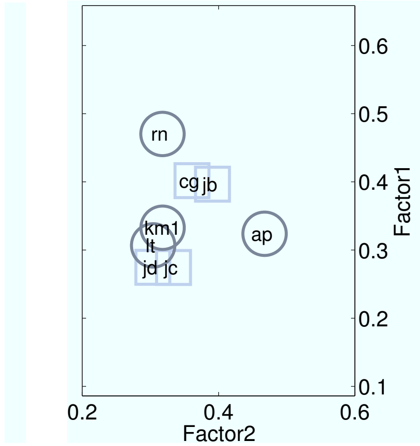
[33] Which was diagnosed by the "core consistency" diagnostic amounting to roughly 65% indicating deviance from strict trilinear structure. For more background on this measure see Bro und Kiers (2003).

[34] Note that the amount of explained variance is not affected by the degeneracy, which rather concerns the factor structure obtained, with degenerated solutions exhibiting highly correlated factors.

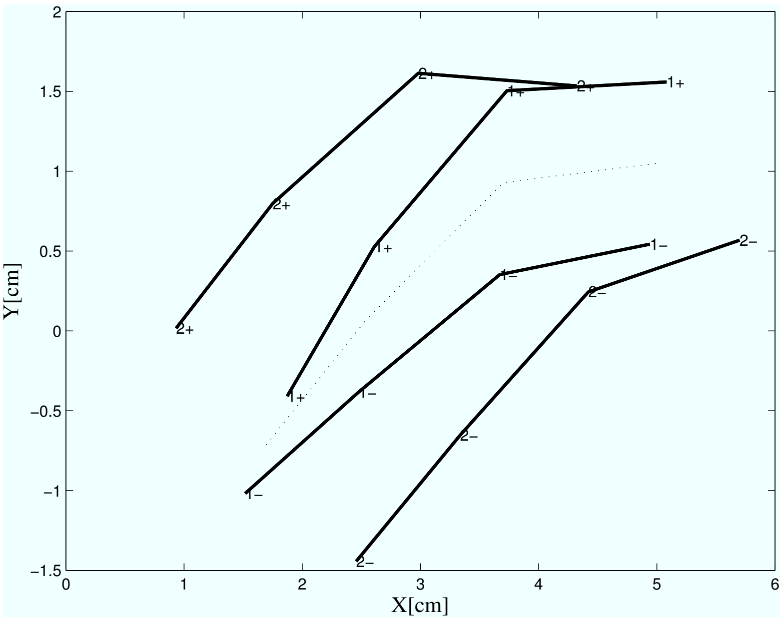
[35] While factor order matters in Principal Component analysis with the first factor always capturing most of the variance, this is not the case for PARAFAC matters, where the order is arbitrary.



(a) consonant-vowel space



(b) speaker weigths

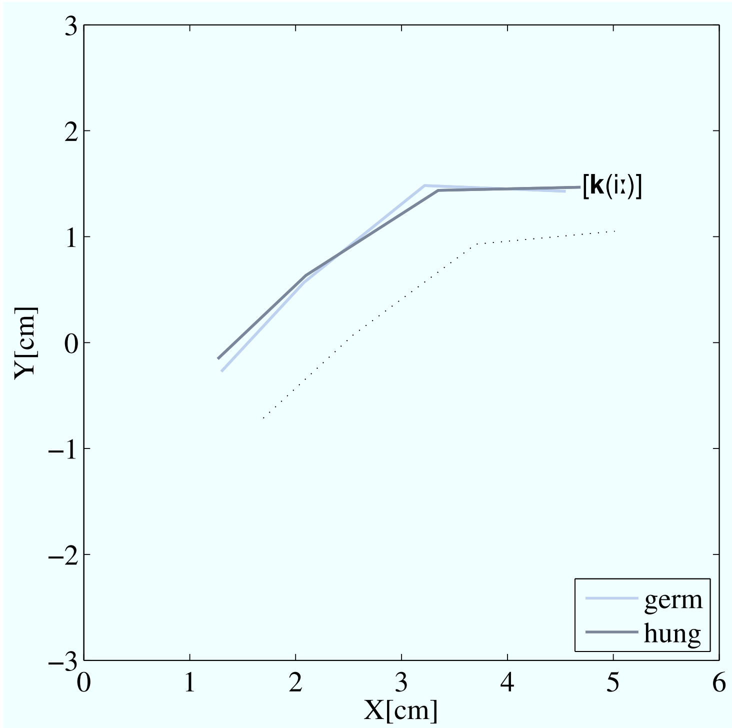


(c) factor shapes

Figure 3.38.: PARAFAC2 solution for the crosslinguistic analysis comparing velar stops in front vowel contexts. Top left panel: projections of linguistic objects into factor space. Top right panel: speaker weights, dark circles encoding Hungarian, lightgrey squares German speakers. Bottom panel: Tongue shapes related to the factors of the two factors of the model. Displacement using mean speaker weights from mean tongue position (also shown as dotted line) caused by setting each factor in turn to ± 2 standard deviations. As an example: 1+ means the shape associated to the positive pole of Factor 1. More anterior locations are to the left.

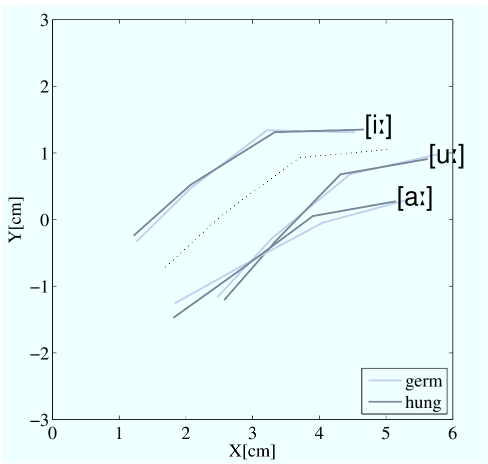
$X_k = A_k S_k V^T$, where the greater flexibility is achieved by altering the matrix A to now containing subspaces for each speaker separately in the A_k. Recall that each of these A_k is expressed as $A_k = P_k A$, i.e. A is constant over all speakers. Now the articulatory shapes are stored in their articulatory profiles, the P matrices, separately for each speaker. This property allows, in contrast to PARAFAC1, to incorporate nontrivial shape changes which might prove useful to capture as was already shown in Geng und Mooshammer (2000). In other words, it has a device for "putting someone's tongue in someone else's mouth". This works as follows: Separate A_k matrices are calculated for the German and the Hungarian subsets of the analysis by taking the mean of articulatory profiles separately for each language, i.e. separate $A_{Hungarian}$ and A_{German} are derived by premultiplying the $P_{Hungarian}$ and the P_{German} with the matrix A common for both languages. In a similar vein, the mean speaker matrices could also be pooled over the distinct languages. This will not be undertaken, rather the mean over all speaker weights will be used for the construction of language-specific profiles. Given the fact that the speaker weights of the factors extracted do not separate between the two languages (0.37 and 0.35 for Factors 1 and 2, German subset, 0.36 and 0.35 for the Hungarian subset), the configurations are not expected to be substantially different. Finally, and maybe most important, the mean configurations onto which these profiles are "superimposed", is the mean configuration for the whole dataset, i.e. no separate mean configurations for the two languages are used.

Figure 3.39 shows these projections. Hungarian configurations are always drawn in darkgrey and the German ones in lightgrey. The top panel (3.39a) compares initial velar projections for both languages. The pattern suggests a meaningful result: The rear part of the tongue, or more precisely the two back sensors are in fact retracted for the Hungarian subset. Furthermore, the rearmost sensor is higher for the Hungarian than for the German projection. This could point to an actual constriction location further back for the Hungarian than for the German data. Another interesting feature of these configurations is that the parts of the tongue obviously not actively involved in forming the constriction - i.e. the laminal and predorsal regions - seem to be higher for the Hungarian than for the German projection. Apart from the objection that these projections are based only on four speakers per language, - which is actually trivial and valid for almost all articulatory movement studies in this field - there might be other, more serious objections against such a display. For example, one might explain the more retracted [kiː] for Hungarian as a phenomenon of the whole articulatory space. That this is not the case is evidenced in figure 3.39b which shows the equivalent partial projections for the corner vowels: Here, the relationship is reversed for [uː] and [aː] contexts while a similar pattern is maintained for the palatal vowel. Another, in some sense more specific critique could argue that all high front configurations are affected and shifted backwards. But reconsider the consonant-vowel space in figure 3.38a: The solution also contains the medial configuration, and the configuration shown as [iːk] still is in the vicinity of the the front vowel and the initial velar stops rather than close to the back vowel. The configurations for the medial velar stops for both languages are shown in figure 3.39c. The crosslinguistic comparison here yields no interpretable difference, on the contrary, it is completely neutralized. Therefore, the velar backing for the Hungarian subset is a specific of the initial velar, not of all high front configurations. A further issue also underpinning the validity of the solution is illustrated in figure 3.40. It shows the initial velar stop data as already displayed in figure 3.39a and the high front vowel as also already shown in figure 3.39b together. It becomes evident that the stop configurations are modelled as more

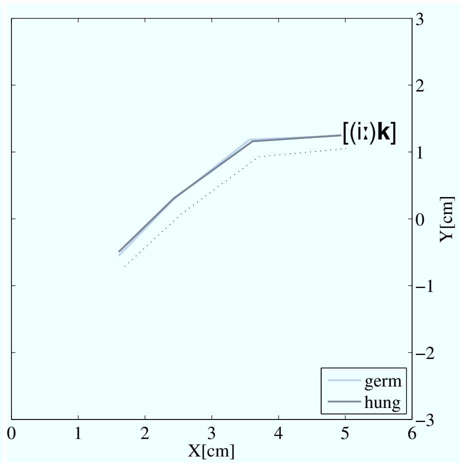


(a) k(iː)

(a) k(iː)



(b) corner vowels



(c) [(iː) k]

Figure 3.39.: Partial projections in the articulatory space of a modal speaker. Darkgrey projections encode Hungarian, lightgrey German data.

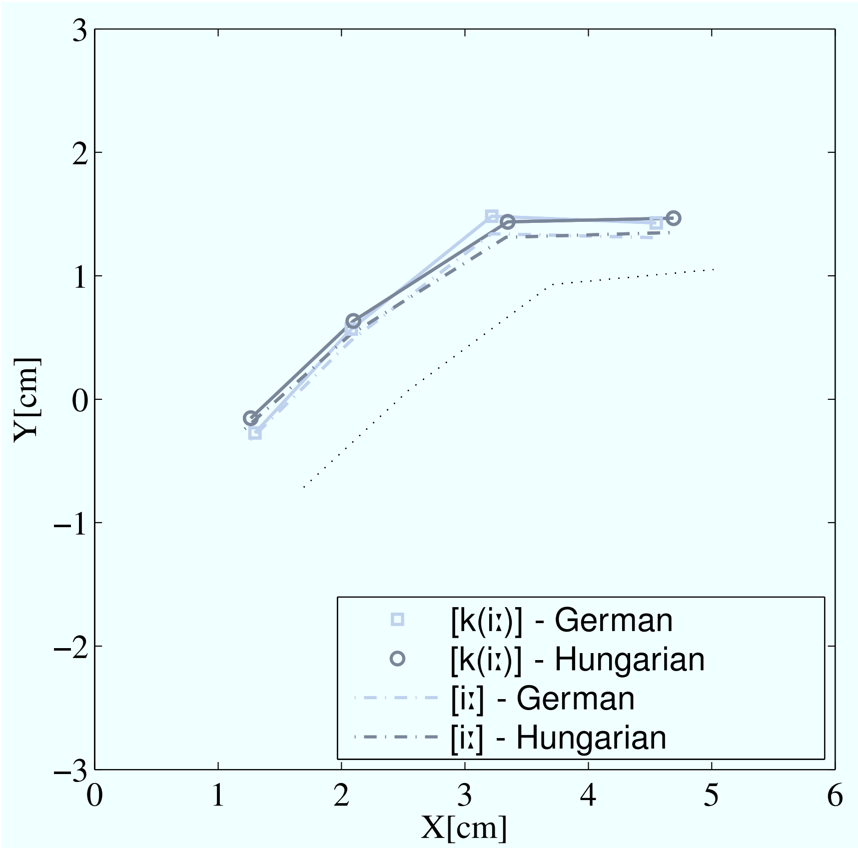


Figure 3.40.: Projections of the initial velar stop in front vowel context and the configurations for [iː]. The darkgrey configurations are for the Hungarian subset, lightgrey projections for the German subset again.

constricted than the vowels which conforms perfectly to phoneticians' expectations. Further, and less trivial, is that there are differential effects of consonant versus vowel in both languages implying that the procedure not just does model one general high front configuration, but that it encodes substantial aspects of tongue postures. A final demonstration of the plausibility of the projections is shown in figure 3.41. In order to generate this figure, an equivalent model to the one just described was fitted. A more thorough model description is omitted, because the results of the fitting process were fairly similar to the solution just discussed in greater depth, i.e. a hard degeneracy for PARAFAC1 and a fully functioning solution with about 98% of explained variance for PARAFAC2 was obtained. The only difference consists in the selection of input data, more precisely, the velar configurations were excluded and replaced by the palatal fricatives after [iː] for the German data and the palatal stop for the Hungarian subset. The results again conform to phonetic expectations: The speaker-independent stop configuration for the Hungarian data is substantially more constricted than the fricative configuration for the German dataset, the constriction location appears to be relatively stable in contrast.

11.1. Discussion

Taken together, the patterns obtained are phonetically interpretable and meaningful, and therefore there are no principled reasons why the velar backing effect of stops in the context of front vowels observed for Hungarian in comparison to German should be invalid. The objection that the results of this study are biased by physiological artifacts and artifacts due to tract geometry, because the number of subjects is too small to assert that these influences are indeed reduced to random fluctuation is of course valid. But still, it is possible to argue against such claims adopting a methodological argumentation: The algorithms used have no access to the mean shape which one can hope to remove a fair amount of morphological information from the raw data during preprocessing. Further, more sophisticated preprocessing schemes could potentially be applied which better respect more issues like the length of the vocal tract. This was suggested for the PARAFAC family by Zheng; Hasegawa-Johnson und Pizza (2003). Our own previous evaluation of these ideas (Geng und Hoole, 2005) suggested only minor relevance of preprocessing which was evidenced by correlating solutions on identical data sets but using additional scalings in the sense of Zheng et al. (2003). The correlations between the factors typically were close to one. A further methodological point which can be made against the objection of morphologically biased study is that PARAFAC2 was highly successful in comparison to PARAFAC1, providing almost perfect fit to the data and a non-degenerate solution: PARAFAC2 incorporates a Procrustes-like notion to achieve its flexibility, and therefore uniform changes in the morphology are unlikely to affect the results. Therefore, the results are in my view substantial and interpretable with respect to all aspects discussed.

For example, the introduction discussed the view that palatal fricatives such as [ç,ʝ] are treated as noncoronals in phonology, and that on the contrary, the existence of noncoronal palatal stops is universally denied. Figure 3.41 contrasted these sounds in a single display and suggests that these classifications have no strong phonetic basis.[36] This result is only seen as a spin-off though. Anyhow, the main question of this section had been whether velar stops in front vowel contexts

[36] However, such projections are only meaningful for front vowel contexts and in medial position, due to the blending and distributional properties of the German dorsal fricative.

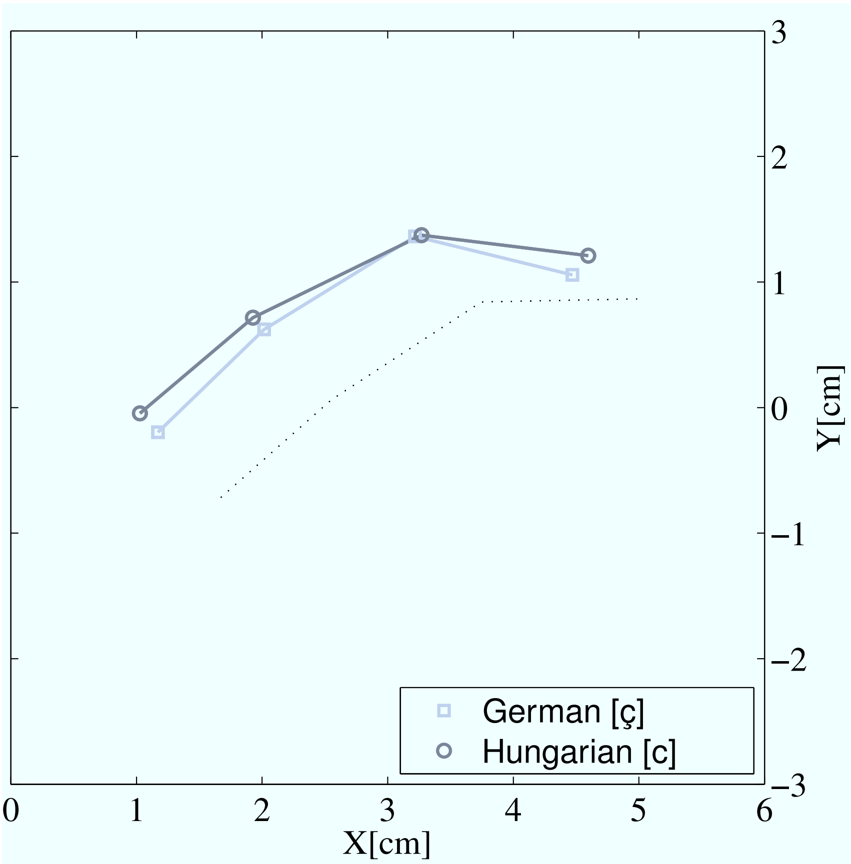


Figure 3.41.: Projections of the medial palatal fricative (for German) and the medial palatal stop (for Hungarian), both after the high front vowel. The darkgrey configurations are for the Hungarian palatal stop, the lightgrey projections for the German dorsal fricative.

are retracted before front vowels in order to avoid the clash with the palatal stop. In my view this clearly is the case. Instead of speculating on the phonetic purpose of this behaviour and its origins right in this place, it seem wise to proceed to a more general discussion of the current work as a whole, because a discussion on the origins of this behaviour has to involve broader perspectives and cannot reasonably be limited to the articulatory domain.

Part III.

General discussion

4. Summary of the main results

1. Perceptual experimentation

Perceptual experiments were carried out with synthetic CV stimuli with manipulated onset frequencies of the second and third formants. In the experimental design, great care was taken to not allow for contextual adjustments made to vocalic context, because vocalic context is known to interact with the relative weightings assigned to transitional patterns on the one hand and burst properties on the other hands. These weighting differences could have seriously hampered cross-linguistic comparisons. Therefore, the target of the CV sequences in this study was equated to be the neutral vowel.[1] These stimuli were presented in a CP paradigm with combined identification and discrimination (AX) tasks to about 20 French and Hungarian listeners each as representants of "three- and four category languages". In the identification tasks, French listeners had response alternatives corresponding to labial, alveolar and velar places of articulation, while Hungarian listeners hat an additional fourth alternative corresponding to the palatals.

The results indicate that first of all for the Hungarian listeners, the category boundary palatal-alveolar was not as robust as the remaining boundaries. This was evidenced by signs of degeneracy of the CP framework, i.e. the traditional CP procedure of predicting discrimination performance from identification performance did not result in a very prominent instance of the famous simple structure obtained in numerous studies.[2] Rather, results suggested the presence for perceptual conflict between alveolar and palatal regions for the Hungarian subset which was demonstrated by noisy identification behaviour. Second, if boundaries for these regions were lumped together, the result was a topology fairly similar to the topology of a language with three places of articulation only - like French in the current study. Third, another essential result which was revealed by a comparison of the territorial mappings of the two languages under consideration is that the velar region is shrunk in the Hungarian sample in comparison to the French sample.

2. Articulatory study

The articulatory studies were in the first place designed to measure the degree to which contextual variation in velar stop production can be influenced by the system of phonemic contrasts. Put the other way around, the leading question initially was whether the most parsimonious

[1] This move is not ad hoc, the neutral vowel is known to have a prominent development role Kent und Murray (1982) and additionally is often taken as the neutral reference in extrinsic speaker normalisation procedures. For example, according to Adank; Smits und van Hout (2004), such a neutral reference frame accounts for listeners' life-long experiences with listening to different types of speakers (Adank et al., 2004, p. 3105).

[2] Which requires aligned predicted and observed discrimination peaks, see section 2.3.

explanation - which would claim that the degree of contextually induced variability for velar stops is a mere side effect of the physiological periphery - is tenable. Concerning the origin of contextual velar variation and the looping patterns in particular, several hypotheses were put forward: Looping patterns have been conceived as a passive forward movement of the tongue due to airstream mechanisms (Kent und Moll, 1972), as a result of an active gesture aiming at the maintenance of voicing (Houde, 1968), synonymously 'cavity enlargement' (Ohala, 1983). Findings by Hoole et al. (1998) were in favour of the airstream mechanism hypothesis. These authors contrasted normal versus ingressive speech, with ingressive speech resulting in size reduction of the - forward - looping patterns. Therefore aerodynamic influences seem to be at work, but it is neither clear when and how they operate, nor can they alone account for the data on loops. Löfqvist und Gracco (2002) tried to explain looping patterns in more general principles of motor control, postulating the entire movement to be planned in terms of cost minimisation principles. In a recent modelling study, Perrier; Payan; Zandipour und Perkell (2003) focused on tongue biomechanics moderated by place of articulation. In contrast to Löfqvist und Gracco (2002), Perrier et al. (2003) conclude that looping patterns can be explained in terms of biomechanics alone and the trajectory as a whole does not have to be preplanned. Theories which emphasise linguistic factors were also described in the current dissertation: Articulatory Phonology[3] which explains velar coarticulation in terms of their blending concept, and the reasoning of Keating were described. The latter models velar coarticulation as an outcome of underspecification for the feature back. In order to make the hypothesised effect of the system of phonemic contrasts testable, a language which contains neighbouring palatal articulations which have the potential to influence the coarticulatory patterns of the velars was required. According to suggestions by Keating und Lahiri (1993) Hungarian was chosen. The study of the Hungarian palatal in isolation was a necessary first step. This topic then constituted a substudy in its own right. Note that the term "substudy" is not used with any pejorative connotation - the isolated treatment of the Hungarian palatals and velars is informative per se, for example with respect to sound change and to approaches at the Phonetics-Phonology-Interface like Keating's underspecification approach. Still, an understanding of Hungarian dorsal articulations can also be seen in a more instrumental fashion - for arriving at a judgement whether the crosslinguistic research questions concerning the velars are at all testable with Hungarian data. Taken together, the presentation of the Hungarian substudy were structured in order to answer several questions: First, is the Hungarian palatal the desired mediopalatal articulation in the terminology of Recasens (1990)? Of course, the EMA movement data as acquired here do not provide explicit contact information as does palatography. This problem was - at least partly - overcome by qualitatively describing single speakers' articulatory profiles and resulted in a confirmation of a dorsopalatal implementation for the Hungarian stops. A further important result is that the Hungarian data suggest that the fronting of velars can indeed converge with palatal articulation. This was the case for three out of four speakers in the current study where the initial velar before the high front vowel was dissolved in the agglomeration of palatal articulations. Further, the results concerning movement amplitudes of the palatal and velars in Hungarian do not conform to an interpretation of an "underspecification scenario" for the velar, i.e. the "transparency effect of the velar with respect to backness": The presence of partly very large amplitudes - during the whole VCV movement

[3] AP does not make detailed predictions on looping patterns.

and also during closure - for the palatal stops which has been specified as a complex segment in this approach does not favour such an explanation. A positive side-effect of the relative similarity of the Hungarian palatal with the velar is that this sound has the *potential* to limit the coarticulatory behaviour of the velar, and that the aspired crosslinguistic comparison of velar stops is a reasonable undertaking.

The intralinguistic research question concerning the German obstruents was whether the fricatives characterised by two resp. three allophonic variants show a greater amount of allophonic variation in contrast to the German stops where no such allophones are postulated for Standard German, and the contextual variation is captured by a single velar stop phoneme. As a precondition, it had to be justified that the German dorsal fricatives obey to the "rough" blending scenario in the original task dynamics sense. This clearly was the case as evidenced by the vowel formants of V_1 of the /VCVɐ/ sequences. A more essential point though is the result that the amount of contextually induced variability was in fact greater for the fricative allophones than for the stops - in terms of two different operationalizations of articulatory distance.

As mentioned above, in some sense, the core question of the articulatory studies of the current work was the crosslinguistic comparison of velar stop realization between languages without a palatal stop in the inventory like German and a language with such a "perturbing" phoneme. The current work proposed a method for crosslinguistic shape comparison based on an elaboration of our own previous work (Geng und Mooshammer, 2000; Geng und Hoole, 2005) which showed that there is in fact such an effect of palatal "perturbation" on the production of velar stops in the sense that the Hungarian velars are somewhat retracted in comparison to their German counterparts. Of course, formant frequencies depend on the area function of the whole vocal tract (see also sections 2.1), so a retraction of the Hungarian velar cannot count as strong evidence for such a retraction as for example being guided by "acoustic goal regions". Furthermore, there were cross-linguistic differences in the correlation structure between the distances travelled during closure and the horizontal and vertical positions during the preceding vowel. While correlations for German vowels were quite consistent - and also robust over different acquisition sessions - this consistency could not be achieved for the Hungarian dataset. This indicates that in German the movement during the velar stop is mainly determined by the place of articulation of the preceding vowel while this is not so consistently the case for Hungarian.

5. Conclusions

As mentioned in earlier sections, much of the experimental work in this dissertation can be characterised by multiple interpretability in terms of ideas and hypotheses triggered from distinct theoretical sources. This will be accommodated by modularity of the concluding discussion: The first section of the following concluding remarks will discuss the phonetic properties of the Hungarian palatal stop segments as recorded in the articulatory study and summarise the most prominent results in the context of their relative infrequency in the sound inventories of the world's languages. Another recurrent issue in the current work has been concerned with the match of phonological and phonetic representations. A summarising section will be devoted particularly to this problem and another one to the implications which arise from the crosslinguistic results obtained in the present dissertation. This latter section will digress most from the material covered experimentally and touch issues in the acquisition of coarticulation. Further note that the division between articulatory and perceptual modalities will be abandoned from now on in favour of a more aggregate perspective. Where appropriate, possible future research directions will be proposed.

1. Sound inventories

This dissertation was written in the context of a research project with the title "Articulatory economy and perceptual discriminability". Therefore it seems natural to also judge the status the Hungarian palatal in these terms. The results are fairly clearcut: A segment

(i) which perceptually is in conflict with the principal places of articulation and competes with them on portions of the stimulus space,

(ii) which has a strong syntagmatic impact on neighbouring segments such that essential cues to the paradigmatic contrasts of neighbouring (vocalic) segments might be fully obscured,

(iii) which, depending on vocalic context, requires the articulators to travel large distances, and finally,

(iv) which attempts to usurp phonetic space usually populated by the front allophones of the velar stops in turn leading to their articulatory reorganization

can hardly be considered as optimal with respect to articulatory economy and perceptual discriminability. On the contrary, such a segment is likely to be avoided in the sound inventories of the world's languages.[1] These findings also further explain the reservations many researchers had against the standard feature geometric representation of velar palatalization requiring a palatal

[1] which actually is the case.

as an intermediate state in a process changing /k/ to /tʃ/ (see figure in 3.6). This might not only have - at least partly - provoked the emergence of alternative feature geometries (as in Lahiri und Evers, 1991, figure 3.5), but also the more recent frequent occurrence of publications relying on perceptual mechanisms driving sound change (e.g. Guion, 1996, 1998). This is not to be misunderstood as my personal theoretical perspective, rather intended as an explanation for the recent interest of phonology in perceptual issues. On the contrary, the gradual velar softening hypothesis (in the sense of Recasens, 2003) is not ruled out by the data presented, rather the convergence of velar fronting with the palatal class conforms well to such "neogrammarian" ideas. Still, this is not the place to decide on causal mechanisms for certain sound change mechanisms. The present work rather was data-driven and empirical in spirit. Therefore making some suggestions for further empirical research seems to be more in place.

Little is known about the functional division of the genioglossus. While the genioglossus is a single muscle morphologically, speech production researchers have argued for a functional division into at least two motor units. This move has also been made by speech modellers, e.g. the model by Dang und Honda (2001) divides the genioglossus into three different functional levels.[2] The only study I am aware of tackling this problem experimentally was a study by Kakita; Hirose; Ushijima und Sawashima (1976). These authors studied muscular activity of the tongue by five different insertions in the genioglossus and confirmed a selectively decreased activity for the posterior fibers of the genioglossus for the /i-j/-contrast in Japanese. This study to some extent also constitutes an ideal starting point for also testing hypotheses about gradual velar fronting with a stronger explanatory claim. In my view, it could be informative to repeat such a study with material similar to the one recorded for the current work, and results might lend further articulatory plausibility for the relative infrequence of this articulatory class in the sounds of the world's languages.

2. Phonology and phonetic manifestations

A second major issue is the question whether different levels of phonological and phonetic manifestations - phoneme, allophone and pure phonetic variation - must have its reflexes in phonetic correlates. In the current work, this issue is most clearly reflected in the comparison of the German dorsal fricative allophones with the contextual stop variants. The fit of phonologist's intuitions was quite good, e.g. Wiese's claim is substantiated by the observation that there is indeed more variation for the fricative series. But what if these results were generated by a genuinely phonetic mechanism? For example, the velar fricative could be produced "more comfortably" in back vowel contexts at the soft palate than the velar stop. It could be that it is easier to make a critical constriction at the soft palate than a completely sealed closure. Maybe this could also account for why [x] may become uvular while the back stop allophone of /k/ never shows this behaviour. Still, *any* defensible alternative hypothesis puts the relationship between phonological entities and phonetic behaviour into question. It should be in order to quote An-

[2] again motivated pragmatically: for most physiological tongue models, it was impossible to raise the tongue tip and have a flat shape in the velar region without introducing a three-fold functional division (P. Perrier, personal communication).

derson (1981) from an earlier section (section 6.5) again refuting the association of fricative assimilation with different articulatory control regimes:

> "On this view, it is still very much part of the business of phonologists to look for "phonetic explanations" of phonological phenomena, but not in order to justify the traditional hope that all phenomena of interest can be exhaustively reduced this way. Rather, [...] the reason is to determine what sorts of facts the linguistic system proper is *not* responsible for: to isolate the core of features whose arbitrariness from other points of view makes them a secure basis for assessing the properties of a language faculty itself." (Anderson, 1981, p. 497, after Fowler, 1983, p. 314)

Taken together, I view the findings concerning these variabilities as purely correlational in nature. They still are valuable findings, further validating results by Pompino-Marschall und Mooshammer (1997) and Ambrazaitis und John (2004) by means of different acquisition techniques and different operationalizations of articulatory distances. But the fit of phonologist's intuitions with phonetic data is not necessarily as impressing as for the data on phonemic versus allophonic conditioning of phonetic surface behaviour just presented. In the same vein one could also phonetically test feature specifications. One such instance was mentioned very early in this dissertation, in section 1, when palatal fricatives such as [ç,ʝ] were reported to pattern phonologically as noncoronal, while the palatal stops [c,ɟ] were classed as [+coronal]. Data on this contrast were presented more or less as a spin-off while validating cross-linguistic factor analyses in section 11. Figure 3.41 compares the German medial palatal fricative with the medial palatal stop in Hungarian in comparable vocalic environments. It is hard to conceive that these cross-linguistic projections reflect a featural difference in feature specifications for [coronal] or [dorsal]. By contrast, an AP interpretation in terms of higher constriction degree for the Hungarian stop in comparison to the German fricative and identical constriction locations for both obstruents seems to be much more in line with these articulatory results. Taken together, phonetic patterns need not converge with phonological ones, and even in the case that they do, this does not warrant a tight bond to motor planning processes. This statement is not to be confused with naive agnosticism, rather is in line with the cautions from Anderson / Fowler quoted above, and is intended to remind of the importance of the careful choice of method: For example, Weber (1998) used phoneme detection tasks to compare German and Dutch - the latter without fricative assimilation. The major results consists in the finding that "German listeners detected the target fricative faster when the German fricative assimilation was violated than when no violation occurred" (Weber, 1998, p. 101). Without furthering this discussion, psycholinguistic methods are presumably necessary for obtaining answers to this kind of question.

3. Crosslinguistic trends in coarticulation

Coarticulation denotes the modification of a speech sound due to adjacent segments which has often been assumed to be an automatic consequence of speech physiology. If such a view holds, there is no room for language-specifics of coarticulatory patterns. Such a view seems hardly to be consistent with the results obtained in the present study: While there are numerous publications which have explored hypotheses for the large amount of velar coarticulation originating from motor control and even more "peripheral" mechanisms - as repeatedly displayed discussing approaches to the loop phenomenon - the current work attempted to approach the phenomenon of articulatory loops from the genuinely linguistic perspective of sound inventories. Given the crosslinguistic differences in German versus Hungarian velar stop production observed in this study, it is hard to avoid the question of how these differences in behaviour arise. This at the same time suggests that a consistent interpretation of the results of the current study cannot be achieved by refining the discussion to behaviour of the mature speech production apparatus and must respect acquisitional aspects of coarticulatory patterns. Concerning the acquisition of coarticulation, one of the few undisputed facts is that children exhibit more articulatory variability than adults do. But how are these facts to be interpreted? Kühnert und Nolan (1997) broadly distinguish between two different theoretical positions: A first perspective postulates that children produce speech more segmentally than adults do (Kent, 1983; Katz; Kripke und Tallal, 1991, after Kühnert und Nolan, 1997). This belief is triggered by the theoretical stance that the motor skill for the temporal sequencing of phones is acquired first, while the finer details of the temporal coordination are developed later. The consequence is that coarticulation is less prominent for children than for adults. The second position (Nittrouer; Studdert-Kennedy und McGowan, 1989; Nittrouer und Whalen, 1989, after Kühnert und Nolan 1997) holds the opposite, i.e. that children show stronger coarticulation than adults. This approach conforms better with Articulatory Phonology in the sense that it originates from syllable-based speech production units which at earlier stages exhibit larger spatiotemporal overlap which then is gradually narrowed. The few studies - most of them dealing only with acoustic recordings - which in fact have investigated the extent and degree of coarticulation in child productions have yielded inconsistent results though. Without reporting these results in detail, it seems obvious that the question raised in these studies, i.e. whether articulation develops from segmental to coarticulated or vice versa only repeats the notorious discussion between gesturalists and segmentalists already reviewed in extenso in the current work. This also illustrates the limitations of the current work: While it is possible to demonstrate that core linguistic aspects like the segmental inventory of a language have its reflex in "low-level" coarticulatory behaviour, there still appears to be no clear picture of how this behaviour originates. Presumably, in order to arrive at a clearer picture of how crosslinguistic patterns like the ones observed in the present study arise would have to incorporate its interdependency with the the development of the lexicon. But this is far beyond the scope of the present work.

A. Consonant Inventories Hungarian / German

	Bilabial	Lab. dent.	Dental	Alveolar	P-alveo.	Palatal	Velar	Uvular	Glottal
Plosive	p b			t d			k g		ʔ
Affricate		pf		ts	tʃ dʒ				
Nasal	m			n			ŋ		
Fricative		f v		s z	ʃ ʒ	ç		χ ʁ	h
Approx						j			
Lat. appr.				l					

Table A 1.: German

	Bilabial	Lab. dent.	Dental	Alveolar	P-alveo.	Palatal	Velar	Uvular	Glottal
Plosive	p b			t d		c ɟ	k g		
Affricate			ts dz		tʃ dʒ	(cç) (ɟʝ)			
Nasal	m			n			ŋ		
Fricative		f v		s z	ʃ ʒ				h
Approx						j			
Lat. appr.				l					

Table A 2.: Hungarian

B. Postures Hungarian

The left panels of the figures show ellipse plots for *initial* palatal and velar stop consonants, vowel means of the following vowel are added in darkgrey for palatal and in lightgrey for velar contexts. Middle panels: 2-σ-ellipses for vowel midpoints. Right panels show ellipse plots for *medial* palatal and velar stop consonants, vowel means of the preceeding vowels are also added, again in darkgrey for palatal and in lightgrey for velar contexts. Mean vowel contours and color coding is identical in all subplots. Consonant places-of-articulation are coded by triangles for palatal and circles for velar stops (see legend).

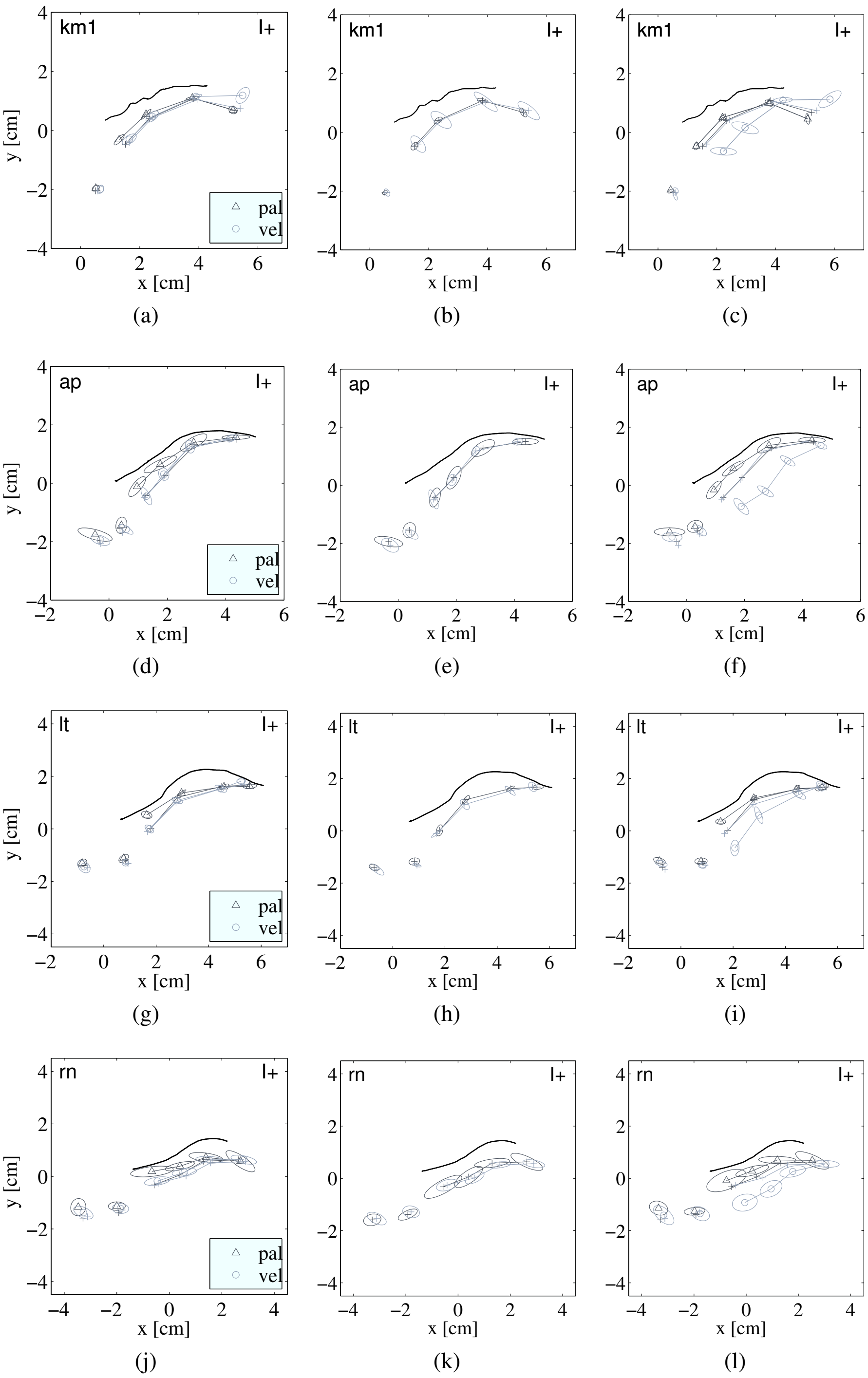


Figure A 1.: Tongue contours [CVCa]-sequences (Hungarian corpus). Vowel context: **[i:]**.

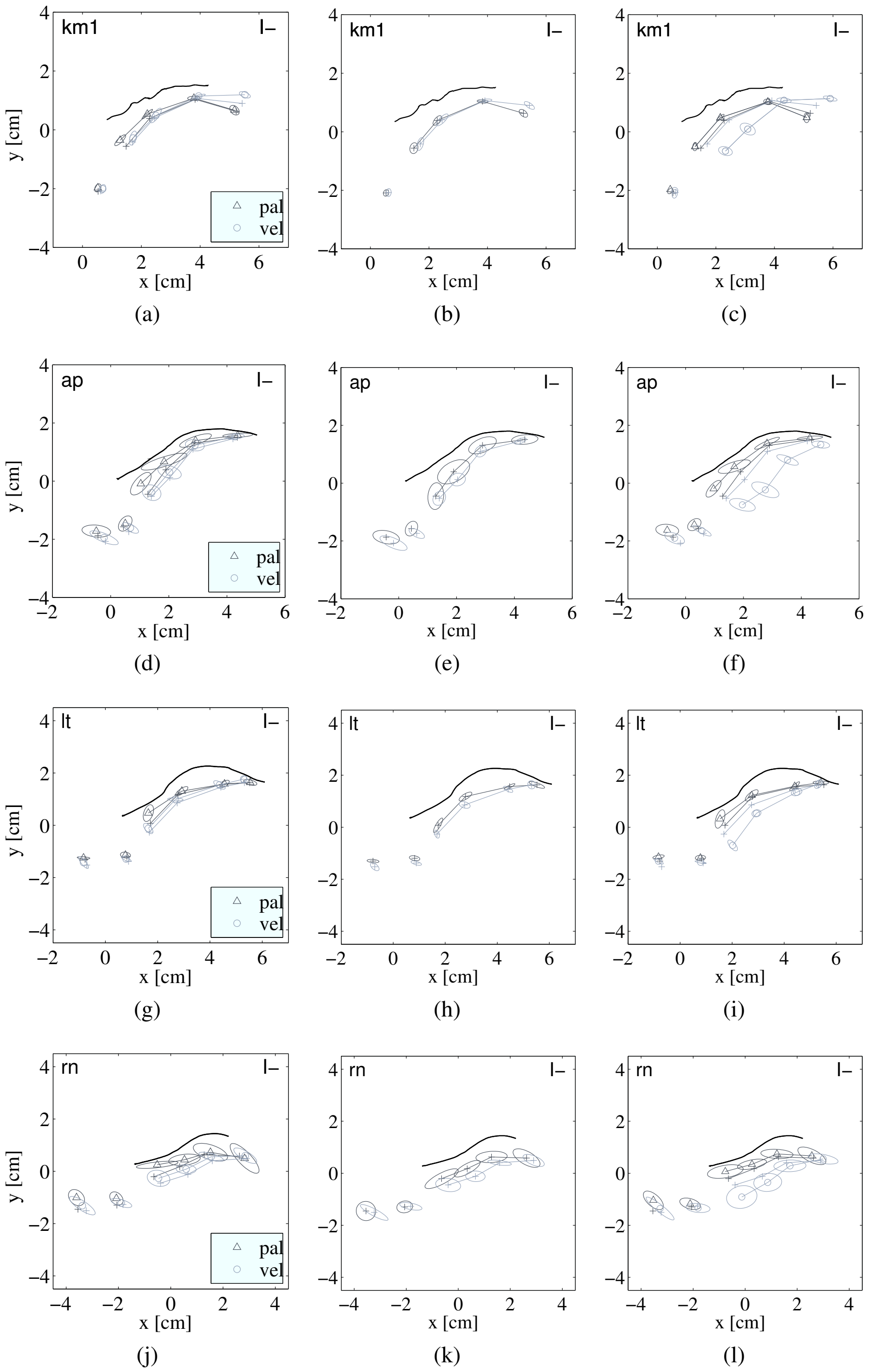
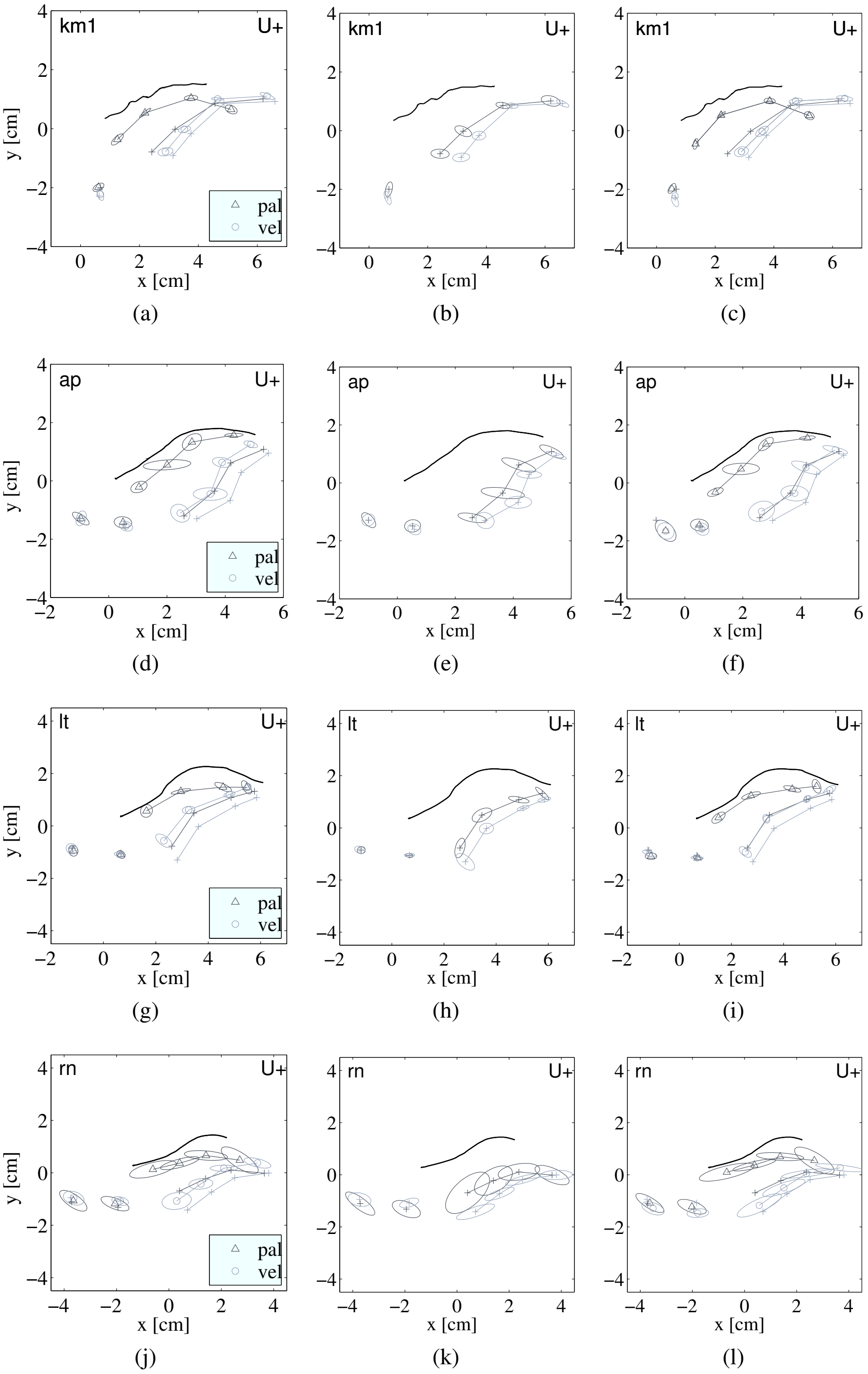


Figure A 2.: Tongue contours [CVCa]-sequences (Hungarian corpus). Vowel context: **[i]**.



(a) (b) (c)
(d) (e) (f)
(g) (h) (i)
(j) (k) (l)

Figure A 3.: Tongue contours [CVCa]-sequences (Hungarian corpus). Vowel context: **[uː]**.

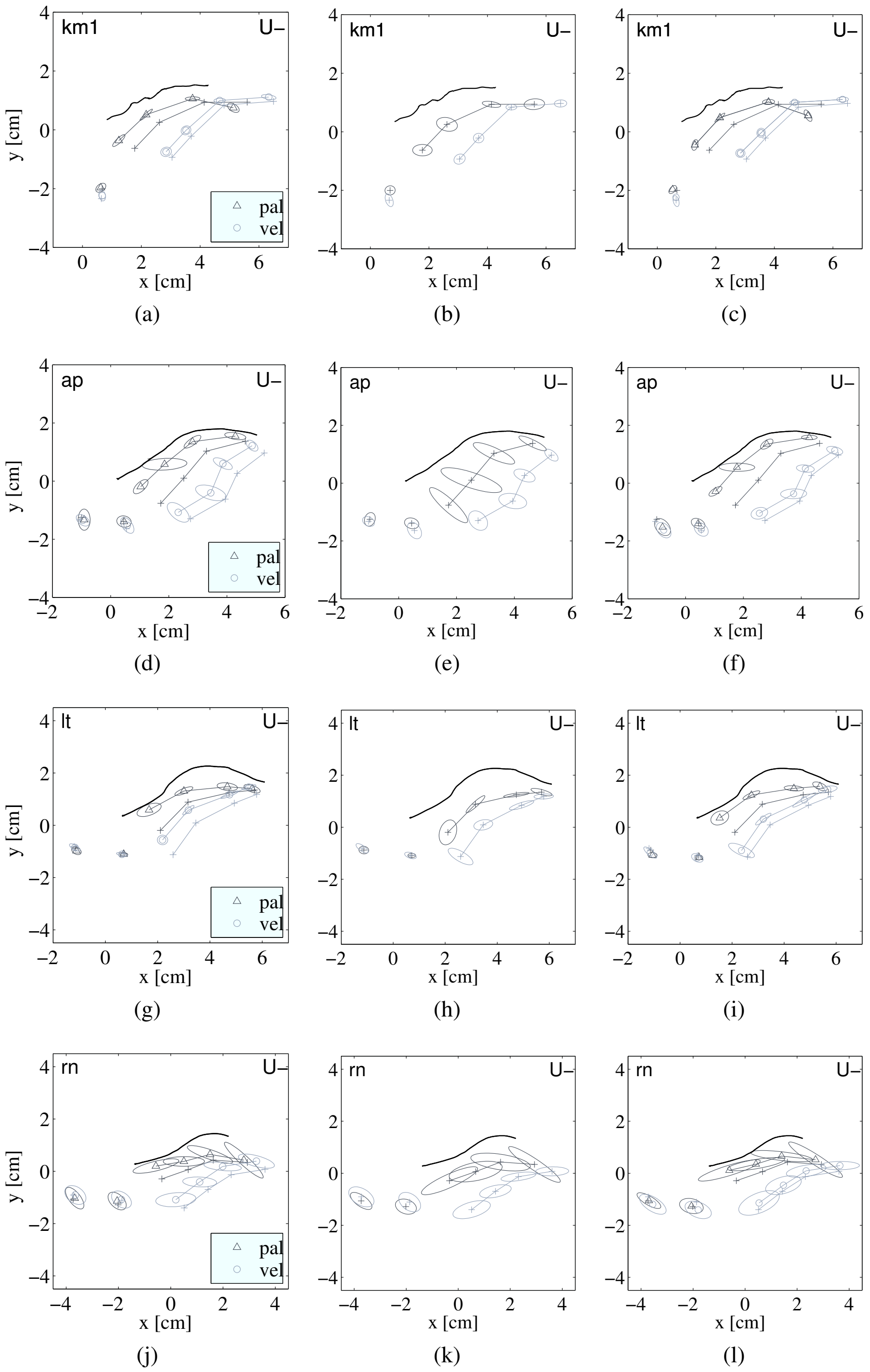


Figure A 4.: Tongue contours [CVCa]-sequences (Hungarian corpus). Vowel context: **[u]**.

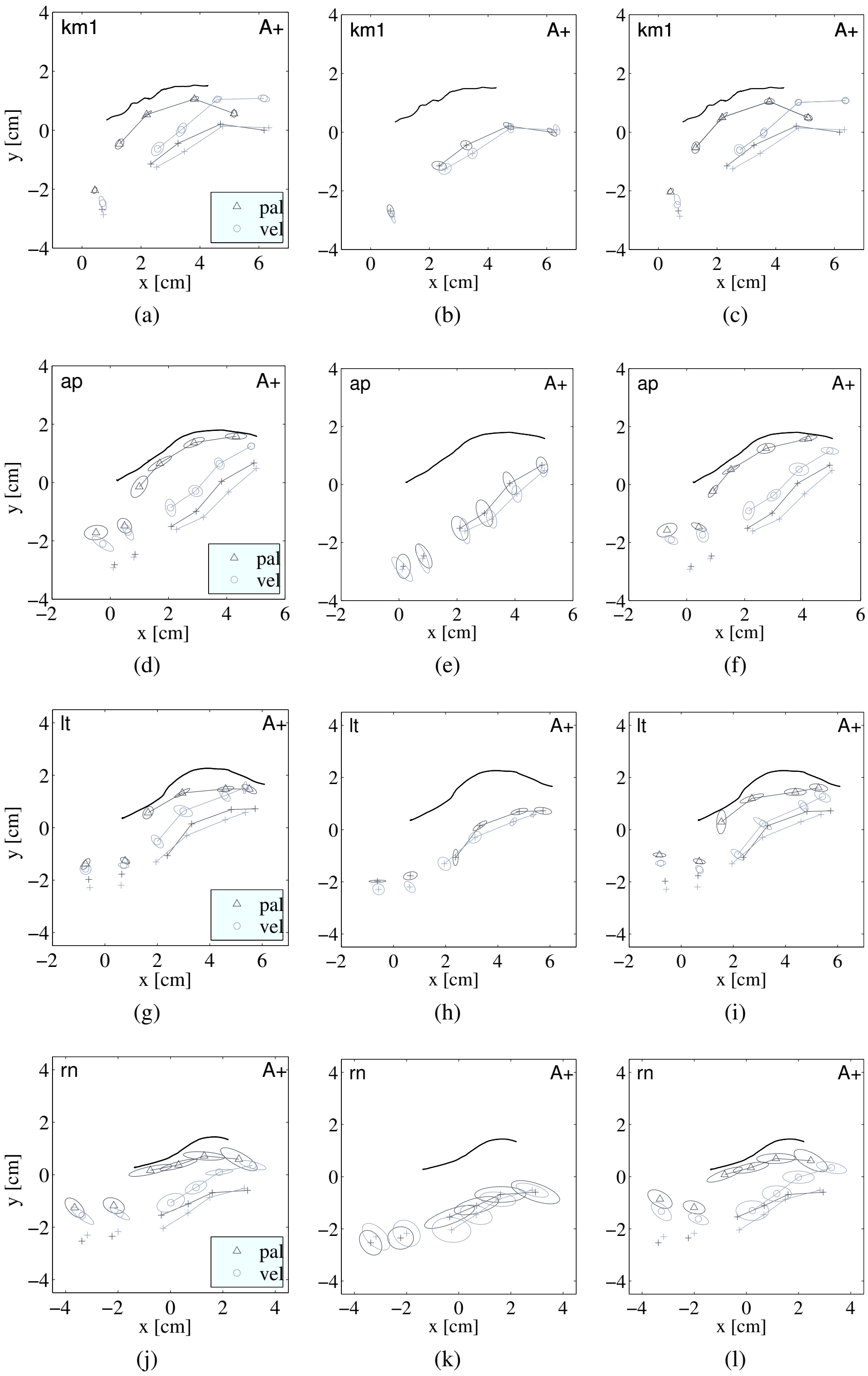


Figure A 5.: Tongue contours [CVCa]-sequences (Hungarian corpus). Vowel context: **[a:]**.

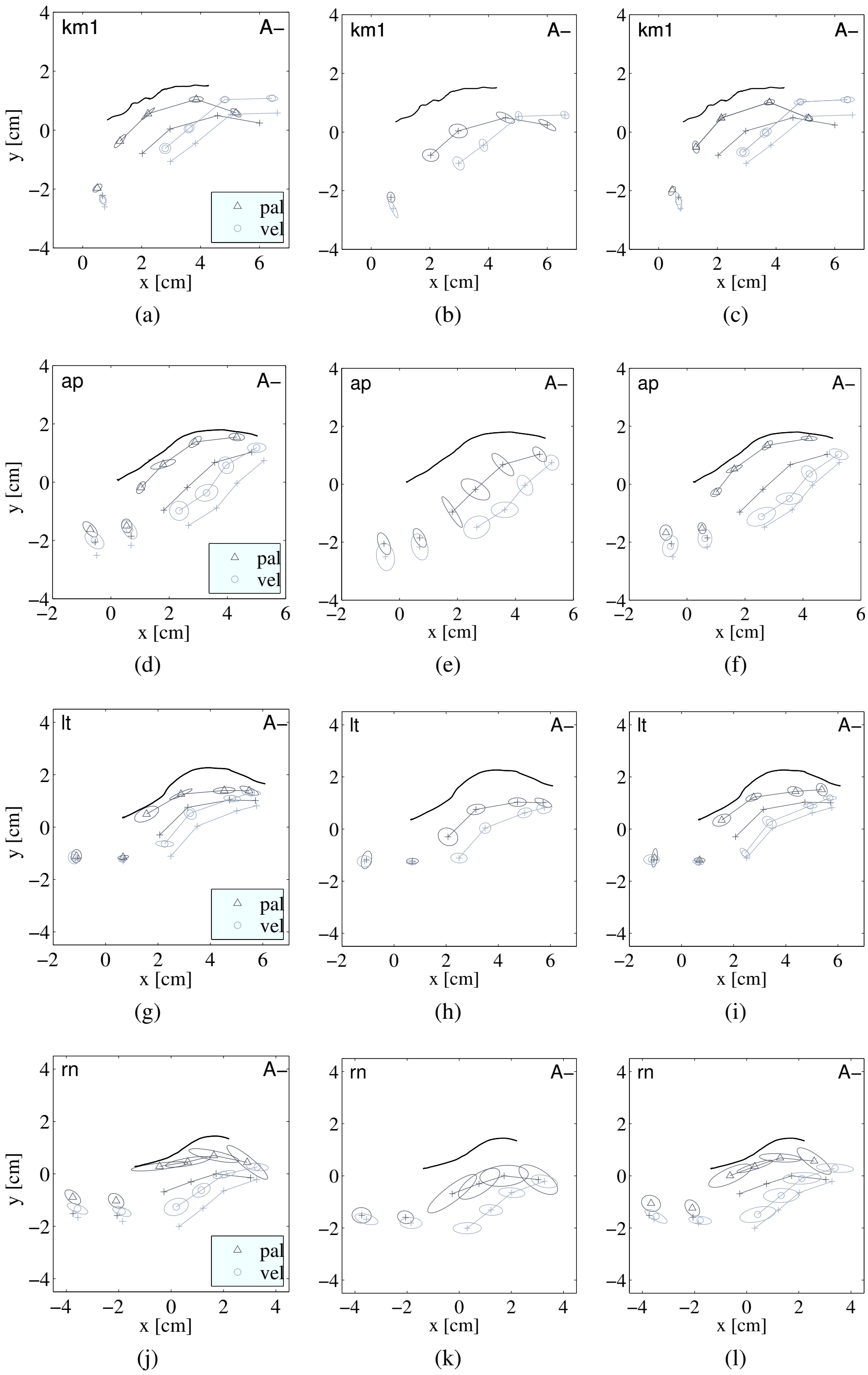


Figure A 6.: Tongue contours [CVCa]-sequences (Hungarian corpus). Vowel context: **[a]**.

C. Postures German

Figures A 7 to A 12 show ellipse plots for the German study. The figures are organized according to appendix B which contains ellipse plots for the Hungarian study. An overview over recorded materials was already given in table 3.6, where the main asymmetries of the design of the current work were displayed.

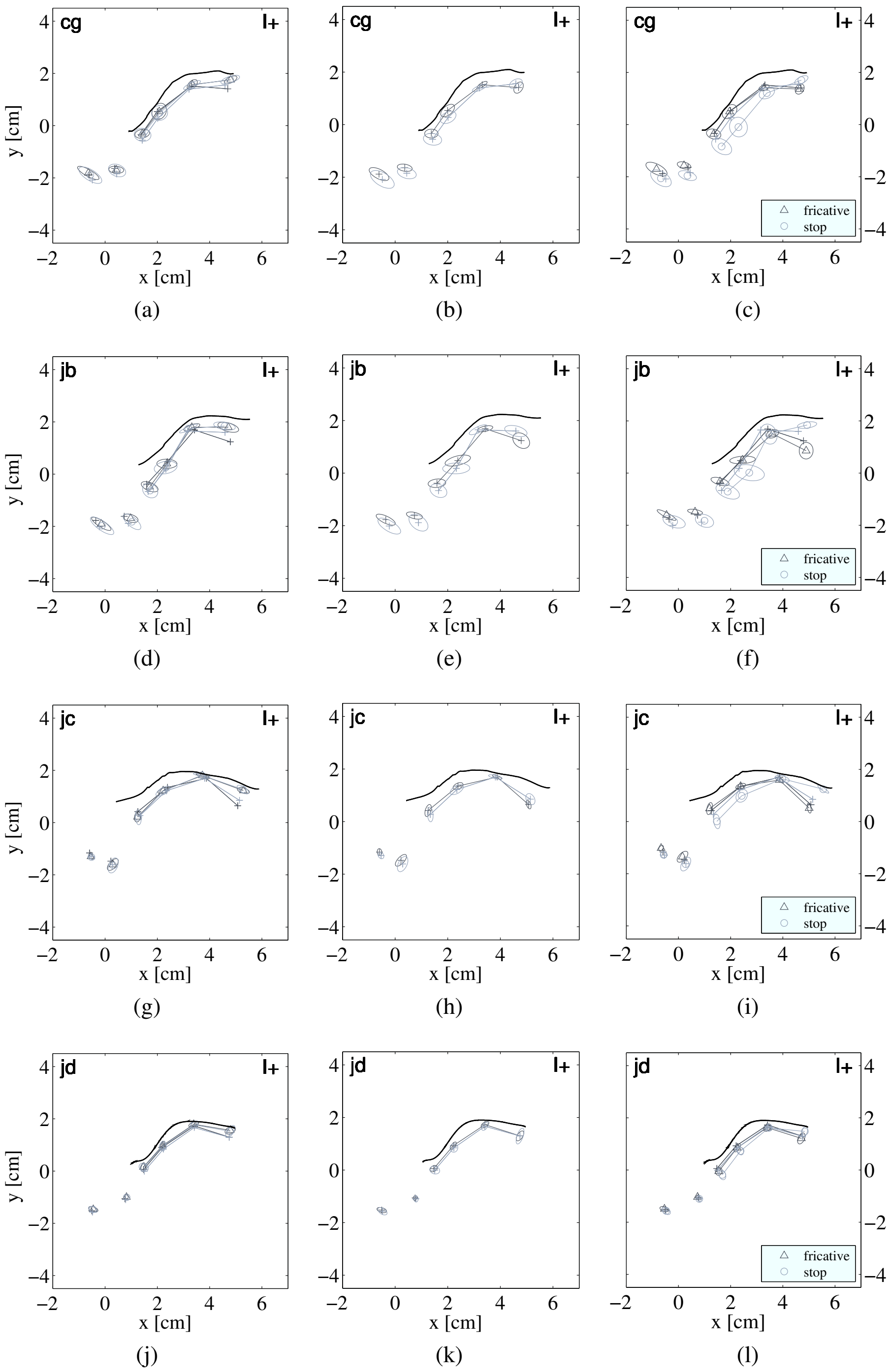


Figure A 7.: Tongue contours [CVCɐ]-sequences (German dataset). Vowel context: **[iː]**.

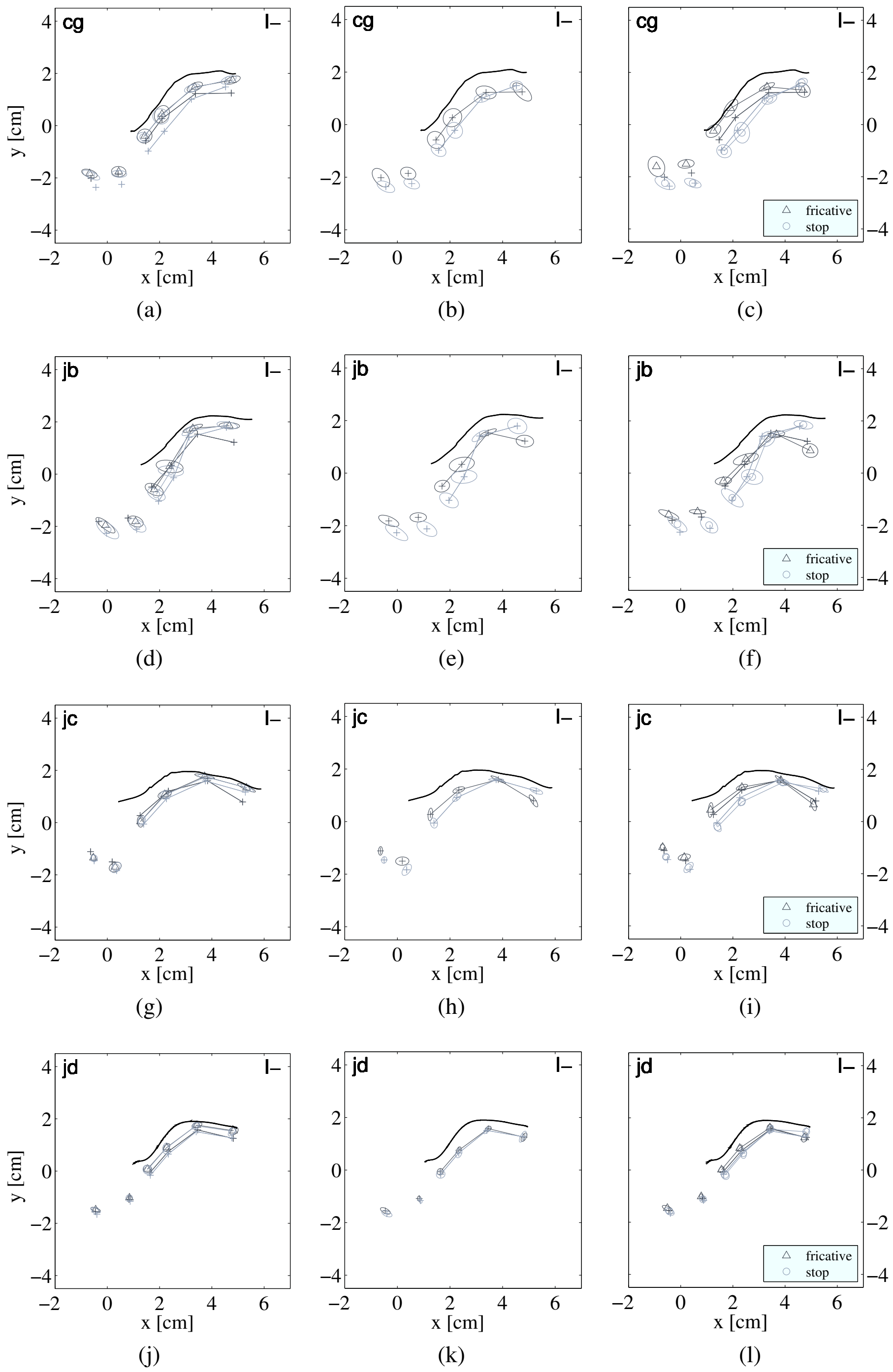


Figure A 8.: Tongue contours [CVCɐ]-sequences (German dataset). Vowel context: **[i]**.

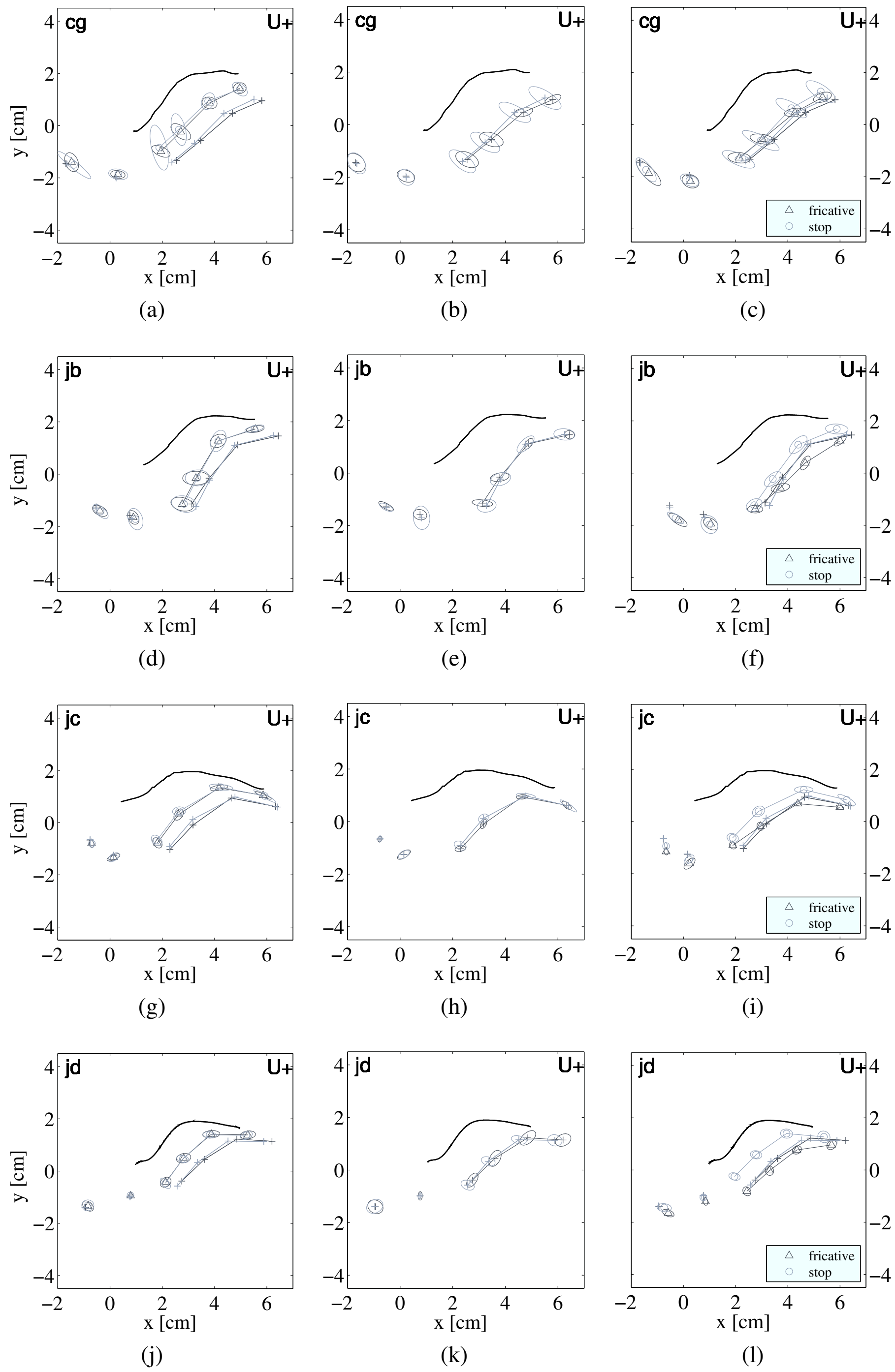


Figure A 9.: Tongue contours [CVCɐ]-sequences (German dataset). Vowel context: **[uː]**.

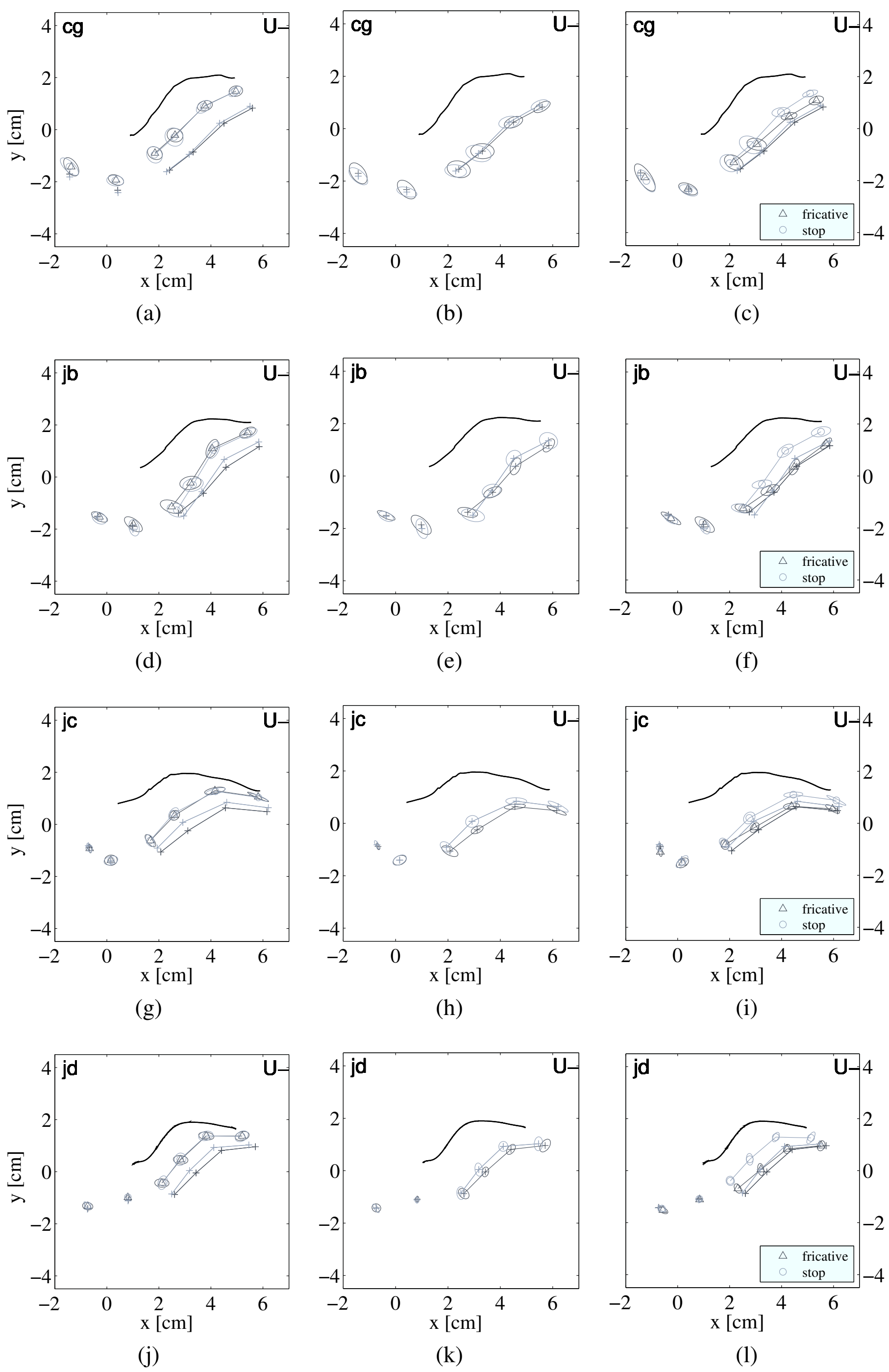


Figure A 10.: Tongue contours [CVCɐ]-sequences (German dataset). Vowel context: **[u]**.

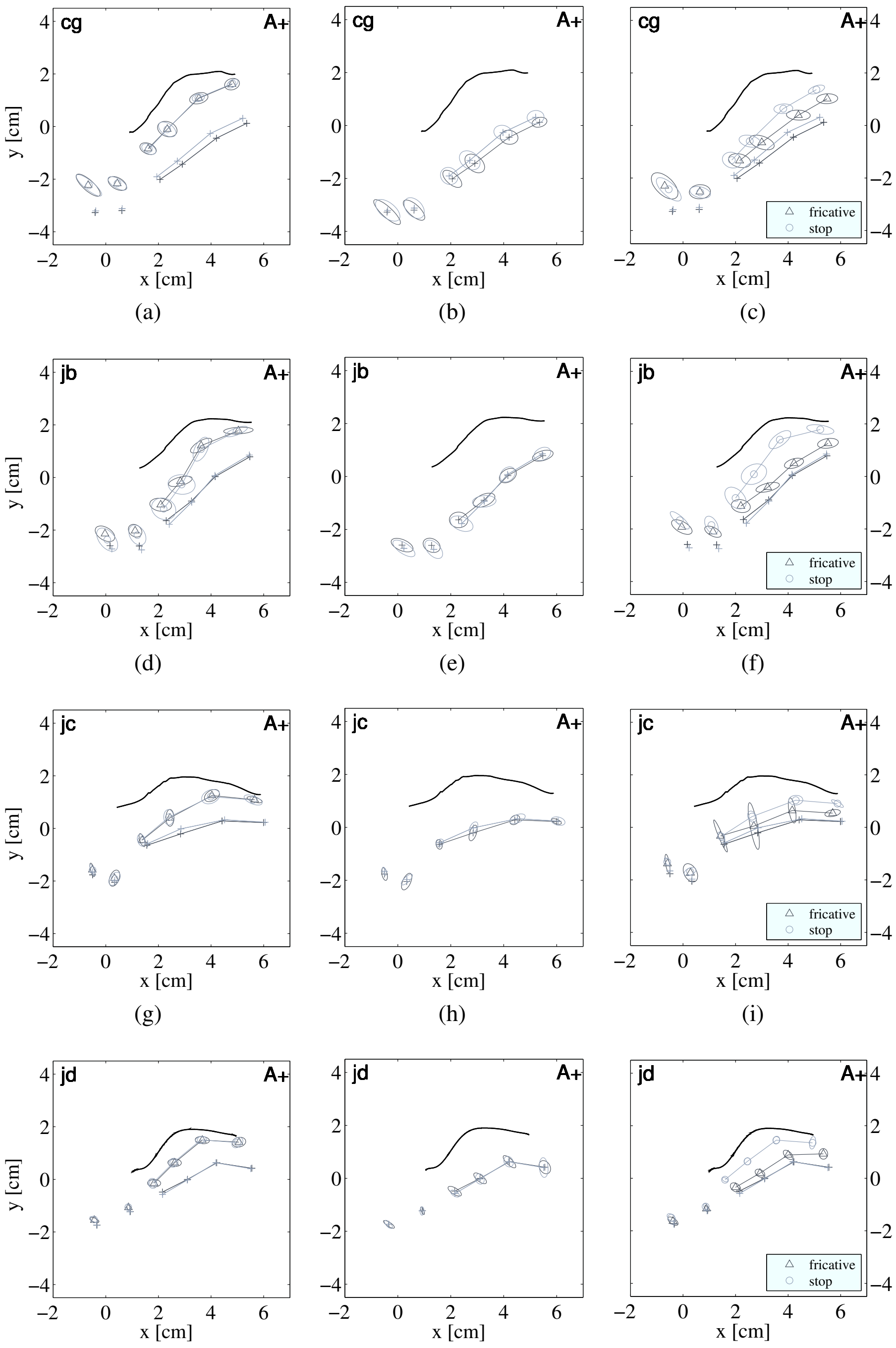


Figure A 11.: Tongue contours [CVCɐ]-sequences (German dataset). Vowel context: **[aː]**.

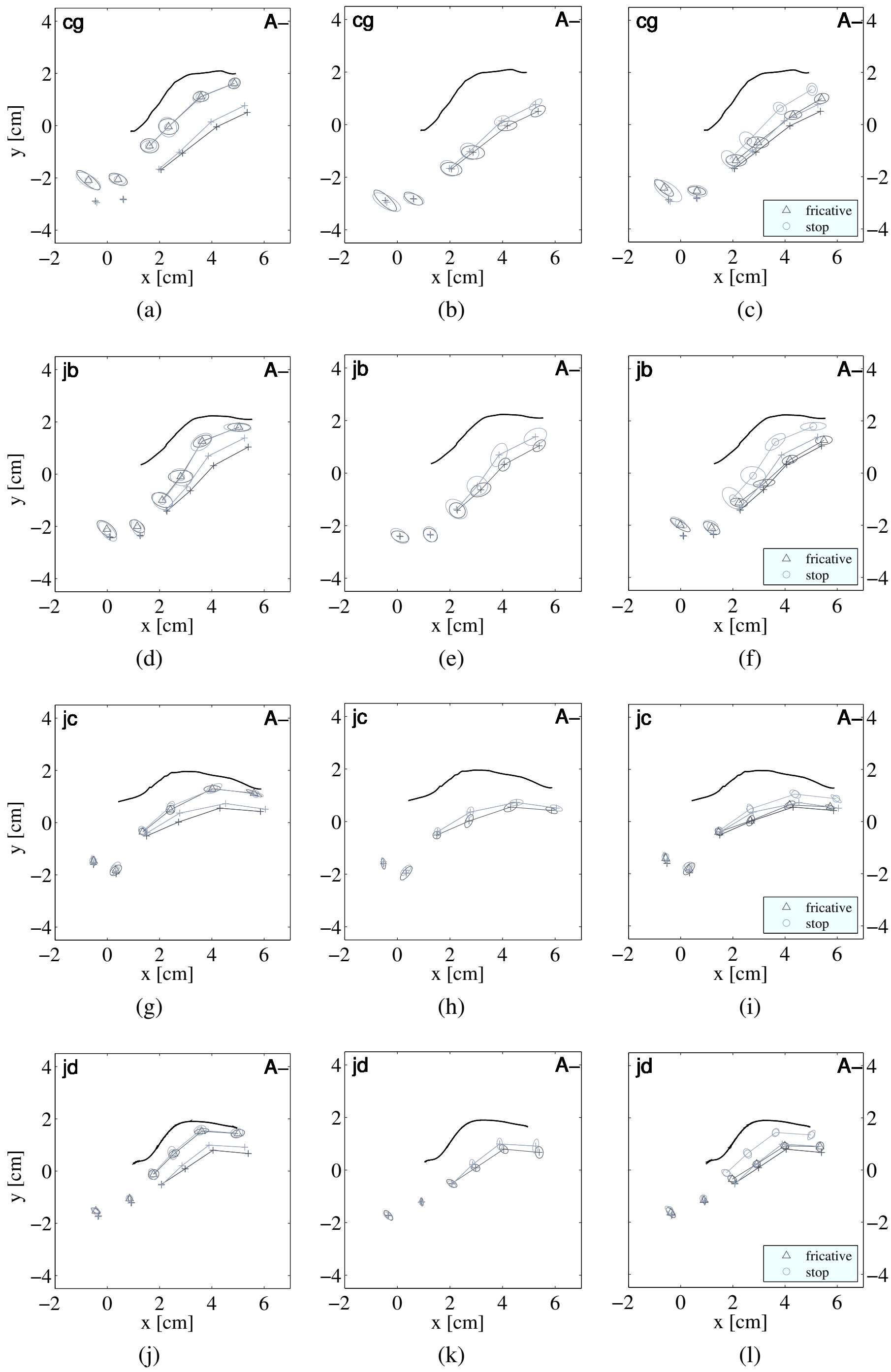


Figure A 12.: Tongue contours [CVCɐ]-sequences (German dataset). Vowel context: **[a]**.

D. German formant analyses

Figures A 13 to A 16 show formants measured at the temporal midpoint of V_1 of the /CV_1Cɐ/ sequences.

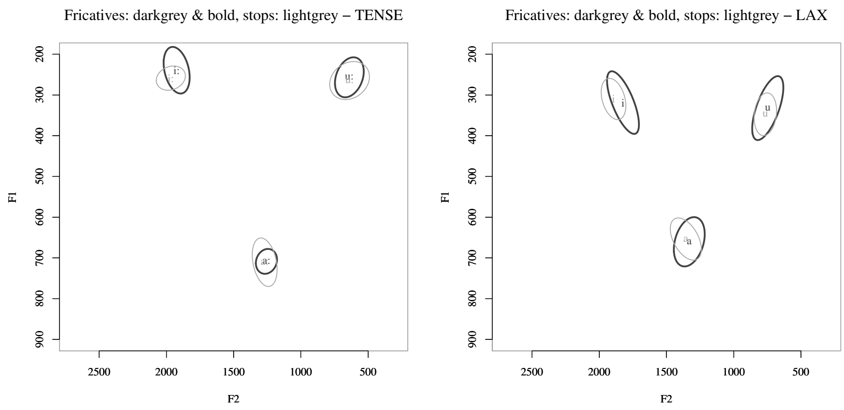


Figure A 13.: The first two formants at the temporal midpoint of the corner vowels for the German data set. This graph shows the data for speaker cg. Bold, darkgrey lines indicate that the following sound is a realisation of the palatal fricative /ç/ or one of its allophones ([x,χ]). Normal linetype refers to the contextual variants of the stop /k/.

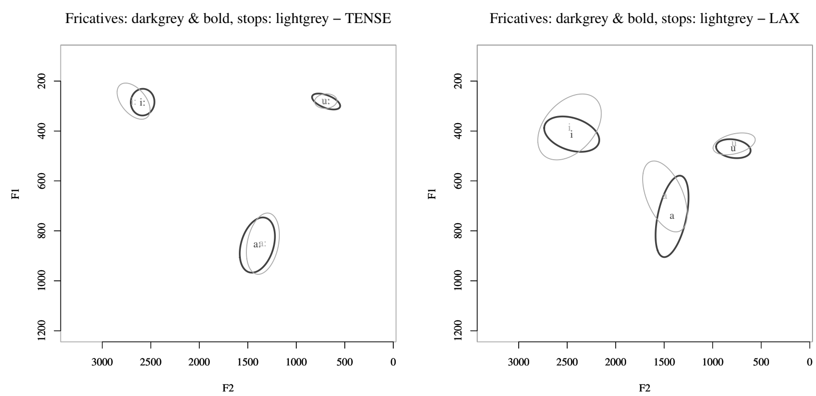


Figure A 14.: The first two formants at the temporal midpoint of the corner vowels for the German data set. This graph show the data for speaker jb. Bold, darkgrey lines indicate that the following sound is a realisation of the palatal fricative /ç/ or one of its allophones ([x,χ]). Normal linetype refers to the contextual variants of the stop /k/.

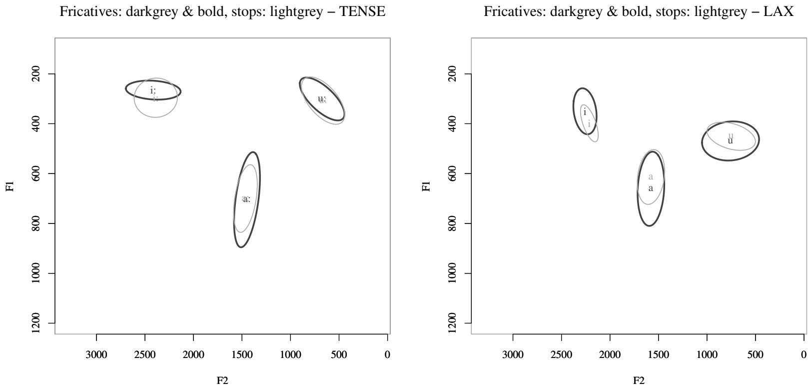


Figure A 15.: The first two formants at the temporal midpoint of the corner vowels for the German data set. This graph show the data for speaker jc. Bold, darkgrey lines indicate that the following sound is a realisation of the palatal fricative /ç/ or one of its allophones ([x,χ]). Normal linetype refers to the contextual variants of the stop /k/.

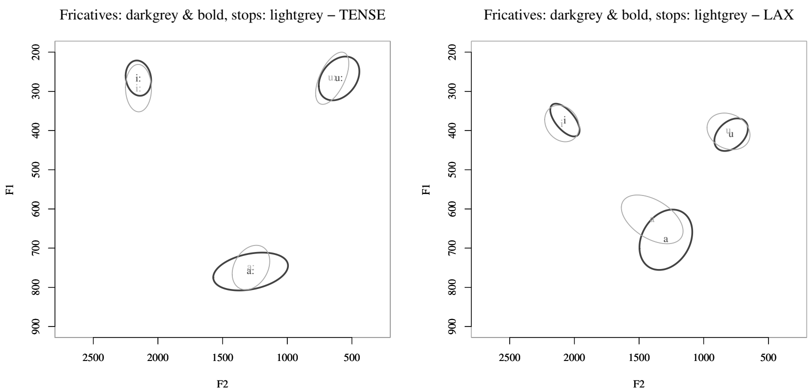


Figure A 16.: The first two formants at the temporal midpoint of the corner vowels for the German data set. This graph show the data for speaker jd. Bold, darkgrey lines indicate that the following sound is a palatal fricative /ç/ or one of its allophones ([x,χ]). Normal linetype refers to the contextual variants of the stop /k/.

Bibliography

Abbs, J.H. und Connor, N.P.: Motor sensory mechnisms of speech motor timing and coordination. In: *Journal of Phonetics*, Band 19: S. 333–342, 1991.

Abry, C.: [b]-[d]-[g] as a universal triangle as acoustically optimal as [i]-[a]-[u]. In: Solé, M.J; Recaesens, D. und Romero, J. (Hg.) *Proc. 15th Int.Cong. Phonetic Sciences*. 2003, S. 727–730.

Adank, P.; Smits, R. und van Hout, R.: A comparison of vowel normalization procdures for language variation research. In: *J. Acoust. Soc. Am.*, Band 116: S. 3099–3107, 2004.

Alfonso, P. und Baer, T.: Dynamics of vowel articulation. In: *Language and Speech*, Band 25: S. 151–173, 1982.

Ambrazaitis, Gilbert und John, Tina: On the allophonic behaviour of German /x/ vs /k/ - an EPG investigation. In: *AIPUK*, Band 36: S. 1–15, 2004.

Anderson, S.: Why phonology isn't natural. In: *Linguistic Inquiry*, Band 12: S. 493–539, 1981.

Archangeli, D. und Pulleyblank, D.: *Grounded Phonology*. Cambridge/London: MIT Press, 1994.

Barry, M.: Palatalization, assimilation and gestural weakening in connected speech. In: *Speech Communication*, Band 11: S. 393–400, 1992.

Beautemps, D.; Badin, P. und Bailly, Gérard: Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. In: *J. Acoust. Soc. Am.*, Band 109: S. 2165–2180, 2001.

Beckman, M.E.: Implications for phonological theory. In: Hardcastle, W. J. und Hewlett, N. (Hg.) *Coarticulation; Theory, Data and Techniques*, Cambridge: University Press, S. 199–228. 1999.

Bernstein-Ratner, N.: Phonological rule usage in mother-child speech. In: *Journal of Phonetics*, Band 12: S. 245–254, 1984.

Best, C.: Learning to perceive the sound patterns of English. In: Rovee-Collier, C. und Lipsitt, L.P. (Hg.) *Advances in Infancy Research*, Norwood, NJ: Ablex, S. 217–304. 1995.

Best, C.; McRoberts, G. und Sithole, N.: Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. In: *J. Exp. Psych.: Hum. Percept.Perform.*, Band 14: S. 345–360, 1988.

Bhat, D.: A general study of palatalization. In: Greenberg, J.H. (Hg.) *Universals of Human Language. Vol. 2.: Phonology*, Stanford: Stanford University Press, S. 47–92. 1978.

Bladon, R. A. W. und Nolan, F. J.: A video-fluorographic investigation of tip and blade alveolars in English. In: *Journal of Phonetics*, Band 5: S. 185–193, 1977.

Bladon, R.A. und Al-Bamerni, A.: Coarticulation reistance in English /l/. In: *Journal of Phonetics*, Band 4: S. 137–150, 1976.

Bloomfield, Leonard: German ç and x. In: *Maître Phonétique*, Band III 29: S. 27–28, 1930.

Blumstein, S. und Stevens, K.: Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. In: *J. Acoust. Soc. Am.*, Band 66: S. 1001–1017, 1979.

Blumstein, S. und Stevens, K. N.: Perceptual invariance and onset spectra for stop consonants in different vowel environments. In: *J. Acoust. Soc. Am.*, Band 67: S. 648–662, 1980.

Blumstein, S.E.: On acoustic invariance in speech. In: Perkell, Joseph S. und Klatt, Dennis H. (Hg.) *Invariance and Variability in Speech Processes*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, London, S. 179–197. 1986.

Boë, L.J.; Perrier, P.; Guérin, B. und Schwartz, J.L.: Maximal vowel space. In: *Eurospeech 89*, Band 2: S. 281–284, 1989.

Boersma, P. und Weenink, D.: Praat, a system for doing phonetics by computer. 1992–2007.

Bogliotti, C.: *Perception categorielle et perception allophonique: Incidences de l'âge, du niveau de lecture, et des couplages entre predispositions phonétiques*. Dissertation, Université Paris 7 - Denis Diderot, 2005.

Bradlow, A.: A comparative study of English and Spanish vowels. In: *J. Acoust. Soc. Am.*, Band 97: S. 1916–1924, 1995.

Bro, R.: *Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications*. Dissertation, Department of Dairy and Food Science, Royal Veterinary and Agricultural University Denmark, 1998.

Bro, R. und Kiers, H.: A new efficient method for determining the number of components in PARAFAC models. In: *Journal of Chemometrics*, Band 17: S. 274–286, 2003.

Browman, Catherine und Goldstein, Louis: Towards an articulatory phonology. In: *Phonology Yearbook*, Band 3: S. 219–252, 1986.

Browman, Catherine und Goldstein, Louis: Articulatory gestures as phonological units. In: *Phonology Yearbook*, Band 6: S. 201–251, 1989.

Browman, Catherine und Goldstein, Louis: Gestural specification using dynamically-defined articulatory structures. In: *Journal of Phonetics*, Band 18: S. 299–320, 1990a.

Browman, Catherine und Goldstein, Louis: Tiers in articulatory phonology, with some implications for casual speech. In: Kingston, John und Beckman, Mary E. (Hg.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, Cambridge, Cambridge, UK, S. 341–376. 1990b.

Browman, Catherine und Goldstein, Louis: Articulatory phonology: an overview. In: *Phonetica*, Band 49: S. 155–180, 1992.

Browman, Catherine P. und Goldstein, Louis: Dynamics and articulatory phonology. In: Port, Robert und van Gelder, Timothy (Hg.) *Mind as Motion: Explorations in the Dynamics of Cognition*, MIT Press, Cambridge, MA, S. 175–193. 1995.

Bybee, J.: *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.

Byrd, D. und Saltzman, E.: The elastic phrase: modeling the dynamics of boundary-adjacent lenghening. In: *Journal of Phonetics*, Band 31: S. 149–180, 2003.

Carré, R. und Mrayati, M.: Vowel-vowel trajectories and region modelling. In: *Journal of Phonetics*, Band 19: S. 433–443, 1991.

Carré, R.; Liénard, S.; Marsico, E. und Serniclaes, W.: On the role of the "schwa" in the perception of plosive consonants. In: *7th International Conference on Spoken Language Processing*, S. 1681–1684, 2002.

Cassidy, S. und Harrington, J.: Multi-level annotation in the EMU speech database management system. In: *Speech Communication*, Band 33: S. 61–77, 2001.

Cattell, R.B. und Cattell, A.: Factor rotation for proportional profiles: analytical solution and an example. In: *British Journal of Statistical Psychology*, Band 8: S. 83–92, 1955.

Cercignani, F.: *The consonants of German: Synchrony and diachrony*. Milano: Cisalpino-Goliardica, 1979.

Chomsky, Noam und Halle, Morris: *The Sound Pattern of English*. Harper and Row, New York, 1968.

Clements, G. N.: Affricates as noncontoured stops. In: Fujimura, O. und Palek, B. (Hg.) *Proceedings of the Fourth International Linguistics and Phonetics Conference, Columbus,OH*, Prague: Charles University Press, S. 271–299. 1999.

Clements, G. Nick und Hume, Elizabeth V.: The internal organization of speech sounds. In: Goldsmith, J.A. (Hg.) *The handbook of phonological theory*, Oxford: Blackwell, S. 245–306. 1995.

Clements, G.N.: The geometry of phonological features. In: *Phonology Yearbook*, Band 2: S. 225–252, 1985.

Connell, B.: *Phonetic aspects of lower cross languages and their implications for sound change*. Dissertation, University of Edinburgh, 1991.

Cribari-Neto, F. und Zarkos, S. G.: R: Yet another econometric programming environment. In: *Journal of Applied Econometrics*, Band 14: S. 319–329, 1999.

Dang, J. und Honda, K.: A physiological articulatory model for simulating speech production process. In: *Acoust. Sci. & Tech.*, Band 22: S. 415–425, 2001.

Daniloff, R. und Hammarberg, R.: On defining coarticulation. In: *Journal of Phonetics*, Band 1: S. 239–248, 1973.

Davis, K. und Kuhl, P.K.: Tests of the perceptual magnet effect for American English /k/ and /g/. In: *127th ASA Meeting, 4pSP17 (1994)*, 1994.

DeLacy, P.: *Markedness. Reduction and preservation in phonology*. Cambridge: Cambridge University Press, 2006.

Delattre, P: Coarticulation and the locus theory. In: *Studia Linguistica*, Band 23: S. 1–26, 1969.

Delattre, P.C.; Liberman, A.M. und Cooper, F.S.: Acoustic loci and transitional cues for consonants. In: *Journal of the Acoustical Society of America*, Band 27: S. 769–773, 1955.

Derkach, M.; Fant, G. und deSerpa Leitao, A.: Phoneme coarticulation in Russian hard and soft VCV-utterances with voiceless fricatives. In: *Speech Transmission Laboratory Quarterly Progress and Status Report*, Band 2–3: S. 1–7, 1970.

Diehl, Randy L. und Kluender, Keith R..: On the objects of speech perception. In: *Ecological Psychology*, Band 1: S. 121–144, 1989.

Diehl, R.L.; Lindblom, B. und Creeger, C.P.: Increasing realism of auditory representations yields further insights into vowel phonetics. In: *Proceedings of the 15th International Congress of Phonetic Sciences*. Adelaide: Causal Publications, 2003, Band 2, S. 1381–1384.

Dorman, M.; Studdert-Kennedy, M. und Raphael, L: Stop-consonant recognition: release bursts and formant transitions as functionally equivalent, context-dependent cues. In: *Perception and Psychophsics*, Band 22: S. 109–122, 1977.

Ellis, L. und Hardcastle, W.: Categorical and gradient properties in alveolar and velar sequences: evidence from EPG and EMA data. In: *Journal of Phonetics*, Band 30: S. 373–396, 2002.

Farnetani, E. und Recasens, D.: Coarticulation in recent speech production theories. In: Hardcastle, W. J. und Hewlett, N. (Hg.) *Coarticulation; Theory, Data and Techniques*, Cambridge: University Press, S. 31–65. 1999.

Feldman, A.: Once more on the equilibrium point hypothesis (λ model) for motor control. In: *Journal of Motor Behavior*, Band 18: S. 17–54, 1986.

Féry, C.: Konsonantische allophonische Alternationen in OT. in prep. URL `http://www.sfb632.uni-potsdam.de/homes/fery/`.

Fischer-Jørgensen, E.: Tape-cutting experiments with Danish stop consonants in initial position. In: *Annu. Rep. Inst. Phon., Univ. Copenhagen*, Band 6: S. 104–168, 1972.

Flanagan, J. L.: *Speech analysis, synthesis, and perception*. Springer, New York, 1972.

Flemming, E. S.: *Auditory Representations in Phonology*. Routledge: New York, London, 2002.

Fowler, C.: Coarticulation and theories of extrinsic timing. In: *Journal of Phonetics*, Band 8: S. 113–133, 1980.

Fowler, C.: Realism and unrealism: A reply. In: *Journal of Phonetics*, Band 11: S. 303–322, 1983.

Fowler, C.: Real objects of speech perception: a commentary on Diehl and Kluender. In: *Ecological Psychology*, Band 1: S. 145–169, 1989.

Fowler, C.A.: Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. In: *Perception & Psychophysics*, Band 55(6): S. 597–610, 1994.

Geng, C.; Fuchs, S.; Mooshammer, C. und Pompino-Marschall, B.: How does vowel context influence loops? In: *Proceedings of the 6th Speech Production Seminar, Sydney*. 2003, S. 67–72.

Geng, C. und Hoole, P.: Some comments on the reliability of three-index factor analysis models in speech research. In: *ZAS Papers in Linguistics*, Band 42: S. 219–239, 2005.

Geng, C.; Mády, K.; Bogliotti, C.; Messaoud-Galusi, S.; Medina, V. und Serniclaes, W.: Do palatal consonants correspond to the fourth category in the perceptual F2-F3 space? In: Hazan, V. und Iverson, P. (Hg.) *ISCA Workshop on Plasticity in Speech Perception*. 2005, S. 219–222.

Geng, C. und Mooshammer, C.: Modeling the German stress distinction. In: *Proc. 5th Speech Production Seminar*. 2000, S. 161–164.

Geng, C. und Mooshammer, C.: The Hungarian palatal stop: Phonological considerations and phonetic data. In: *ZAS Papers in Linguistics*, Band 37: S. 221–243, 2004.

Gibson, J.J.: *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum (originally published: Boston: Houghton-Mifflin,1979), 1986.

Giegerich, H.: *Metrical Phonology and Phonological Structure: German and English*. Cambridge: Cambridge University Press, 1985.

Goldinger, S.D.; Pisoni, D.B. und Luce, P.A.: Speech perception and spoken word recognition: Research and theory. In: *Research on Speech Perception, Progress Report*, Band 16: S. 3–64, 1990.

Goldsmith, J.: *Autosegmental Phonology*. Bloomington: Indiana University Linguistics Club, 1976.

Goldstein, L.; Byrd, D. und Saltzman, E.: The role of vocal tract gestural action units in understanding the evolution of phonology. In: *Action to language via the Mirror Neuron System*, Cambridge: Cambridge University Press, S. 215–249. 2006.

Goldstein, L. und Fowler, C.: Articulatory phonology: a phonology for public language use. In: Schiller, Niels O. und Meyer, Antje S. (Hg.) *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, Mouton de Gruyter, Berlin, S. 159–207. 2003.

Guion, S.: *Velar Palatalization, Coarticulation, Perception, and sound change*. Dissertation, University of Texas at Austin, 1996.

Guion, S.: The role of perception in the sound change of velar palatalization. In: *Phonetica*, Band 55: S. 18–52, 1998.

Gussenhoven, C. und Jacobs, H.: *Understanding Phonology*. Hodder Arnold, 1998.

Hall, T.A.: Lexical phonology and the distribution of German [ç] and [x]. In: *Phonology*, Band 6(1): S. 1–17, 1989.

Hall, T.A.: *The Phonology of Coronals*. Amsterdam: John Bejmains, 1997.

Hall, T.A.: Whose German? the ach/ich alternation and related phenomena in 'standard' and 'colloquial' (review). In: *Phonology*, Band 19: S. 293–298, 2002.

Halle, M.: Features. In: Bright, W. (Hg.) *Oxford International Encyclopedia*, Oxford: Oxford University Press. 1988.

Hammarberg, R.: The metaphysics of coarticulation. In: *Journal of Phonetics*, Band 4: S. 353–363, 1976.

Hammarberg, R.: On redefining coarticulation. In: *Journal of Phonetics*, Band 10: S. 123–137, 1982.

Hardcastle, William J.: *Physiology of speech production : an introduction for speech scientists*. Academic Press, London; New York, 1976. ISBN 0123249503 9780123249500.

Harshman, R. A.; Ladefoged, P. und Goldstein, L.: Factor analysis of tongue shapes. In: *J. Acoust. Soc. Am.*, Band 62: S. 693–707, 1977.

Harshman, R.A. und Lundy, M.E.: Data preprocessing and the extended PARAFAC model. In: Law, H.G. (Hg.) *Research Methods for Multimode Data Analysis*, Prager, New York, S. 216–284. 1984.

Hombert, Jean-Marie; Ohala, John J. und Ewan, William G.: Phonetic explanations for the development of tones. In: *Language*, Band 55: S. 37–58, 1979.

Hoole, P.: Issues in the acquisition, processing, reduction and parameterization of articulographic data. In: *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, Band 34: S. 158–173, 1996.

Hoole, P.: On the lingual organization of the German vowel system. In: *J. Acoust. Soc. Am.*, Band 106(2): S. 1020–1032, 1999.

Hoole, P.; Munhall, K. und Mooshammer, C.: Do air-stream mechanisms influence tongue movement paths? In: *Phonetica*, Band 55: S. 131–146, 1998.

Houde, R.: A study of tongue body motion during selected consonant sounds. Speech Communications Research Laboratory, Santa Barbara,SCRL Monograph 2, 1968.

Hyman, L.: The feature [grave] in phonological theory. In: *Journal of Phonetics*, Band 1: S. 329–337, 1973.

International Phonetic Association (Hg.): *Handbook of the International Phonetic Association*. Cambridge University Press, Cambridge, UK, 1999.

Itô, J. und Mester, A.: The core-periphery structure of the lexicon and contraints on reranking. In: Beckman, J.; Walsh-Dickey, L. und Urbanczyk, S. (Hg.) *Papers in Optimality Theory*, Amherst: GLSA, S. 181–210. 1995.

Itô, J. und Mester, A.: Realignment. In: Kager, H., R. van der Hulst und Zonneveld, W. (Hg.) *The Prosody-Morphology Interface*, Cambridge: Cambridge University Press, S. 188–217. 1999.

Jackson, M. T. T.: Analysis of tongue positions: Language-specific and cross-linguistic models. In: *J. Acoust. Soc. Am.*, Band 84: S. 124–143, 1988.

Jacob, F.: Evolution and tinkering. In: *Science*, Band 196: S. 1161–1166, 1977.

Jakobson, Roman; Fant, Gunnar und Halle, Morris: *Preliminaries to Speech Analysis*. MIT Press, Cambridge, 1952.

Jessen, M.: Phonetic implementation of the distinctive auditory features [voice] and [tense]. In: *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung, Lehrstuhl für Experimentelle Phonetik Universität Stuttgart: Research Papers from the Phonetisc Labs*, Band 6(3): S. 11–62, 2000.

Jessen, Michael: Die dorsalen Reibelaute [ç] and [x] im Deutschen. In: *Linguistische Berichte*, Band 117: S. 371–396, 1988.

Johnson, Keith: *Acoustic and Auditory Phonetics*. Blackwell, Oxford, UK; Cambridge, MA, 1997.

Jongman, A.; Fourakis, M. und Sereno, J.A.: The acoustic vowel space of modern Greek and German. In: *Language and Speech*, Band 32: S. 221–248, 1989.

Kager, R.: *Optimality theory*. Cambridge: Cambridge Univ. Press, 1999.

Kakita, K.; Hirose, H.; Ushijima, T. und Sawashima, M.: A preliminary report on the electromyographic study of the production of Japanese semivowels. In: *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, Band 10: S. 37–46, 1976.

Kálmán, B.: Hungarian historical phonology. In: Benkö, Lorand und Imre, Samu (Hg.) *The Hungarian Language*, Mouton, The Hague. 1972.

Katz, W.; Kripke, C. und Tallal, P.: Anticipatory coarticulation in the speech of adults and young children: acoustic, perceptual and video data. In: *Journal of Speech and Hearing Research*, Band 34: S. 1222–1232, 1991.

Keating, P.A.: Palatals as complex segments: X-ray evidence. In: *UCLA Working Papers in Phonetics*, Band 69: S. 77–91, 1988a.

Keating, P.A.: Underspecification in phonetics. In: *Phonology*, Band 5: S. 275–292, 1988b.

Keating, P.A.: Phonetic representations in a generative grammar. In: *Journal of Phonetics*, Band 18: S. 321–334, 1990a.

Keating, P.A.: The window model of coarticulation: articulatory evidence. In: Kingston, John und Beckman, Mary E. (Hg.) *Papers in Laboratory Phonology I*, Cambridge University Press, Cambridge, S. 451–470. 1990b.

Keating, P.A.: Coronal places of articulation. In: Paradis, C. und Prunet, Jean-François (Hg.) *Phonetics and Phonology 2: The special status of coronals: Internal and external evidence*, Academic Press, San Diego, Calif., S. 29–48. 1991.

Keating, P.A.: Phonetic representation of palatalization versus fronting. In: *UCLA Working Papers in Phonetics*, Band 85: S. 6–21, 1993.

Keating, P.A. und Lahiri, A.: Fronted velars, palatized velars, and palatals. In: *Phonetica*, Band 50: S. 73–101, 1993.

Keating, P.A.; Lindblom, B.; Lubker, J. und Kreimann, J.: Variability in jaw height for segments in English and Swedish VCVs. In: *Journal of Phonetics*, Band 22: S. 407–422, 1994.

Kent, R.: The segmental organization of speech. In: MacNeilage, P. (Hg.) *The Production of Speech*, New York: Springer, S. 57–89. 1983.

Kent, R und Minifie, F.: Coarticulation in recent speech production models. In: *Journal of Phonetics*, Band 5: S. 115–133, 1977.

Kent, R. und Moll, K.: Cinefluographic analyses of selected lingual consonants. In: *Journal of Speech and Hearing Research*, Band 15: S. 453–473, 1972.

Kent, R.D. und Murray, A.D: Acoustic features of infant vocalic utterances at 3, 6, and 9 months. In: *J. Acoust. Soc. Am.*, Band 72: S. 353–364, 1982.

Kewley-Port, D.: *Reference Manual for KLTEXC (Version 3)*, 1978.

Kewley-Port, D.; Pisoni, D. B. und Studdert-Kennedy, M.: Perception of static and dynamic cues to place of articulation in initial stop consosnants. In: *J. Acoust. Soc. Am.*, Band 73: S. 1779–1793, 1983.

Kewley-Port, Diana: Measurement of formant transitions in naturally produced stop consonant-vowel syllables. In: *J. Acoust. Soc. Am.*, Band 72: S. 379–389, 1982.

Kiers, H. A. L.; ten Berge, J. M. F. und Bro, R.: PARAFAC2-Part I. A direct fitting algorithm for the PARAFAC2 model. In: *Journal of Chemometrics*, Band 13: S. 275–294, 1999.

Kingston, J. und Diehl, R.: Phonetic knowledge. In: *Language*, Band 70: S. 419–454, 1994.

Kingston, J. und Diehl, R.: Intermediate properties in the perception of distinctive feature values. In: Connell, B. und Arvantini, A. (Hg.) *Papers in Laboratory Phonology IV: Phonology and Phonetic evidence*, Cambridge: Cambridge University Press, S. 7–27. 1995.

Kiparsky, P.: Some consequences of Lexical Phonology. In: *Phonology Yearbook*, Band 2: S. 85–138, 1985.

Kiparsky, P.: Phonological change. In: Newmeyer, F.J. (Hg.) *Linguistics: The Cambridge Survey*, University Press, Cambridge, S. 363–415. 1988.

Kiparsky, P.: The phonological basis of sound change. In: Goldsmith, J. (Hg.) *Handbook of Phonological Theory*, Cambridge, MA: Blackwell, S. 640–670. 1995.

Kipp, Andreas; Wesenick, M.-B. und Schiel, Florian: Automatic detection and segmentation of pronunciation variants in German speech corpora. In: *Proceedings of the 6th International Conference on Spoken language Processing*. Philadelphia, 1996, Band 1, S. 106–109.

Kochetov, A. und Lobanova, A.: Palatal stops and affricates in Komi-Permyak. In: McRobbie, Z. und So, C. (Hg.) *Proceedings of the Thirteenth Conference of the Finno-Ugric Studies Association of Canada*. Simon Fraser University, Burnaby, British Columbia., 2003, S. 73–81.

Kohler, K.: German. In: *Journal of the International Phonetic Association*, Band 20: S. 48–50, 1990a.

Kohler, K.: Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In: Hardcastle, W.J. und Marchal, A. (Hg.) *Speech Production and Speech Modeling*, Kluwer Academic Publishers, S. 69–92. 1990b.

Kohler, Klaus J.: *Einführung in die Phonetik des Deutschen*. Erich Schmidt Verlag, Berlin, zweite Auflage, 1995. (1st edition 1977).

Kovács, M.: Az affrikátaák időszerkezetéről. In: *Kísérleti Laboratóriumi Fonológia*, S. 39–54, 2002.

Kuhl, P.K.; Williams, K.A.; Lacerda, F.; Stevens, K.N. und Lindblom, B.: Linguistic experiences alters phonetic perception in infants by 6 months of age. In: *Science*, Band 255: S. 606–608, 1992.

Kühnert, B. und Hoole, P.: Speaker-specific kinematic properties of alveolar reductions in English and German. In: *Clinical Linguistics & Phonetics*, Band 18(6–8): S. 559–575, 2004.

Kühnert, Barbara und Nolan, Francis: The origin of coarticulation. In: *FIPKM*, Band 35: S. 61–75, 1997.

Lahiri, A.; Gewirth, L. und Blumstein, S. E.: A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. In: *J. Acoust. Soc. Am.*, Band 76: S. 391–404, 1984.

Lahiri, Aditi und Blumstein, Sheila E.: A re-evaluation of the feature coronal. In: *Journal of Phonetics*, Band 12(2): S. 133–146, 1984.

Lahiri, Aditi und Evers, V.: Palatalization and coronality. In: Paradis, C. und Prunet, Jean-François (Hg.) *Phonetics and Phonology 2: The special status of coronals: Internal and external evidence*, San Diego, California: Academic Press, S. 79–100. 1991.

Lane, Harlan: The Motor Theory of Speech Perception: A critical overview. In: *Psychological Review*, Band 72: S. 275–309, 1965.

Lass, R.: *Phonology: An introduction to basic concepts*. Cambridge: Cambridge University Press, 1984.

Leppänen, P. H. T. und Lyytinen, H: Auditory event-related potentials in the study of developmental language-related disorders. In: *Audiology and NeuroOtology*, Band 2(5): S. 308–340, 1997.

Liberman, A. M.; Cooper, F. S.; Shankweiler, D. P. und Studdert-Kennedy, M.: Perception of speech code. In: *Psychological Review*, Band 74: S. 431–461, 1967.

Liberman, Alvin M.: Some results of research on speech perception. In: *J. Acoust. Soc. Am.*, Band 29: S. 117–123, 1957.

Liberman, Alvin M.; Delattre, Pierre C.; Cooper, Fraklin S. und Gerstman, Louis J.: The role of consonant-vowel transitions in the perception of the stop and nasal consonants. In: *Psychological Monographs: General and Applied*, Band 68: S. 1–13, 1954.

Liberman, Alvin M.; Harris, K.S.; Hoffman, H.S. und Griffith, B.C.: The discrimination of speech sounds within and across phoneme boundaries. In: *Journal of Experimental Psychology*, Band 54: S. 358–368, 1957.

Liberman, Alvin M. und Mattingly, Ignatius G.: The motor theory of speech perception revised. In: *Cognition*, Band 21: S. 1–36, 1985.

Liljencrants, J. und Lindblom, B.E.F.: Numerical simulation of vowel quality systems: the role of perceptual contrast. In: *Language*, Band 48: S. 839–862, 1972.

Lindblom, B. und Sundberg, J.: Acoustical consequences of lip, tongue, jaw and larynx movement. In: *J. Acoust. Soc. Am.*, Band 50: S. 1166–1179, 1971.

Lindblom, B und Sussman, H.M.: Articulatory and acoustic bases of locus equations. In: *Proceedings of Fonetik 2004, the XVII Swedish Phonetics Conference*. Stockholm, 2004, S. 8–16.

Lindholm, J. M.; Dorman, M.; Taylor, B. E. und Hannley, M. T.: Stimulus factors influencing the identification of voiced stop consonants by normal-hearing and hearing-impaired listeners. In: *J. Acoust. Soc. Am.*, Band 83: S. 1608–1614, 1988.

Livijn, P.: Acoustic distribution of vowels in differently sized inventories - hot spots or adaptive dispersion? In: *Perilus*, Band 23: S. 93–96, 2000.

Local, J.: Modeling assimilation in nonsegmental, rule-free synthesis. In: *Papers in Laboratory Phonology II*, Cambridge: Cambridge University Press, S. 190–223. 1992.

Löfqvist, A. und Gracco, V.: Tongue body kinematics in the velar stop production: Influences of consonant voicing and vowel context. In: *Phonetica*, Band 51: S. 52–67, 1994.

Löfqvist, A. und Gracco, V. L.: Control of oral closure in lingual stop consonant production. In: *J. Acoust. Soc. Am.*, Band 111(6): S. 2811–2827, 2002.

MacFarland, T. und Pierrehumbert, J.: On ich-laut, ach-laut and Structure Preservation. In: *Phonology*, Band 8: S. 171–180, 1991.

Maeda, S.: Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: Hardcastle, W. und Marchal, A. (Hg.) *Speech production and speech modelling*, Kluwer Academic Publishers, Dordrecht,The Netherlands, S. 131–149. 1990.

Maeda, S. und Honda, K.: From EMG to formant patterns of vowels: The implications of vowel spaces. In: *Phonetica*, Band 51: S. 17–29, 1994.

Manuel, S.: The role of contrast in limiting vowel-to-vowel coarticulation in different languages. In: *J. Acoust. Soc. Am.*, Band 88: S. 1286–1298, 1990.

Manuel, S. und Krakow, R.: Universal and language particular aspects of vowel-to-vowel coarticulation. In: *Haskins Laboratories Status Report on Speech Research*, Band 77/78: S. 69–78, 1984.

Massaro, D. M.: *Speech Perception by Eye and Ear: A Paradigm for Psychological Inquiry*. Erlbaum, Hillsdale, NJ, 1987.

Massaro, D. M. und Cohen, M. M.: The paradigm and the fuzzy logical model of perception are alive and well. In: *J. Exp. Psychol.: General*, Band 122: S. 115–124, 1993.

Mayo, C. und Turk, A.: Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. In: *J. Acoust. Soc. Am.*, Band 115: S. 3184–3194, 2005.

McCarthy, J. und Prince, A.: Generalized Alignment. In: Booij, G. und van Marle, Jaap (Hg.) *Yearbook of Morphology*, Kluwer, Dordrecht, S. 79–153. 1993.

McNeilage, P.F.: The frame/content theory of evolution of speech production. In: *Behavioral and Brain Sciences*, Band 21: S. 499–546, 1998.

Menzerath, Paul und de Lacerda, A.: *Koartikulation, Steuerung und Lautabgrenzung*. Phonetische Studien 1. Ferd. Dümmlers, Berlin and Bonn, 1933.

Merchant, J.: Alignment and fricative assimilation in German. In: *Linguistic Inquiry*, Band 27(4): S. 209–219, 1996.

Modarresi, G.; Sussman, H.M.; Lindblom, B. und Burlingame, E.: Locus equation encoding of stop place: revisiting the voicing/VOT issue. In: *Journal of Phonetics*, Band 33: S. 101–113, 2005.

Mooshammer, C.; Hoole, P. und Kühnert, B.: On loops. In: *Journal of Phonetics*, Band 23: S. 3–21, 1995.

Moulton, William G.: *The Sounds of English and German*. University of Chicago Press, Chicago, 1962.

Munhall, K.G.; Ostry, D.J. und Flanagan, J.R.: Coordinate spaces in speech planning. In: *Journal of Phonetics*, Band 19: S. 293–307, 1991.

Nádasdy, A. und Siptár, P.: Issues in Hungarian Phonology: Preliminary queries to a new project. In: *Acta Linguistica Hungarica*, Band 39: S. 3–27, 1989.

Nam, H.; Goldstein, L.; Saltzman, E. und Byrd, D.: TADA: An enhanced, portable task dynamics model in MATLAB. In: *Journal of the Acoustical Society of America*, Band 115(5): S. 2430, 2004.

Neagu, Adrien und Bailly, Gerard: Relative contributions of noise burst and vocalic transitions to the perceptual identification of stop consonants. In: *Proc. Eurospeech '97*. Rhodes, Greece, 1997, S. 2175–2178.

Nearey, T.: A double-weak view of trading relations: comments on Kingston and Diehl. In: Connell, B. und Arvanti, A. (Hg.) *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*, Cambridge University Press, Cambridge, S. 28–40. 1995.

Nearey, T.: Speech perception as pattern recognition. In: *J. Acoust. Soc. Am.*, Band 101: S. 3241–3254, 1997.

Nearey, T.M: Critique: Phonological contrast and articulatory properties. In: *J. Acoust. Soc. Am.*, Band 99: S. 1714–1717, 1996.

Newton, B.E.: *The generative interpretation of dialect: a study of modern Greek phonology*. Cambridge: Cambridge University Press, 1972.

Nguyen, N.; Gibbon, F. und Hardcastle, W. J.: Articulatory and perceptual aspects of fricative-stop coarticulation: a pilot study. In: *QMC Working Papers in Speech and Language Sciences*, Band 1: S. 13–25, 1996.

Nittrouer, S.: Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. In: *Journal of Phonetics*, Band 20: S. 1–32, 1992.

Nittrouer, S.; Studdert-Kennedy, M. und McGowan, R.: The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. In: *Journal of Speech and Hearing Research*, Band 32: S. 120–132, 1989.

Nittrouer, S. und Whalen, D.: The perceptual effects of child-adult differences in fricative vowel coarticulation. In: *Journal of the Acoustical Society of America*, Band 86, 1989.

Nix, D. A.; Papcun, M. G.; Hodgen, J. und Zlokarnik, I.: Two cross-linguistic factors underlying tongue shapes for vowels. In: *J. Acoust. Soc. Am.*, Band 99: S. 3707–3718, 1996.

Noske, M.: Feature spreading as dealignment: the distribution of [ç] and [x] in German. In: *Phonology*, Band 14: S. 221–234, 1997.

Odden, D.: Further evidence for the feature [grave]. In: *Linguisitc Inquiry*, Band 9: S. 141–144, 1978.

Ohala, John J.: Experimental historical phonology. In: Anderson, J. M. und Jones, C. (Hg.) *Historical linguistics II. Theory and description in phonology. Proc. of the 1st Int. Conf. on Historical Linguistics, Edinburgh, 2 - 7 Sept. 1973*, North Holland, Amsterdam, S. 353–389. 1974.

Ohala, John J.: The origin of sound patterns in vocal-tract constraints. In: MacNeilage, P. F. (Hg.) *The Production of Speech*, New York: Springer, S. 189–216. 1983.

Ohala, John J.: The phonetics of sound change. In: Jones, C. (Hg.) *Historical linguistics: Problems and perspectives*, Longman, London, S. 237–278. 1993.

Ohde, R.N. und Sharf, D.J.: Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study. In: *J. Acoust. Soc. Am.*, Band 69: S. 297–300, 1981.

Öhman, S.E.G.: Coarticulation in VCV utterances: Spectrographic measurements. In: *J. Acoust. Soc. Am.*, Band 39: S. 151–168, 1966.

Öhman, S.E.G.: Numerical model of coarticulation. In: *J. Acoust. Soc. Am.*, Band 41: S. 310–320, 1967.

Olsson, M.: *Hungarian Phonology and Morphology*. Lund: Lund University Press, 1992.

Olsson, M.: The basic Hungarian allophone system: Structure and rules. In: *Lund University, Dpt. of Linguistics, Working Papers*, Band 40: S. 157–184, 1993.

Padgett, J. und Zygis, M.: The evolution of sibilants in Polish and Russian. In: *ZAS Papers in Linguistics*, Band 32: S. 145–174, 2003.

Pagliuca, W. und Mowrey, R.: On certain evidence for the feature [grave]. In: *Language*, Band 56: S. 503–514, 1980.

Paul, H.: *Deutsche Grammatik*. Tübingen: Niemeyer. [Unveränderter Nachdruck der ersten Auflage: 1950], 1916.

Payan, Y. und Perrier, P.: Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. In: *Speech Communication*, Band 22(2–3): S. 185–205, 1997.

Pegg, J. und Werker, J.: Adult and infant perception of two English phones. In: *J. Acoust. Soc. Am.*, Band 102: S. 3742–3753, 1997.

Perkell, J. S.: *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge: Cambridge University Press, 1969.

Perkins, W. H. und Kent, R. D.: *Functional Anatomy of Speech, Language and Hearing*. College-Hill Press, San Diego, California, 1986.

Perrier, P.; Payan, Y.; Zandipour, M. und Perkell, J.: Influences that shape tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. In: *J. Acoust. Soc. Am.*, Band 114(3): S. 1582–1599, 2003.

Phillips, C.: Levels of representation in the electrophysiology of speech perception. In: *Cognitive Science*, Band 25(5): S. 711–731, 2001.

Pierrehumbert, J.: Phonological and phonetic representations. In: *Journal of Phonetics*, Band 18: S. 375–394, 1990.

Pierrehumbert, J.: Beyond contrast and ease. In: *ZAS - conference on the Phonetics-Phonology Interface: Abstracts*, 2001. URL `www.zas.gwz-berlin.de/events/phon_interfaces/abstracts/pierrehumbert.pdf`.

Pisoni, D.: *On the Nature of Categorical Perception of Speech Sounds*. Dissertation, The University of Michigan, 1971.

Pisoni, D. B. und Tash, J.: Reaction times to comparisons within and across categories. In: *Perception & Psychophysics*, Band 15(2): S. 285–290, 1974.

Plauche, M.; Delogu, C. und Ohala, J. J.: Asymmetries in consonant confusion. In: *Eurospeech 97*, S. 2187–2190, 1997.

Pollack, I. und Pisoni, D.: On the comparison between identification and discrimination tests in speech perception. In: *Psychonomic Science*, Band 24(6), 1971.

Pompino-Marschall, B. und Mooshammer, C.: Factors of variation in the production of the German dorsal fricative. In: *Eurospeech 97*, S. 375–378, 1997.

Port, R.F. und Leary, A.M.: Against formal phonology. In: *Language*, Band 81: S. 927–964, 2005.

Prince, A. und Smolensky, P.: *Optimality Theory: Constraint interaction in generative grammar*. Linguistic Inquiry Monograph. Cambridge: Cambridge University Press, 1993.

Ramers, K.H. und Vater, H.: *Einführung in die Phonologie*. Hürth-Efferen: Gabel, 1992.

Recasens, D.: The articulatory characteristics of palatal consonants. In: *Journal of Phonetics*, Band 18: S. 267–280, 1990.

Recasens, D.: A model of lingual coarticulation based on articulatory constraints. In: *J. Acoust. Soc. Am.*, Band 102(1): S. 544–561, 1997.

Recasens, D.: An EMA study of VCV coarticulatory direction. In: *J. Acoust. Soc. Am.*, Band 111(6): S. 2828–2840, 2002.

Recasens, D.: Articulation and sound change in Romance. In: *Proceedings of the 15th International Congress of Phonetic Sciences*. 2003, S. 231–234.

Recasens, D.: Acoustics and perception of velar softening for unaspirated stops, 2007. Unpublished Ms.

Recasens, D. und Espinosa, A.: Articulatory, positional and contextual characteristics of palatal consonants: Evidence from Majorcan Catalan. In: *Journal of Phonetics*, Band 34: S. 295–318, 2006.

Recasens, D.; Farnetani, E.; Fontdevila, J. und Pallares, .M.D.: An electropalatographic study of alveolar and palatal consonants in Catalan and Italian. In: *Language and Speech*, Band 36(2–3): S. 213–234, 1993.

Recasens, D.; Fontdevila, J. und Pallarès: Linguopalatal coarticulation and alveolar-palatal correlations for velarized and nonvelarized /l/. In: *Journal of Phonetics*, Band 24: S. 165–185, 1994.

Recasens, Daniel: Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. In: *Language and Speech*, Band 28: S. 97–114, 1985.

Repp, B.H. und Liberman, A.M.: Phonetic category boundaries are flexible. In: Harnad, S. (Hg.) *Categorical Perception*, Cambridge: Cambridge University Press, S. 89–112. 1987.

Repp, Bruno: Categorical perception: Issues, methods, and findings. In: Lass, Norman J. (Hg.) *Speech and Language: Advances in Basic Research and Practice*, Academic Press, New York, Band 10, S. 243–335. 1984.

Robinson, O.: *Whose German? The ach/ich Alternation and related phenomena in 'standard' and 'colloquial'*. Amsterdam: John Benjamins, 2001.

Rubin, R. und Baer, T.: An articulatory snthesizer for perceptual research. In: *J. Acoust. Soc. Am.*, Band 70(2): S. 321–328, 1981.

Sagey, E.: *The Representation of Features and Relations in Nonlinear Phonology*. New York: Garland, 1990.

Saltzman, Elliot L. und Munhall, Kevin G.: A dynamical approach to gestural patterning in speech production. In: *Ecological Psychology*, Band 1(4): S. 333–382, 1989.

Sanguineti, V.; Laboissière, R. und Ostry, D.J.: A dynamic biomechanical model for neural control of speech production. In: *J. Acoust. Soc. Am.*, Band 103: S. 1615–1627, 1998.

Sanguineti, V.; Laboissière, R. und Payan, Y.: A control model of human tongue movements in speech. In: *Biological Cybernetics*, Band 77: S. 11–22, 1997.

Schiel, F.: Automatic Phonetic transcription of non-prompted speech. In: *Proceedings of the 14th International Congress of Phonetic Sciences (San Francisco)*. 1999, S. 607–610.

Schwartz, J. L.; Boë, L. J.; Vallée, N. und Abry, C.: The dispersion-focalization theory of vowel systems. In: *Journal of Phonetics*, Band 25(3): S. 255–286, 1997.

Serniclaes, W.: *Etude expérimentale de la perception du trait de voisement des occlusives du français*. Dissertation, Université Libre de Bruxelles, 1987.

Serniclaes, W.; Van Heghe, S.; Mousty, P.; Carré, R und Sprenger-Charolles, L.: Allophonic mode of speech perception in dyslexia. In: *Journal of Experimental Child Psychology*, Band 87: S. 336–361, 2004.

Shepherd, G. M.: The significance of real neuron architectures for neural network simulations. In: Schwartz, E.L. (Hg.) *Computational Neuroscience*, Cambridge: MIT Press, S. 82–96. 1990.

Siptár, P.: How many affricates are there in Hungarian? In: *Proceedings of the Speech Research '89 International Conference*. 1989, Band 20 von *Hungarian Papers in Phonetics*, S. 123–126.

Siptár, Péter und Törkenczy, Miklós: *The phonology of Hungarian*. Oxford: Oxford Univ. Press, 2000.

Smits, R.: Context-dependent relevance of burst and transitions for perceived place in stops: It's in production, not perception. In: *Proc. ICSLP '96*. Philadelphia, PA, 1996, Band 4, S. 2470–2473.

Smits, R.; ten Bosch, L. und Collier, R.: Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. perception experiment. In: *J. Acoust. Soc. Am.*, Band 100: S. 3852–3864, 1996a.

Smits, R.; ten Bosch, L. und Collier, R.: Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. II. modeling and evaluation. In: *J. Acoust. Soc. Am.*, Band 100: S. 3865–3881, 1996b.

Sondhi, M. M. und Schroeter, J.: A hybrid time frequency domain articulatory speech synthesizer. In: *IEEE Transactions on Acoustics, Speech and Signal Processing*, Band 35: S. 955–967, 1987.

Stevens, Kenneth N.: *Acoustic Phonetics*. Cambridge: Cambridge University Press, 1998.

Stevens, Kenneth N. und Blumstein, S. E.: The search for invariant acoustic correlates of phonetic features. In: Eimas, P. D. und Miller, J. L. (Hg.) *Perspectives on the Study of Speech*, Erlbaum, Hillsdale, S. 1–38. 1981.

Stevens, Kenneth N. und Blumstein, S.E.: Invariant cues for place of articulation in stop consonants. In: *J. Acoust. Soc. Am.*, Band 64: S. 1358–1368, 1978.

Stevens, Kenneth N. und House, A. S.: Perturbation of vowel articulation by consonantal context: an acoustical study. In: *Journal of Speech and Hearing Research*, Band 6: S. 111–128, 1963.

Stone, M.: Toward a three dimensional model of tongue movement. In: *Journal of Phonetics*, Band 19: S. 309–320, 1991.

Stone, M.: How the tongue takes advantage of the palate during speech. In: Bell-Berti, F. und Raphael, L.J. (Hg.) *Producing Speech: Contemporary Issues*, AIP Press, S. 143–153. 1995.

Studdert-Kennedy, M.: Mirror neurons, vocal imitation, and the evolution of particulate speech. In: Gallese, V. und Stamenov, M. (Hg.) *Mirror Neurons and the Evolution of Brain and Language*, Amsterdam: John Benjamins, S. 207–227. 2002.

Suga, N.: Amplitude-spectrum representation in the doppler-shifted CF-processing area of the auditory cortex of the mustache bat. In: *Science*, Band 196: S. 64–67, 1977.

Suga, N.: Multi-function theory for cortical processing of auditory information: Implications for single unit and lesion data for future research. In: *Journal of Comparative Physiology*, Band 175: S. 135–144, 1994.

Suga, N.: Basic acoustic patterns and neural mechanisms shared by humans and animals for auditory perception: a neuroethologist's view. In: *ISCA Workshop on the Auditory Basis of Speech Perception*. 1996, S. 31–38.

Sussman, H.; Fruchter, D.; Hilbert, J. und Sirosh, J.: Linear correlates in the speech signal: The orderly output constraint. In: *Behavioral and Brain Sciences*, Band 21: S. 241–299, 1998.

Tabain, M. und Perrier, P.: An articulatory and acoustic study of /u/ in preboundary position in French: The interaction of compensatory articulation, neutralization avoidance and featural enhancement. In: *Journal of Phonetics*, Band 35: S. 135–161, 2007.

Thurstone, L.L.: A law of comparative judgement. In: *Psychological Review*, Band 34: S. 273–286, 1927.

Vago, R.M.: More evidence for the feature [grave]. In: *Linguistic Inquiry*, Band 7: S. 671–674, 1976.

Vallée, N.; Boë, L. J.; Schwartz, J. L.; Badin, P. und Abry, C.: The weight of phonetic substance in the structure of sound inventories. In: *ZAS Papers in Linguistics*, Band 28: S. 145–168, 2002.

Vennemann, T. und Ladefoged, P.: Phonetic features and phonological features. In: *Lingua*, Band 32: S. 61–74, 1973.

Wakita, H. und Fant, G.: Toward a better vocal tract model. In: *STL-QPSR*, Band 19(1): S. 9–29, 1978.

Weber, A.: Listening to non-native language which violates native assimilation rules. In: *Proceedings of the European Scientific Communication Association workshop: Sound patterns of Spontaneous Speech*. Aix-en-Provence, France, 1998, S. 101–104.

Werker, J.F.: Cross-language speech perception: developmental change does not involve loss. In: Nusbaum, H.C. und Goodman, J. (Hg.) *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, MA: MIT Press, S. 93–120. 1994.

Werker, J.F. und Tees, R.C.: Cross-language speech perception: evidence for perceptual reorganization during first year of life. In: *Infant Behavior and Development*, Band 7: S. 49–63, 1984.

Wickelgren, W.: Context-sensitive coding, associative memory, and serial order in (speech) behaviour. In: *Psychological Review*, Band 76: S. 1–15, 1969.

Wickelgren, W.: Discussion paper on speech perception. In: Gilbert, J. (Hg.) *Speech and Cortical Functioning*, New York: Academic Press, S. 237–262. 1972.

Wiese, R.: Schwa and the structure of words in German. In: *Linguistics*, Band 24: S. 697–724, 1986.

Wiese, R.: *The Phonology of German*. The Phonology of the world's languages. Oxford: Oxford University Press, 1996.

Wilhelms-Tricarico, R. W.: Physiological modelling of speech production: Methods for modelling soft-tissue articulators. In: *J. Acoust. Soc. Am.*, Band 97: S. 3085–3098, 1995.

Winitz, H.; Scheib, M.E. und Reeds, J.A.: Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. In: *J. Acoust. Soc. Am.*, Band 51: S. 1309–1317, 1972.

Wood, C.C.: Discriminability, response bias, and phoneme categories in discrimination of voice onset time. In: *J. Acoust. Soc. Am.*, Band 63: S. 1381–1389, 1976.

Wood, S.: X-ray and model studies of vowel articulation. In: *Lund: Working Papers in Linguistics*, Band 23, 1982.

Wood, S.: Assimilation or coarticulation? evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops. In: *Journal of Phonetics*, Band 24: S. 139–164, 1996.

Zheng, Y.; Hasegawa-Johnson, M. und Pizza, S.: Analysis of the three-dimensional tongue shape using a three-index factor analysis model. In: *J. Acoust. Soc. Am.*, Band 113(1): S. 478–486, 2003.

Žygis, M.: Phonetic and phonological aspects of Slavic sibilant fricatives. In: *ZAS Papers in Linguistics*, Band 32: S. 175–213, 2003.

List of Figures

List of Tables

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 28.12.2009 Christian Geng