Forschungsberichte des
Instituts für Phonetik und
Sprachliche Kommunikation
der Universität München
(FIPKM) 38 (2001) 265-393

Invariance and variability in articulation and acoustics of natural perturbed speech

Anja Geumann
Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3
D-80799 München

Abstract

The goal of this study was to evaluate invariance vs. variability in both articulation and acoustics of speech production units. To keep interaction of controlled variables manageable, only a very simple subrange of speech productions was studied. Three different vowel qualities and six different consonants were examined in a VCV sequence embedded in an utterance. Beside coarticulation vocal effort was a further factor of perturbation occuring in natural speech.

The set of consonants comprised various modes of articulation (stop, fricative, nasal, lateral) all produced at virtually the same place of articulation, viz. (post-) alveolar. The range of vowel environments /i:/, /e:/, /a:/ was selected for differences in height, in order to vary coarticulatory effects between the segments. Utterances were produced at two different volume levels, viz. normal and loud speech. Experiments by others have demonstrated that higher speech volume is not simply realized as a raised sound pressure level or as raised intensity. For loud speech a number of different correlates were observed, as raised subglottal pressure (see Ladefoged/McKinney 1963), raised fundamental frequency, raised first formant, and change of segmental durations (e.g. Traunmüller/Eriksson 2000). Furthermore an effect on jaw height was observed in vowels, which is that in vowel production in loud speech the jaw has a lower position.

In earlier studies results have been presented for either articulatory (Schulman 1989) or acoustic changes (Traunmüller/Eriksson 2000) associated with higher volume. The present study examines effects of higher volume level on vowels as well as on consonants, in the articulatory as well as the acoustic channel. Data from six German speakers (5 male, 1 female) were recorded and analyzed. In the

articulatory channel jaw and tongue-tip movements were analyzed, in the acoustic domain segmental characteristics as formants, duration, intensity and fundamental frequency.

The main results can be described as follows:

- *Jaw height in vowels depends on vowel height, in the vowel production of loud speech the jaw is lowered significantly.*
- *Jaw height in consonants depends on the type of consonant (very high for /s/, /ʃ/, /t/, fairly low for /n/, /l/). Speaking at higher volume level does not have a significant effect on jaw height during (post-) alveoloar consonant production, coarticulatory effect of vowel context is mainly found with /n/ and /l/.*
- *In loud speech jaw gestures have higher amplitude.*
- *Acoustic segmental duration is changed: Vowels are lengthened and consonants are shortened.*
- *Fundamental frequency in vowel segments is raised significantly.*
- *In all vowels the first formant is raised.*
- *The second formant of the non-front vowel /a:/ is raised.*

This work has demonstrated that jaw articulation in a number of alveolar consonants is remarkably precise and that motor equivalence only plays a minor role. Moreover, it has been shown that in the face of the generally larger variability of acoustic and articulatory parameters, the results are best considered in terms of perceptual invariants. The findings also substantiate the complexity of articulatory and acoustic reorganisation in loud speech.

Zusammenfassung

Titel: Invarianz und Variabilität in Artikulation und Akustik natürlicher, perturbierter Sprache

Es war das Ziel der vorliegenden Untersuchung, Invarianz vs. Variabilität lautlicher Einheiten hinsichtlich akustischer und artikulatorischer Merkmale zu testen. Damit die Einflussgrößen bei einer solchen Untersuchung kontrollierbar bleiben, muss der zu untersuchende Ausschnitt aus der Sprachproduktion notwendigerweise klein sein. Untersucht wurden 3 verschiedene Vokale und 6 unterschiedliche Konsonanten in einer VCV Sequenz. Neben Koartikulation wurde der Faktor Lautstärke als weiterer Störfaktor, der in natürlicher Sprache auftritt, gewählt.

Die Konsonanten umfassten eine Reihe möglicher Artikulationsmodi (Plosiv, Frikativ, Nasal, Lateral), wurden jedoch alle praktisch am gleichen Artikulationsort gebildet, nämlich (post-) alveolar. Die umgebenden Vokale (/i:/, /e:/, /a:/) waren verschieden hoch, um diverse koartikulatorische Einflüsse der Segmente aufeinander zuzulassen. Als ein zusätzlicher Einflussfaktor wurden die Äußerungen mit variierender Lautstärke produziert. Einige Untersuchungen haben gezeigt, dass erhöhte Lautstärke gesprochener Sprache nicht einfach nur mit höherer Schallintensität des Sprachsignals gleichzusetzen ist. Es wurde festgestellt, dass sowohl eine Erhöhung des subglottalen Drucks (Ladefoged/McKinney 1963) als auch Effekte, wie erhöhte Grundfrequenz, erhöhter erster Formant, Dauerveränderungen einzelner Segmente zu beobachten sind (vgl. z.B. Traunmüller/Eriksson 2000). Ebenfalls beobachtet wurde ein Effekt auf die Höhe des Unterkiefers in Vokalen; beim lauten Sprechen ist der Unterkiefer weiter geöffnet. Nichtsdestotrotz bleiben lautliche Kategorien von diesen Veränderungen unberührt.

Bisher wurden Teilergebnisse zu entweder artikulatorischen Veränderungen (Schulman 1989) oder akustischen Parametern (Traunmüller/Eriksson 2000) vorgestellt. Hier sollten nun insbesondere die Effekte lauten Sprechens auf sowohl Vokale als auch Konsonanten, sowohl artikulatorisch als auch

akustisch untersucht werden. Daten wurden von sechs Sprechern des Deutschen (5 männlich, 1 weiblich) erhoben und ausgewertet. Zum einen wurden Kiefer- und Zungenspitzenbewegungen, zum anderen akustische Merkmale wie Formanthöhe, Dauer, Intensität und Höhe der Grundfrequenz untersucht.

Die Ergebnisse lassen sich wie folgt beschreiben:

- *Die Kieferhöhe im Vokal ist abhängig von der Vokalhöhe, der Kiefer wird in der Produktion der Vokale beim lauten Sprechen deutlich abgesenkt.*
- *Die Kieferhöhe im Konsonant ist abhängig vom Konsonanttyp (sehr hoch für /s/, /ʃ/, /t/, ziemlich niedrig für /n/, /l/), lautes Sprechen führt zu keiner signifikanten Veränderung in der Kieferhöhe bei der(post-)alveolaren Konsonantproduktion, koartikulatorischer Einfluss der Vokalumgebung ist nur für /n/ und /l/ feststellbar.*
- *Die Kieferbewegungen beim lauten Sprechen werden im Schnitt größer.*
- *Im akustischen Signal kommt es zu deutlichen Dauerveränderungen. Beim lauten Sprechen werden Vokale länger und Konsonanten kürzer.*
- *Ferner wird die Grundfrequenz der Vokale deutlich erhöht.*
- *Der erste Formant wird für alle Vokaltypen erhöht.*
- *Beim nichtvorderen /a:/ kommt es ebenfalls zur Erhöhung des zweiten Formanten*

Im Rahmen dieser Arbeit konnte gezeigt werden, dass Kieferartikulation in alveolaren Konsonanten erstaunlich präzise ist und motorische Äquivalenz einen geringeren Stellenwert einnimmt. Ferner, dass bei ansonsten insgesamt starker Variabilität in akustischen wie artikulatorischen Parametern offensichtlich am ehesten von perzeptiven Konstanten gesprochen werden kann. Zudem konnte die komplexe Natur artikulatorischer, wie auch akustischer, Reorganisation beim lauten Sprechen untermauert werden.

Preface

This thesis was inspired by the theoretical assumptions of a project on „the nature of phonetic targets“, funded by the German Research Council as part of a large group of projects on speech and language production. The project was supervised by Phil Hoole, a further collaborator was Christian Kroos, principal investigator was Hans G. Tillmann. I would like to thank all of them for enabling this project. Within the data processing, especially of the articulatory data, the perpetual help of Phil Hoole is acknowledged, who as well developped the software, for processing and analysing the EMA data.

I would like to thank all the people at the phonetics lab in Munich for the nice atmosphere, making it enjoyable just to be there.

A lot of people helped in different ways writing the thesis of whom I would especially like to mention Klaus Härtl, Christiane Hofbauer, Hartmut Pfitzinger, Uwe Reichel, Sigurd Rosenau, Felix Schaeffler, Rolf Siepmann, Robert Summers, Karl Weilhammer, Anke Werani, Andreas Zierdt.

For moral support I would like to thank my parents and Markus Hiller.

The final printed version benefitted from helpful comments on form and content especially by Phil Hoole and Bernd Pompino-Marschall.

Diese Arbeit ist die überarbeitete Fassung meiner Dissertation eingereicht im März 2001 bei der Philosophischen Fakultät für Sprach- und Literaturwissenschaften II der Ludwig-Maximilians-Universität München für das Fachgebiet "Phonetik und Sprachliche Kommunikation".

Content

1. Introduction

The search for invariant empirically observable units has a long tradition in the field of phonetics. When coming from linguistic phonological descriptions that deal in strictly divided entities the next plausible step is to look in the field of phonetics for observable and concrete entities which form an equivalent to or spellout of the abstract unit.

Phonetic experiments have revealed that there is a high variability in the acoustic signal depending on context. Yet, people usually understand speech even in an extremely noisy surrounding (e.g. while car driving), but analysis of the acoustic speech signal then is extremely difficult if not impossible. So, an obvious step in scientific progress was - as techniques improved - to look more deeply into articulatory movements and their suitability for detection of units. This is motivated not least by the long tradition in using articulatory terms for an abstract description (e.g. Bell (1867), Sweet (1877), Sievers (1881), Passy (1890), Rousselot (1897-1901), Jespersen (1904)).

To test the characteristics of potential articulatory units experiments were conducted which perturbed the usual articulatory behaviour. The most well known experiments were those with bite blocks (e.g. Lindblom/Lubker/Gay 1979). Here, the position of the jaw was fixed with a small block. The interest was then to find out whether during sound production the (missing) jaw movement was compensated for by an adapted tongue movement. Results should reveal information about the nature of planning units.

This precise knowledge is highly important not only from a theoretical perspective. Detailed knowledge about the characteristics of motor planning will be the basis for anthropomorphic speech synthesis and may help to improve speech recognition in a multimodal framework (e.g. Blackburn/Young (1995), Zlokarnik (1995a, b)).

The fundamental problem with such bite block experiments is that the whole setup is extremely artificial. One of the very few real life situations where one might actually have something like a bite block condititon is a pipe or cigar smoker talking while keeping the cigar or pipe in his mouth.

And even if the number of cigar smokers is increasing, the impression lasts that talking while smoking is a pretty rare phenomenon. So, obviously one should look for more natural sources of perturbation that are likely to exist in all or most communicative situations. Some of them, viz. loud speech (see Lindblom 1990, Schulman 1989) and coarticulation (see Edwards 1985) are looked into in this work.

In chapter 2 the role of speech production in a more general model of language production is outlined, with focus on some relevant results about motor control, motor equivalence and speech specific compensatory articulation.

Chapter 3 briefly discusses coarticulation, as a major source of variability.

Chapter 4 presents results from the literature about loud speech, a further source of variability.

Chapter 5 - 7 will present own experimental data and results, which will be discussed in chapter 8.

Chapter 9 presents some general conclusions of the results and how they fit into a more general model of speech production.

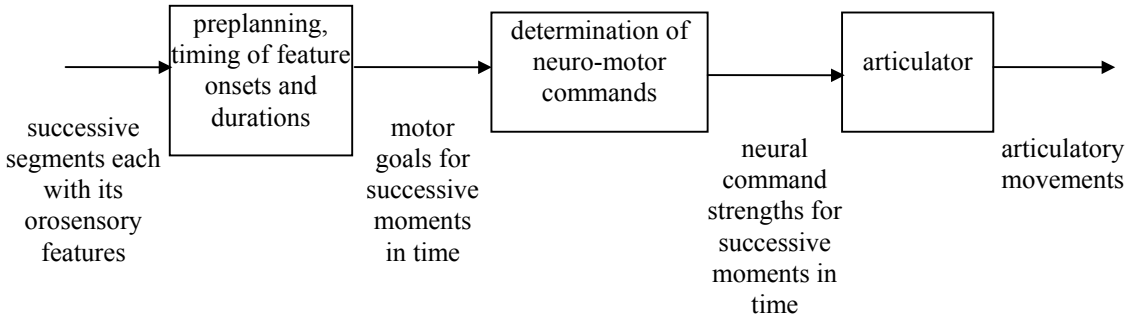
2. Speech production in a model of language production

If we consider the search for invariance in kinematics or acoustics as the attempt to find a measurable quantifiable correlate of a segment, we should look closer into the role that this segment does play in a model of language and speech production. Or in shorter words, the part of phonetics in a production model. A current popular model is Levelt's (1989) speech production model, it being one of the most comprehensive models, so far. His work of 1989 has been recently revised (Levelt et al. 1999) but the part relevant for us here is more extensively focussed in the older version, and has not been revised. Levelt's model of speech production is attractive for this study since it is cognitively oriented and insofar interested in psychologically adequate modelling. Other more specific models of speech production such as Guenther et al. (1998) will be discussed further below. They do not contradict the general model, but supply more detail.

While it has been the topic of a variety of phonetic theories to discuss the nature of the representation, it goes without saying, that speech at some stage has to be executed motorically and is transferred to the listener mainly in the acoustic channel as modulated sound waves. There are recent proposals to account for other than acoustic input to the receiver (or output of the sender of spoken language), as indicated by Jackendoff (1997). This would account for the fact that especially paralinguistic information is often encoded in facial expression or body language. It allows as well for findings of audio-visual integration to play a role in speech perception.

More detailed descriptions about the nature of the control of articulatory muscles are presented in the following subchapter, since Levelt remains here somewhat unspecific or cites other authors.

2.1 Oro-sensory targets, motor equivalence, motor control



Perkell (1980) cited in Levelt (1989:443)

Figure 2-1. Model of speech production from Perkell (1980)

We will here try to reveal more about the nature of the motor commands that are stored in the articulatory buffer. It had already been stated that Levelt speaks of one articulatory gesture per phone. It is however clear that these gestures however abstract, are still influenced by gestural context, vary with speech rate, show patterns of compensation, as in bite block speech. It can thus be concluded that gestures have to be context dependent. A framework that especially dealt with these topics is Articulatory Phonology (Browman/Goldstein 1986ff). Articulatory gestures are described as trajectories in time and space that are defined in simple physical terms of a mass-spring model, which accounts for an intrinsic description of timing. These gestures then assume that articulatory movement is oriented towards a certain target. Trajectories of different gestures blend with each other; on the other hand this model states that articulatory gestures can be retrieved by the listener in the spirit of the motor theory of speech perception. In this sense Articulatory Phonology has to be called a purely articulatory framework.

Other approaches have then tried to conclude in the spirit of „we speak in order to be understood“ that articulation should be somewhat more connected to „distal“ auditory targets, that were learned during speech acquisition (Perkell 1980). Then a learned auditory feature has a certain vocal-tract state, that is formed by certain articulatory correlates, which provide tactile and proprioceptive information, so called „oro-sensory information” (see Figure 2-1).

Oro-sensory goals as in Perkell (1980) and oro-sensory feedback are used as well in a similar newer framework by Guenther et al. (1998).

Another similar approach is the internal reference model of Lindblom et al. (1979). Targets are here certain vocal tract states for a specific area function producing an acoustic signal that results in an auditory impression.

2.1.1 Motor equivalence and coordinative structure

All of the above mentioned approaches have in the one way or the other to deal with the fact that articulation is to some extent variable, even if the perceptual result remains constant. Bite block experiments revealed movements of the tongue to compensate for the jaw in vowel production, at least in the presence of some oral afferent information (Hoole 1987). There are as well trading relationships between different articulators reported, that are even less coupled. Perkell et al. (1993 and later Savariaux et al. 1995) showed that tongue back position and lip rounding in the production of /u/ can show some compensatory behavior: they are linked for functional purposes.

A strong view of the interaction of single articulators was introduced by the work on „coordinative structures“ (Fowler 1986, Kelso et al. 1986). They are formed by groups of muscles that form „synergisms“ that carry out complex movements automatically. The muscles in a coordinative structure then work together to achieve functional equivalence.

„Motor equivalence“ and „coordinative structures“ both assume that jaw and tongue act „as compound rather than component gestures to represent functional entities“ (Hertrich/Ackermann 2000: 2236). Other works see the jaw as a more independent structure, which with its cyclic movements forms the underlying basis for syllabic organization of speech. Vowel and consonant articulations are then superimposed on this jaw cycle, with vowels associated with the jaw opening phase and consonants with jaw closing (Keating 1983, Lindblom 1983).

3. Coarticulation

In the following the term coarticulation will be further explained. Firstly, one has to differentiate the types of articulatory processes that may contribute to coarticulation. Next, one has to explore the temporal domain over which coarticulation can show an effect. An important distinction was made first by Menzerath and de Lacerda (1933) who differentiate between „Koartikulation“ and „Steuerung“. The definitions are given in (3-1) and (3-2), the English translation was done by me, a.g.).

(3-1) Koartikulation - Coarticulation

(3-1a) „Synkinese“, „Bewegungsverflechtung“ interlacing of movements of nonhomorganic sounds. The simultaneous activity of independent articulators. (after Menzerath/deLacerda 1933:53)

(3-1b) „Beim Zusammenwirken von Organen unterschiedlicher Artikulationspräzision kommt es zu zeitlichen Verschiebungen im gegenseitigen Bewegungseinsatz, d.h. entweder zu einem Beharren oder zu einer Vorwegnahme in einem Artikulationsparameter. [... Es] wird nicht Bewegung eingespart, sondern sie kommt zu früh oder zu spät“ (Kohler 1977:208)

In the combined action of organs of different articulatory precision temporal postponements in their movement onset occur, either perseveration or anticipation of an articulatory parameter. Articulation is not economized but it starts too early or too late

(3-2) Steuerung - Control (Navigation)

(3-2a) „[D]er Vokal entwickelt sich, vom folgenden Konsonanten gesteuert, in Dauerbewegung bis zu dem betreffenden Konsonanten [...] bei homorganen Lauten.“ (Menzerath/de Lacerda 1933:21)

The vowel develops, controlled by the following consonant, in permanent movement to that consonant in homorganic sounds

(3-2b) „[Zungenspitzen-]Artikulationsreduktion soll Steuerung auf einen Fokus hin genannt werden. So bilden velare und labiale Kontoide, d.h. Zungenmasse und Lippen, einen Fokus, auf den hin die Artikulation gesteuert wird. Interveniert eine Zungenspitzen-artikulation in dieser Steuerungsbewegung, dann kann sie eliminiert werden.“ (Kohler 1977:209)

Articulatory reduction (of tongue-tip movement) shall be called navigation towards a focus. Velar and labial contoids, viz. tongue-body and lips are regarded as a focus on to which that articulation is navigated. An intervening tongue-tip articulation may be eliminated.

Very often both (3-1) and (3-2) are called coarticulation, and it depends on the context if homorganic movements affect each other (3-2a) or if a non-homorganic articulation superposes another (3-1). Following the definition in (3-2a) control („Steuerung“) is regarded as a strongly local effect on the trajectories between two adjacent homorganic segments. Kohler is here representing a different point of view, he is making a functional distinction between vowel and

consonant articulation. The tongue-tip articulation of a consonantal gesture can thus be controlled or navigated by other non-homorganic consonantal articulatory movements.

Coarticulation in the sense of (3-1) can be effective over quite longe time spans. An articulator may spread or move its activitiy onto neighbouring sounds that do not need that articulator. This in fact requires a sometimes problematic distinction between active and not active articulators. The tongue might in a very first broad definition be described as a single articulator. But there are very important reasons from a physiological point of view to differentiate further between tongue-tip and tongue-body articulation, since the first is more evoked by intrinsic muscles, while the tongue-body position is more defined by extrinsic muscles. In the following one has to define with great care which articulator is to be treated as independent.

In a broader sense, coarticulation fundamentally underlies speech production in general. Coarticulatory effects might be recognized by the listener as assimilation of a single feature (e.g. Kühnert 1996) In other examples the effect might be noticable mainly for the experienced listener, as in the control phenomenon [k̟i] vs. [k̠u].

There have been especially lately some experiments testing the time span of coarticulatory effects (West 1999, 2000, Heid/Hawkins 2000). An earlier work of Kohler/van Dommelen/Timmermann (1981) described for Dutch and French that devoicing of phonologically voiced obstruents is influenced by the frequency of voiceless consonants in a phrase. Experimental results for spreading of features like fronting or velarization, induced by /r/ or /l/ reported effects over one syllable or even a foot (Barry/Hawkins 1992), two syllables (West 2000), or even more than two syllables (Heid/Hawkins 2000). A language specific process in continuous speech is the assimilation of voiceless segments to the voicing of adjacent segments in French or Russian. Opposed to that, in English, German, or Swedish voiced segments tend to assimilate to an unvoiced environment (Barry/Hawkins 1992:142). This then can no longer be accounted for as an articulatory process, based on general principles of economy of movement, but represents language specific settings that might not be explained by principles of coarticulation.

3.1 Models of coarticulation

3.1.1 Segmental model

Coarticulation that results in assimilation in a segmental model means the adoption of segmental feature values within a feature matrix. Features represent „dimensions of articulatory control“ (Keating 1985:3). Important means are here underspecification and feature spreading. Feature values of specified segments (‘±’) are spread or copied onto preceding unspecified segments (‘0’). In (3-3) this look-ahead mechanism (after Henke 1966) is described. The result is a complete specification. The completely specified segmental chain is then analysed in the phonetic component of a speech production model, which assigns the abstract features quantitative values.

(3-3)

	S1	S2	S3		S1	S2	S3
F	0	0	+		+	+	+
G	+	0	-	→	+	-	-
H	0	-	+		-	-	+

F, G, H are features

S1, S2, S3 are segments

3.1.2 Two-tier model

This model was developed to explain vowel-to-vowel coarticulation. Öhman (1966) measured the formant values F1 and F2 at the transitions to the consonant in VCV sequences. The results show that the value is influenced by V1, C, and V2. Thus, the vowel-to-vowel coarticulation can be interpreted as bidirectional, or as an interpolation between vowels. Features of vowels and consonants are exclusive and on separate tiers, called by Öhman „channels of articulation". These two channels are complementarily underspecified; this is motivated articulatorily, since articulations of vowels and consonants can be regarded in a way as independent. Even labial consonants may be additionally rounded depending on vowel context. So there is in this two-tier model underspecification as well as in the segmental model, but no feature spreading. Underspecified segments remain unspecified. Keating (1988) analyzed this as „phonetic underspecification", which means that phonological underspecification in some cases is preserved in phonetic realisation. Nevertheless, adjacent segments on the other tier are considered, as for the velar plosive being controlled by the adjacent vowel [k̟i] vs. [k̠u]. In Öhman's (1966, 1967) model vowel movements form an uninterrupted diphthong-like continuum, onto which intervening consonants are superimposed. Then, the effect of vowels on consonants is only a byproduct of the effect of vowels on each other. Gay (1977) noted that this diphthong-analysis might be a slight oversimplification. He noted that even with an intervening labial consonant vowel gestures show some hiatus between them, i.e. temporal coordination of vowel gestures is affected by intervening consonants. Coarticulation between consonants in Öhman's model is assumed to work analogously to that between vowels, but no specific data are presented. Exceptions to the distinction between vowel and consonant articulation are r-coloured vowels and nasalized vowels, they both use additionally the consonantal channel of articulation. Consonants with a secondary articulation, like palatalization, velarization, pharyngealization, or labialization use additionally the vowel tier. An argument (after Keating 1985:6) for this analysis is vowel-to-vowel coarticulation in English and Swedish and its absence in Russian. While English and Swedish do not have a secondary articulation of stops, Russian stops are either palatalized or velarized.

(3-4)

vowel tier	[αF]	[βF]	[γF]
	|	|	|
	V	C	V
		|	
consonantal tier	[δG]		

Following Öhman's model they use additionally the vowel tier. Since coarticulation is then regarded as strongly local, in the sense that it cannot pass over a specification at the same tier, the vowel features [αF] and [γF] in Russian cannot influence each other with an intervening stop specified on the vowel tier [βF].

4. Loud speech

First we should look closer into the literature on different aspects of loud speech.

It was suggested by Schulman (1989) to regard loud speech as a natural source of perturbation. Since speaking at different volume levels is a very common phenomenon in natural speech, we find this a very attractive source of variability in the experimental setup that will be described in the following chapters.

4.1 Why do speakers speak loud?

The reasons why speakers increase their volume can be broadly divided into two categories. First, there is an objective reason to increase vocal effort, either because the listener is known to be somewhat distant, or to be hard of hearing or the speaker simply wants to amplify volume to underline an argument. Second, there is an interesting effect of speaking loud without an actual need by the listener. This is subsumed under the term Lombard effect (after Etienne Lombard's „Le signe de l'élévation de la voix", 1911). A speaker who is speaking under noise, no matter whether the listener is in the same noisy environment, increases volume level considerably, even up to shouting. There has been some amount of literature on the topic of the Lombard effect (e.g. Lane/Tranel 1971, VanSummers et al. 1988). The results in general show, beside a higher volume level, that speakers accommodate rather differently. In addition, experiments (Pick et al. 1989) revealed that it is all but impossible for the speaker to suppress the increase of volume under noise. Since we are interested in loud speech effects that are intentionally produced by the speaker we will in the following focus on loud speech that is produced by the speaker to account for the listener.

4.2 Articulation of loud speech

An early interesting work examining loudness, sound pressure, and subglottal pressure is that of Ladefoged and McKinney (1963). They stated that the volume velocity of air that is flowing through the glottis is proportional to subglottal pressure. The rate of work done upon the air in producing a voiced sound is proportional to the square of subglottal pressure.

Ladefoged and McKinney (1963) showed that listeners' judgements of the loudness of syllables were more closely correlated with the subglottal pressure that was needed than with their SPL. They found further a proportional relationship between subglottal pressure and transglottal volume velocity, the product of the two factors is proportional to the work done on the air. They further concluded that since subglottal pressure and transglottal volume velocity are proportional, one could instead take the square of subglottal pressure to be proportional to work done on the air. As a result it could then be stated that listeners judge the work done by the speaker on the air instead of SPL. This is, however, as Ladefoged and McKinney state, only the case for speech sounds. For non-speech sounds, loudness is directly related to SPL, that is a direct result of the distance to the sound source.

Recently, it has been discussed by Eriksson and Traunmüller (1999) that Ladefoged and McKinney (1963) did not describe loudness, but what they call vocal effort. The literature in

most cases is using the the terms normal and loud voice to refer to the same topic. We will here not argue for the one or the other term, and further use a variety of terms rather liberally. It should however be stated that if we speak in the following chapters of loud speech or higher volume level, it describes a subjective dimension chosen by the speaker to accommodate to a listener who is more distant (viz. as in our experiment in the room next door).

There is not much literature about vocal tract articulation in loud speech besides the work of Schulman (1989). In Swedish vowel production he found a lower jaw position in loud voice condition. This is an interesting finding for our purpose to establish sources of perturbation in natural speech. Since articulation of tongue and jaw are to be varied as much as possible, a natural source for jaw lowering could evoke interesting patterns. More literature is found in the context of production of stress (de Jong/Beckman/Edwards 1993, de Jong 1995), which seems to show similar effects. It has however to be discussed elsewhere in more detail to what extent the same strategies are used for loud speech and stress realisation in a specific language.

4.3 Acoustics of loud speech

More work has been done to examine the acoustic properties of higher volume level. One rather interesting work that has apparently not received too much attention yet, beside the experiments of Traunmüller/Eriksson (2000), is that of Wilkens and Bartel (1977). They examined playback loudness as perceptual cue for listeners to identify original loudness. They could show that listeners are able to identify the original loudness of even an unknown speaker with high precision, independent of playback loudness.

Acoustic characteristics of shouted voice include longer vowel durations and shorter consonants, as was found by Bonnot and Chevrie-Muller (1991) for French data. This work again highlights the problem of comparing studies in the broader field of loud speech. It is certainly problematic to compare shouted and loud voice, although sometimes they might be functionally equivalent. The two modes are at least differing in voice quality, which means that they differ in the aerodynamic aspect of production. Thus, we cannot rely on results obtained for shouted voice to be found as well in loud voice.

Work of Rostolland (1982a, 1982b) refers as well to characteristics of shouted voice. He found the first formant and fundamental frequency to be increased in shouted voice. Newer work by Traunmüller/Eriksson (2000) confirms effects on first formant and fundamental frequency for higher vocal effort as a result of increased speaker-listener distance. Another cue for loud speech is an increase in the prominence of higher frequency components. It can be suggested that this loud speech effect is comparable to patterns of higher frequency spectral emphasis, i.e. differences in spectral slope for linguistic stress (Sluijter et al. 1995ff). Another pattern that might be related to spectral effects for loud speech, is the so called singing formant, a technique used by professional singers.

To conclude, there has been a variety of work on the subject of speaking loud and perceiving loud speech. However, terminology is sometimes somewhat misleading, and one has to be careful to compare different findings, especially that for shouted as opposed to loud voice.

5. Experimental Data

5.1 Experimental setup

Kinematic and acoustic recordings were made of read phrases, produced by 6 German speakers (one female (AW), five male). Pseudo-word 'VCV sequences were embedded in carrier phrases of the type "Hab das Verb ___ mit dem Verb ___ verwechselt". Both target words received contrasting sentence accent.

The target consonants were the alveolar German phonemes differing in manner of articulation /s, ʃ, l, n, d, t/ (/ʃ/ is postalveolar). They were placed in differing symmetric vowel height contexts /i:__i:, e:__e:, a:__a:/; both vowels were long, with main stress on the first vowel. All phrases were produced in loud and normal speech, which was elicited by simple instruction of the speaker. The loud and normal phrases were presented in random order. For each target consonant with given loudness and vowel context 12 repetitions were produced, i.e. 72 repetitions of each consonant over all context and loudness conditions.

For one speaker (hp) a slightly different corpus was used containing symmetrical vowel contexts as well as asymmetrical combinations of /i/ and /a/. Furthermore, this corpus only contained five target consonants /s, ʃ, l, n, d/. For each target consonant with given loudness and vowel context 10 repetitions were produced, i.e. 80 repetitions of each target consonant over all context and loudness conditions.

All speakers were students, graduate students or faculty of the Institute of Phonetics and Speech Communication, University of Munich. They are all native speakers of German, aged between 23 and 31. They were at the time of the recordings not familiar with the details of the experiment. No dental problems were reported. Dental casts of all speakers were manufactured by an external dentist, including a precise cast of the palate contour. In further sessions stuctural MRI images of static sounds were recorded for all speakers, additionally some anatomical information concerning the position of the mandibular condyle was obtained via MRI images. Some results of the MRI data are presented in Hoole et al. (2000). Here, only the information about positioning of mandibular condyle is further regarded, but for some future work this combined information might be rather helpful.

The instruction for different volume levels was given as:

- Modal voice: Speak at your own comfortable volume level. Stimuli on screen have not been marked aditionally.
- Loud voice: Speak as loud as possible (below shouting). Imagine you have to be heard in the control room (see Figure5-2), while the microphone is off. Stimuli have been marked additionally in the next row as (LOUD).

The use of more natural target words was considered. But even extensive lexical searches revealed insufficient German word material for all symmetrical vowel contexts combined with different consonants. A further corpus was then generated, which contained consonants in a context of full vowel in combination with Schwa (V-C-Schwa). This on the other hand can not easily be compared to the data obtained with the nonsense material, simply because of differing vowel contexts. Those data will not be reported on further here.

Although the nonsense material is a bit different from sequences commonly found in German words, many sequences have an equivalent in the real word. Examples are „Solidität“ /idi/; „PDS“ /ede/; „Kanada“ /ana/, /ada/ and so on. The stress pattern with stressed first vowel and unstressed second vowel is typical for German. Only one of the speakers (sr) showed a strong tendency to reduce the vowel quality of the unstressed second vowel. Especially /e/ was very often substituted by /ə/ (for speaker sr).

The midsagittal kinematic signals were recorded with an electromagnetic transduction system with 500Hz sampling rate. (Articulograph AG100, Carstens Medizinelektronik, for more technical details see Hoole [3]). Four sensors were placed on the tongue (referred to as tip, blade, dorsum and back): The tip sensor was placed approx. 1cm posterior to the tongue tip, and was assumed to best track alveolar articulation. The other three followed in equidistant steps up to a point opposite the junction of hard and soft palate (blade, dorsum, back). Three sensors were used to track the jaw movement. One each was placed on the inner (jaw-in) and outer (jaw-out) surface of the gums beneath the lower incisors, a third sensor was placed on the angle of the chin (chin). Reference sensors were located on upper jaw and the nasion.

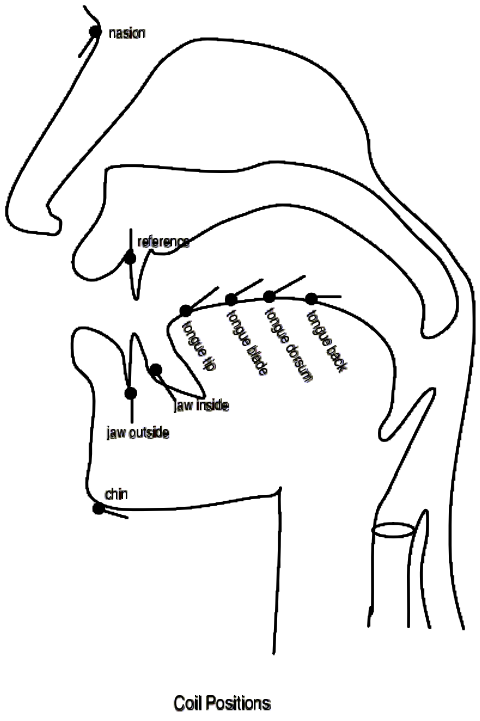


Figure 5-1. Coil positions used for this experiment

Acoustic signals were recorded with a Sennheiser MK H20 P48 microphone with omnidirectional characteristics. Distance between speaker’s mouth and microphone was approx. 30 cm. Sampling rate was 24kHz and data were stored on a 8 track Sony PC 208Ax DAT recorder, together with a synchronizing signal to enable synchronization with articulatory data.

Above the stimuli screen a digital video camera was mounted. The speakers were aware of this. Data were stored digitally and synchronized with the acoustic signal. Video data were not analyzed in this study.

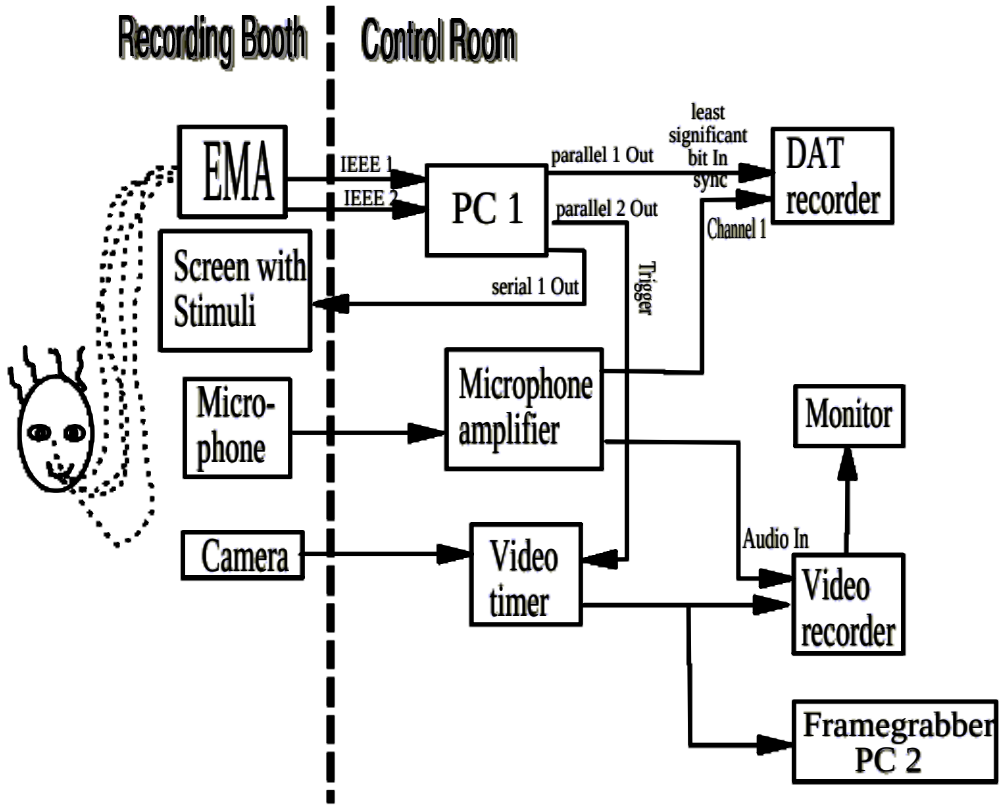


Figure 5-2. Experimental setup. During the recording the investigators were in the control room, while the subject was in the recording booth. The two rooms are separated through a regular not very thick wall, with a small hole for different cables.

5.2 Acoustic segmentation

All acoustical data were in a first step segmented manually by the author. Since those acoustical segmentations form the basis of further articulatory and acoustic analysis, they are outlined at this point.

The segmentation of the acoustic signal is based on the following guidelines.

The start of the sentence „Hab das Verb VCV mit dem Verb VCV verwechselt“ /haːp das VEɐp ___ mIt deːm VEɐp ___ fEɐVEksəlt/ was defined by the offset of initial /h/ friction and the

begin of periodicity of the first vowel /a/. End of sentence is defined by the release of the final stop /t/.

The VCV nonsense words are defined by the sequence of non overlapping segments V1, C, and V2.

V1 begins with the release of the preceding stop /p/ or the release of a prevocalic glottal stop frequently found in German.

C begins with the end of vowel-like periodicity; in cases, viz. /n/ and /l/, where a clear boundary could not easily be detected, the auditory absence of identifiable vowel quality was taken as a second criterion.

V2 begins with the release of the preceding stop (thus includes VOT) or the end of frication in case of /s/ and /ʃ/. The boundary between /l/ or /n/ and V2 is determined in the same way as that between V1 and nasal or lateral. The end of V2 is defined by onset of the following nasal /m/ for the first target word and the onset of frication of /f/ for the second target word.

The VOT part of V2 (for target /t/) was segmented as well, but is not regarded here any further.

6. Articulatory Results

Before the results for the articulatory data are described a few further comments on the processing of the articulatory data have to be made.

The processing of the articulatory data recorded with the Carstens AG 100 system, has been described in detail elsewhere, e.g. Hoole (1993, 1996), Kühnert (1996), Mooshammer (1998) or more generally in Perkell et al. (1992). The raw data were corrected for head movement, and the post-processed data were rotated, so that the x axis is parallel to the occlusal plane, with the origin at the position of the reference coil above the upper incisors. To obtain the occlusal plane each speaker's dental cast (including second molars) of the upper teeth was laid out on a horizontal surface with the teeth facing down. The palate contour from this dental cast was obtained with a sensor coil moved along the midsagittal plane. For each speaker this contour was mapped with the palate contour obtained during the recording session. The actual palate contour was obtained by moving a coil on the investigator's fingertip along the palate. The rotation to the occlusal plane is a standard procedure which is used as well in microbeam data. The processing of the data was done in Matlab, using routines and macros written by Phil Hoole.

Although four EMA coils were placed on the subject's tongue the following results here will focus on the tongue tip and jaw positions. This means that we are using the consonantal place of articulation to analyze the kinematics of the adjacent vowels as well. Although this does probably not reflect the vowel articulation ideally, especially for the /a/, no other place of articulation would have ensured to cover the consonantal articulation. Since we regard vowel articulation less as forming a constriction at one single point but as shaping the whole vocal tract, even the front part of the tongue should reflect a specific vowel target to a wide extent. (Compare to Hoole/Kühnert 1996 or Hertrich 1998). So the lowest position of the tongue tip in proximity to the acoustic vowel segment will be identified with a vowel target position, whereas the maximal vertical position between the two minima is identified with the consonantal target position. As we will see in some of the results, for all speakers the tongue tip coil might be a good estimation of the place of /s, n, l, d, t/ production, but missed at least for most speakers the relevant position at the tongue surface for /ʃ/. However, the adjacent tongue blade coil was positioned too far back. This was not a mistake in the experimental setup, but technical limitations require at least some minimal distance between coils, to avoid interferences.

6.1 Intrinsic tongue

One important point of interest in the analysis of the articulatory data will be the covariation pattern between jaw and tongue height. This investigation unfortunately is not technically trivial at all. Since there is a mechanical coupling between tongue and jaw, the positional data of the tongue always contain a proportion of jaw movement. I.e. the same lingual height position under the condition of different jaw heights, means in effect different tongue movements. To account for the influence of the jaw on the tongue movement, different approaches have been used in the literature. The most simple estimation of the then so-called intrinsic tongue movement is a simple subtraction of the jaw vertical position. It has been shown in great detail

by Edwards (1985) that jaw movement in speech production can be described best as largely rotational around a joint, the mandibular condyle (Figure 6-2). A further component of jaw movement is horizontal and vertical translation. Both are shown in Figure 6-1. For a more recent extensive overview on how to estimate the jaw contribution and more elaborated solutions see Westbury et al. 2002.

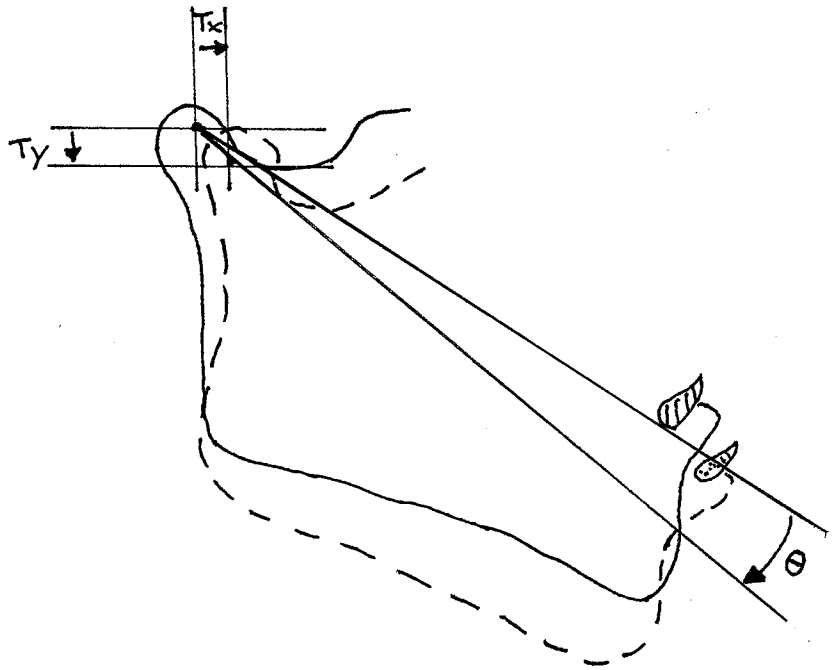


Figure 6-1. Based on Edwards/Harris (1990) Jaw movement as a combination of rotational (θ) and translational (vertical: TY, horizontal: TX) parts.

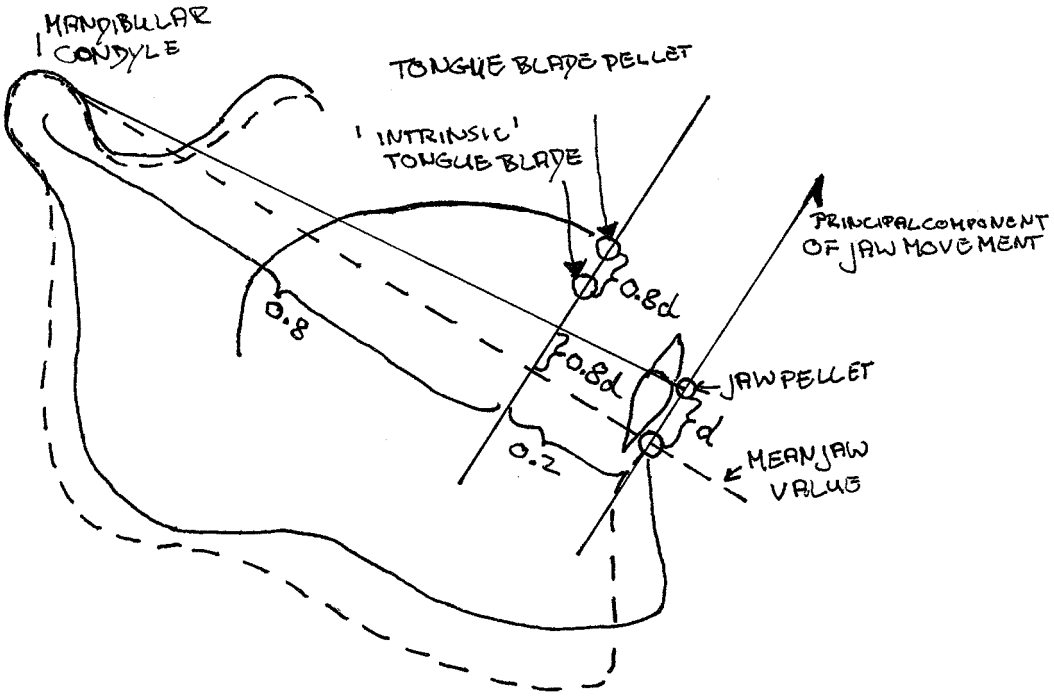


Figure 6-2. Based on Edwards (1985). Jaw movement considered as rotation. Intrinsic tongue movement extracted by subtracting jaw from tip position, whereby jaw is weighted with the „tongue-to-condyle-distance" factor, here 0.8 for blade.

The procedure that was used here to estimate the tongue condyle distances is demonstrated in Figure 6-3. A critical point remains open, viz. with this procedure the tongue jaw coupling is assumed to have the same strength for tongue tip as well as tongue back. It appears that this might underestimate the relative independence of the tongue tip. On the other hand no clear counterarguments could be found that would make it necessary to reject this intrinsic tongue tip extraction method. We have to leave this question open for further research.

It has been further proposed by Edwards (1985) that the effect of jaw movement on tongue movement can be estimated as in Figure 6-2. The distance between tongue coil and condyle in percent of the distance condyle to jaw coil is taken as a weighting factor when subtracting the jaw position from the tongue position. This method was used in our data, since we were able to obtain for every speaker the exact position of the mandibular condyle by structural MRI (Siemens Magnetom Vision, 1.5 Tesla) sequences. These data were mapped by Phil Hoole on to the palate contour information of the EMA data and could thus be quantitatively compared. In Table 6-1 distance information for every speaker and every tongue coil is listed.

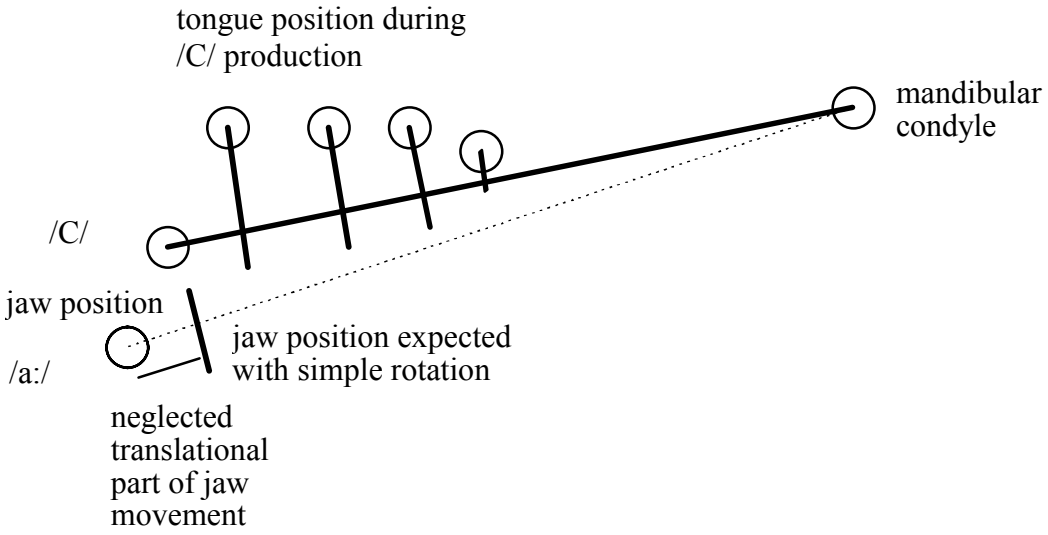


Figure 6-3. Distance between mandibular condyle and jaw coil was determined for each speaker. Positions of the jaw coil and the four tongue coils at the acoustic mid point of consonant were used to estimate the distance between tongue and condyle. The vertical axis was simply put at right angles to the straight line linking jaw and condyle parallel to the maximal observable rotational effect.

Speaker	distance between condyle jaw-out in [mm]	maximal observed rotation in [°]	neglected translational part of jaw movement in [mm]	distance between condyle and tongue tip in [%]	distance between condyle and tongue blade in [%]	distance between condyle and tongue dorsum in [%]	distance between condyle and tongue back in [%]
aw	90.5	6	2.5	76	69	59	67 (sic!)
hp	114	6	4	79	69	61	53
kh	98	2	0	85	78	69	57
rs	98	4	5	78	72	65	57
sr	96	5	8	80	68	63	54
ur	100	3	0	83	73	69	63

Table 6-1. The values for the measures outlined in Figure 6-3. It should be noted that the translational part of jaw movement was only observed in /a:/ production. During /e:/ or /i:/ production the observed translational effect was virtually nonexistent even for speakers sr, hp, and rs.

6.2 Kinematic segmentation

Our further kinematic analysis concentrates upon the data of tongue tip and jaw_out coil's vertical position. The chin coil is ideal with respect to distance to condyle but is more prone to errors as it might show skin movements unrelated to jaw movement. The jaw_out coil seems to represent best the jaw movement as it is further away from the condyle than the jaw_in coil. As we are looking at VCV sequences where the C is a (post-)alveolar consonant and the sequence is embedded in a labial consonantal context the tongue tip is thought to represent the clearest tongue movement to and from the consonant target. While tongue coils positioned further back

are more likely to be closer to the place of closest constriction for front vowels, they will not really catch the (pharyngeal) target region for the back/central vowel. Since the tongue as a whole can be understood to form a characteristic vowel position and since the tongue tip will show the clearest movement for the intervocalic consonant, the tongue tip coil is chosen to represent the tongue positions for consonants and vowels equally. In the following the two positions will be referred to as *jaw* and *tip*. For extraction of all the kinematically relevant points, the two and the derived *intrinsic tip* have been analysed separately from each other. Based on the segmentation of the acoustic signal, the VCV word segments have been used by the author to identify interactively well-defined kinematic positions. To ensure that all kinematically relevant information of the VCV sequence is covered an advanced V1 onset was used, which adapts to the speech rate of the individual speaker for the VCV sequence. The onset for speakers rs, sr, hp was -30ms; for speaker ur -40ms, for speakers aw and kh -50ms. The actual algorithm identified the absolute minimal position (minp1) within the acoustic V1 segment with the individual early onset. Then within the VCV word segment following the first obtained minimal peak (minp1), an absolute maximum (maxp) is found, and then again the following minimal peak (minp2). At the next step the points of inflection of the movement were determined, the first was determined by finding the maximal peak velocity (maxvel) between minimal peak (minp1) and maximal peak position (maxp), the second point of inflection was defined as the negative peak velocity (minvel) between maximal peak position (maxp) and the following minimal peak (minp2). These data were manually checked, and corrected if peak positions of adjacent movements were detected erroneously, which happened quite often. In some cases no clearly defined minima could be detected, in these cases, positions were determined manually by an (almost) zero velocity criterion.

The procedure does not very clearly distinguish between the two different types of (a) relations between acoustic signal and a single kinematic signal of a certain articulator, and (b) different positions within a kinematic signal, defined by position or velocity or acceleration. Although it has been shown in the literature that correlations between articulatorily defined points and acoustically defined points exist the dependency is not straightforward. Sock (1998) and others

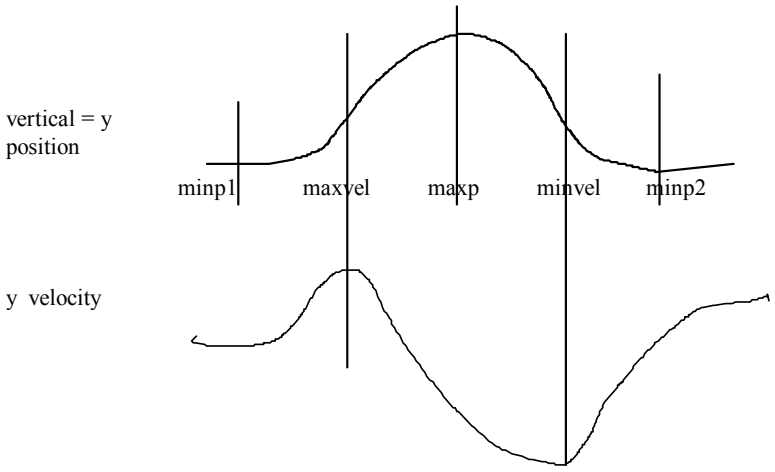


Figure 6-4. Kinematic defined points for analysis

describe peak velocity as often occuring almost synchronously with acoustic segment onset. Thus our method of looking for kinematic positions within certain acoustic boundaries should not be interpreted as much more than a rule of the thumb. On the other hand, defining the search frame for a velocity peak within two peak positions of an articulatory movement, is caused by the necessity to find a certain peak within an articulatorily motivated time frame.

Figure 6-4 outlines the kinematically defined positions, extracted for jaw, tip, and intrinsic tip. This certainly shows an idealized pattern. In a very few cases an almost inverted pattern was found, e.g. with speaker hp jaw movement for /ili/, the kinematic postions were then manually adjusted. In a variety of cases clearly distinct peaks could not be detected. This means that even if an assumed target is missed, the information for the target vertical position is still quite reliable. But durational information in these cases is certainly distorted. To loose not to much information about the vertical position of the single articulator, most of those problematic points have been adopted, since they contain valid information on the height, although not on the point in time.

6.3 Definition of effort in analogy to physical work

There have been several approaches to defining the effort that is needed for an articulatory gesture. A good overview can be found in Kirchner's dissertation (1998). The method we chose is similar to that chosen by Lindblom (1990). See Kirchner (1998: 40f) for a discussion and alternate approaches.

The amount of work necessary to produce the consonantal tongue or jaw movement was estimated with the following formula. For reasons of simplicity we assume that the V1 to C movement as shown in Figure 6-5 can be approximated by a part of a cosine function. The area below the function can be calculated in the following way:

$$W = c/2 \int_0^{t_c} 1\text{-}\cos\left(\frac{t}{t_c}\pi\right) dt$$

simple integration results in

$$W = t_c\, c/2$$

where t_c is the interval of minimal peak (first vowel) to maximal peak (consonant) and c the vertical distance between minimal peak and maximal peak (c = y(maxp) - y(minp)), i.e. formula for the area of a triangle. In the results presented in the following we have only accounted for consonantal movements that go upwards. In fact for one speaker (sr) some lowering movements in high vowel context for the (intrinsic) tongue tip are to be found. We were not sure if those negative movements should be interpreted with the same amount of work involved. Since we have come to the conclusion that the lowering in this case should be interpreted as using the same amount of work as an inverted upward movement, the work algorithm should better use the absolute value of the amplitude |c|.

This model neglects mass differences, a certainly not unproblematic assumption. On the other hand, this is common practice in damped spring-mass equations that are used, e.g. in the task dynamics model (e.g. Saltzman (1986), or see for an introduction Hawkins (1992)) or in the articulatory synthesis model by Kröger (1993).

6.4 Results for consonant articulation

In the following subchapters the influence of four different factors will be considered: (a) speaker, (b) type of consonant, (c) vowel context, (d) volume level. Since a variability between speakers can be expected and has been found in probably all studies on articulation, results will be shown in most examples for individual speakers. The statistical analysis was carried out with the SPSS package (version 10).

6.4.1 Jaw articulation in consonants

A multifactorial analysis of variance (general linear model, four fixed factors) revealed that the jaw position at the consonantal target position is affected by speaker $F=7347$, $p<0.001$, consonant type $F=1543$, $p<0.001$, vowel context $F=498$ $p<0.001$ and volume level $F=79$ $p<0.001$ highly significantly. Interactions between Factors were highly significant as well, with the exception of interaction between volume and vowel context, speaker, volume, vowel context and consonant type, vowel context, volume as well as interaction between all four factors.

6.4.1.1 Effect of consonant type

The error bars in Figure 6-5 and 6-6 now combine results presented in the preceding sub-chapters. The data in Figure 6-5 somewhat blur the different standard deviations for different consonant types, since the data for individual speakers have not been normalized.

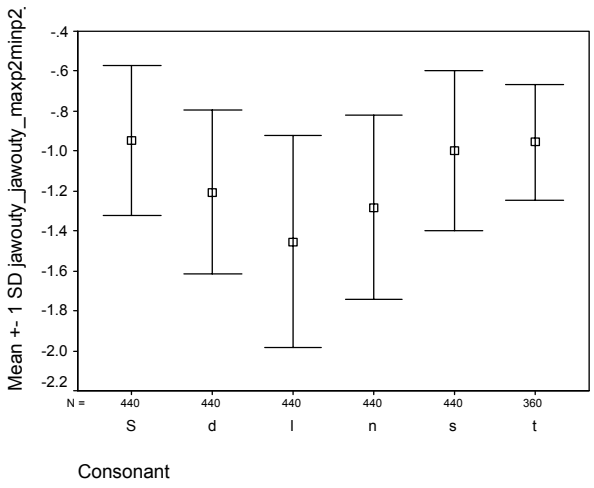


Figure 6-5. Over all speakers, volume levels and vowel contexts, mean jaw height at consonantal target position, and standard deviation in [cm]. $ /ʃ/, D /d/, L /l/, N /n/, S /s/, T /t/.

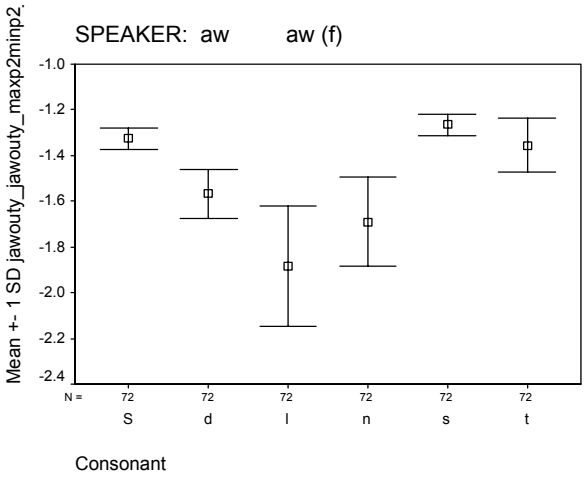


Figure 6-6a. Speaker aw over volume levels and vowel contexts, mean jaw height at consonantal target position, and standard deviation in [cm].

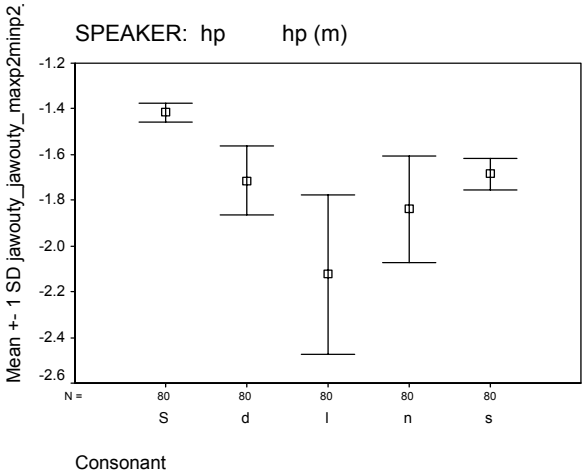


Figure 6-6b. Speaker hp over volume levels and vowel contexts, mean jaw height at consonantal target position, and standard deviation in [cm].

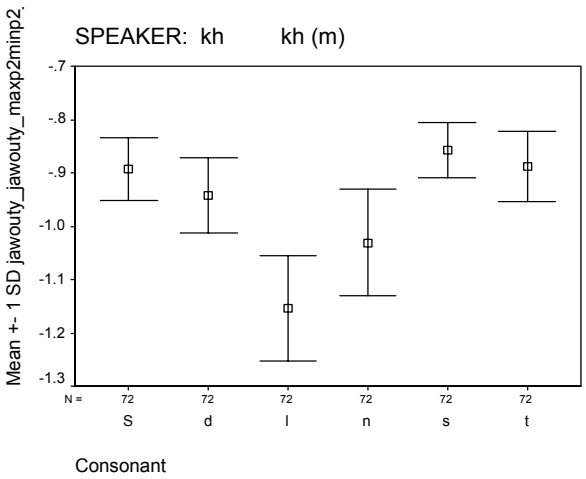


Figure 6-6c. Speaker kh over volume levels and vowel contexts, mean jaw height at consonantal target position, and standard deviation in [cm].

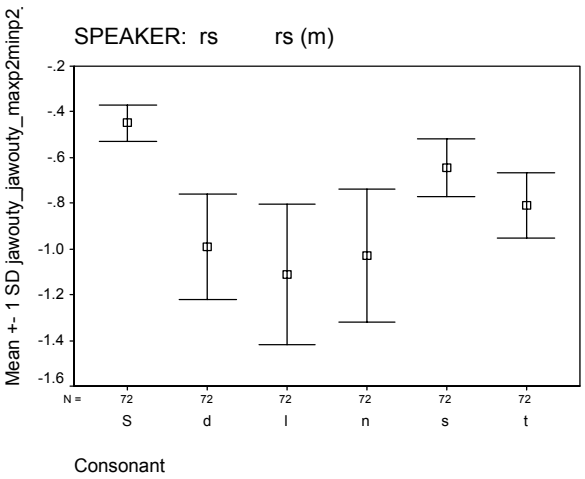


Figure 6-6d. Speaker rs over volume levels and vowel contexts, mean jaw height at consonantal target position, and standard deviation in [cm].

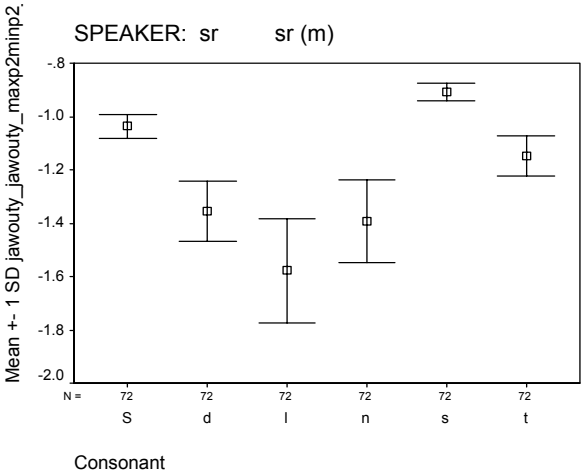


Figure 6-6e. Speaker sr over volume levels and vowel contexts, mean jaw height at consonantal target position, and standard deviation in [cm].

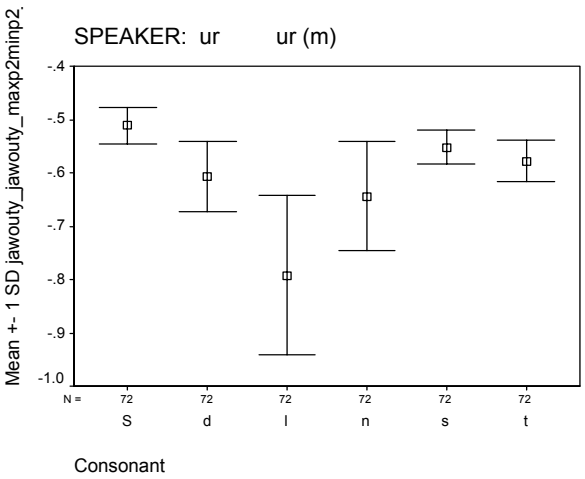


Figure 6-6f. Speaker ur over volume levels and vowel contexts, mean jaw height at consonantal target position, and standard deviation in [cm].

So a general tendency can best be derived from patterns in Figure 6-6 found for all (most) speakers. These findings are:

⇒ Jaw position is especially high for /s/, /ʃ/, /t/. A post-hoc (Bonferroni) comparison of variances of different consonants showed for most speakers no significant differences for these three.

⇒ Jaw position is especially low for /n/, /l/

⇒ Standard deviation is especially small for /s/, /ʃ/

⇒ Standard deviation is relatively small for /t/, /d/

⇒ Standard deviation is highest for /n/, /l/

Cluster analysis results are reported in the following.

Cluster analyses (as available in SPSS 10.0, using Euclidian distances between groups) were performed to test the different patterns of jaw variability in consonant production. The used method was a hierarchical cluster analysis over single speakers, for the variables standard deviation of jaw maximal peak position combined with standard deviation of intrinsic tongue tip position. Analyses over other variables such as VC amplitude of intrinsic tip and jaw were also used, The only reliable result across different variables on a two cluster level, was clustering of /n/ and /l/ as opposed to the other consonants. On a three cluster level for most speakers a pattern like /n/, /l/ vs. /d/, /t/, vs. /s/, /ʃ/ emerged.

These patterns of different variances are quite suggestive. On the other hand as has been noted in Keating et al. (1989), standard deviations may only be compared for groups with a similar mean. As we have shown for different consonants exactly two patterns, viz. differing jaw height (mean) and differing variability (standard deviation) have been found. A method to avoid pseudo effects of assumed higher variability, simply evoked by different means, is the empirical coefficient of variance (s * 100 / mean). Unfortunately this coefficient is only meaningful with positive x_i and not very informative for means close to zero (Clauß/Finze/Partzsch 1995). Our data fall exactly into this group, and a normalization procedure such as a z-transform would still show these counterindications. We can here only point at this problem. An alternative solution to treat jaw variability is to consider the closed jaw as rest position and to multiply jaw height by minus 1 and think of it as vertical distance from rest position.

Another interaction between jaw height and jaw variability might be brought up by the fact that high jaw position is at the limit of anatomically possible jaw height, viz. dental occlusion. So this, too sets a limit on variability. On the other hand, e.g. Edwards/Harris (1990) have shown, that jaw height range for speech purposes is much smaller than for mastication or than is anatomically possible. So, low jaw position during speech is not limited by an anatomical threshold. Thus higher jaw variability can be simply a natural epiphenomenon of more open jaw position.

To sum up: we will still assume that standard deviations can be compared, and that low standard deviations or variances account for a higher targetness of the respective jaw position for the production of the respective sound. On the other hand, we cannot ignore the possibility that especially physiological factors might have an important impact on differences in variability.

6.4.1.2 Influence of vowel context

The influence of vowel context on jaw height in consonants over all speakers was tested with two separate (unifactorial) Anovas. The influence of V1 and V2 is very similar but depends on type of consonant. Neither V1 nor V2 have a significant effect on jaw height in /ʃ/, /s/, /t/. The effect on /d/ is significant on a $p<0.01$ level (V1: F 6.543, $p=0.002$; V2: F 4.919, $p=0.008$). The effect on /n/ and /l/ is even significant on a $p<0.001$ level (/l/ F 26.083/26.160 $p<0.001$; /n/ F 16.651/13.851 $p<0.001$). As was said earlier the speakers show considerable differences, so a further analysis (multifactorial general linear model with vowel context and volume as fixed factors) for individual speakers and consonants showed a highly significant effect of vowel context for all six speakers' /l/ and /n/, for five speakers' /d/ and /t/, three speakers' /s/ and only one speaker's /ʃ/.

6.4.1.3 Influence of volume level

In Figure 6-7 mean values at the consonantal jaw target position are presented. The jaw position is slightly lower in loud condition, but only significant for two speakers (kh $p<0.05$; ur $p<0.001$). The missing significance in the other speakers might probably best be explained again with the highly significant effect of consonant type on jaw height (UniAnova F=103.3, $p<0.001$). The two-factorial (volume level and vowel context) analysis for individual speakers and consonants showed for no consonant a significant effect of volume level in all speakers but again a highly significant effect for all consonant types (/s/ significant with $p=0.018$) in speaker ur, a (highly) significant effect in /n/ and /l/ for speakers sr, rs, and kh. Individual significant results were found for different speakers and different consonants but they don't show a more generalizable pattern. Interaction between vowel context and volume level was never significant.

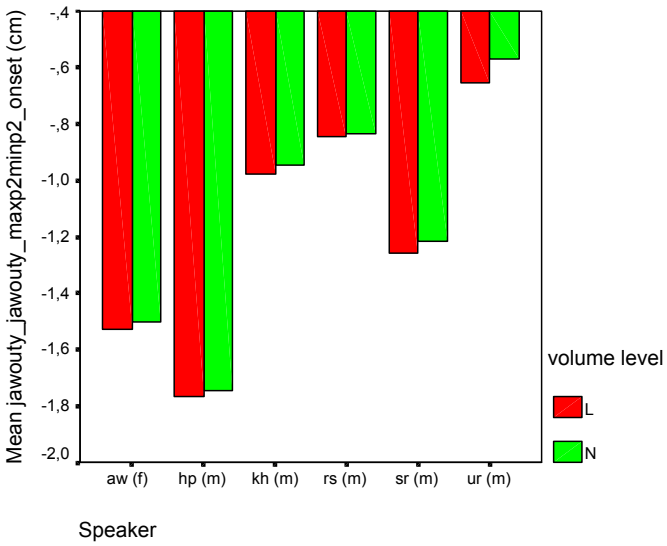


Figure 6-7. Jaw y maximal position (C) means for loud vs. normal volume.

6.4.2 Tongue articulation in consonants

A multifactorial analysis of variance (general linear model, four fixed factors) revealed that the tongue tip position at the consonantal target position is affected by speaker F=1169, $p<0.001$,

consonant type F=675, p<0.001, vowel context F=65 p<0.001 highly significantly, and by volume level F=5.168 p=0.023 at a 5 percent significance level. Interaction between factors were highly significant for speaker, consonant; speaker, vowel context; consonant type, vowel context; speaker, volume level; speaker, consonant type, vowel context. Interactions between consonant type and volume level as well as interaction between speaker, vowel context and volume level are significant at a 5 percent level.

As was described in 6.1 the actual tongue tip articulation corrected for the jaw articulation was calculated at the consonantal target position. The intrinsic tip height is affected by speaker F=2336, p<0.001, consonant type F=1961, p<0.001, vowel context F=519 p<0.001, and volume level F=145 p<0.001 highly significantly. Interactions between factors are highly significant beside combinations of all four factors and consonant type, vowel context and volume level. Interactions between vowel type and volume level and between speaker, consonant type and volume level are significant at a 5 percent level.

6.4.2.1 Effect of consonant type

Different tongue heights at consonantal target position refer to the fact that different consonant modes require different degree of closure. See Figure 6-8 for details. The highest tongue tip position can be found at /t/ target position. The lowest tongue tip height is found in /s/ target position. Higher positions are found for the other four consonants.

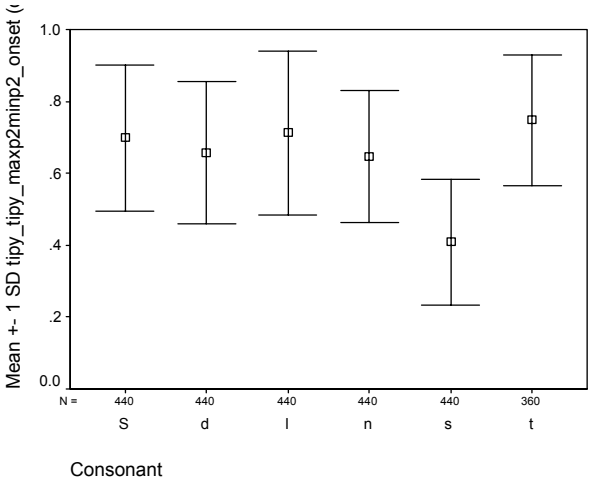


Figure 6-8. Tongue tip at consonantal target position. Means over all speakers for different consonants.

6.4.2.2 Influence of vowel context

A twofactorial univariate analysis of tongue tip height in consonantal target position (factors were volume and vowel context) done separately for individual speakers and consonantal targets showed for four speakers in all consonant targets a significant (in most cases even highly significant) difference, for the other two speakers (rs and kh) four of six vowel contexts were significicantly different (in 50 percent even highly significant). An overview of the vowel context influence on tongue tip height is presented in Figure 6-9 and shows a somewhat surprising pattern. The highest tongue tip position can be found in the context of /e:/. Lower tongue tip positions are found in the context of /a:/ as well as /i:/.

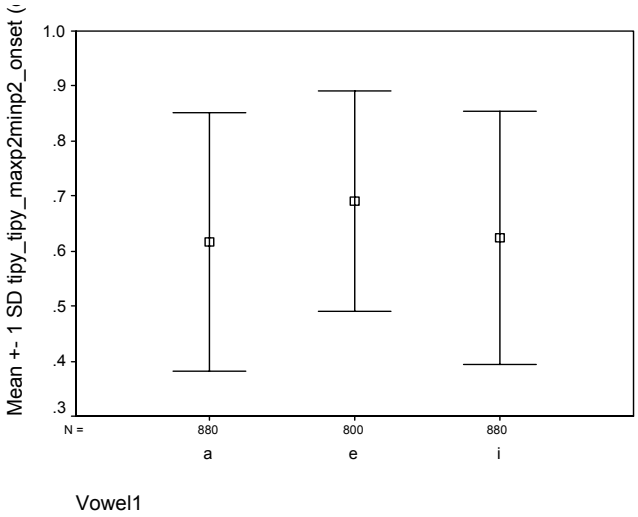


Figure 6-9. Tongue tip height at consonantal target position. Means over all speakers for different vowel contexts.

6.4.2.3 Influence of volume level

In Figure 6-10 mean y values at consonantal target position are presented. Intrinsic tip position is slightly higher in the loud condition, but only for three speakers at a significant level (aw, rs $p<0.05$; ur $p<0.001$). The jaw position is slightly lower in loud condition, but only significant for two speakers (kh $p<0.05$; ur $p<0.001$). A two factorial univariate analysis of intrinsic tongue tip height in consonantal target position (factors were volume and vowel context) done separately for individual speakers and consonantal targets showed the same tendencies with significant volume effects on all consonants for speaker aw, all consonants beside /s/ for speakers ur and rs, all consonants beside /s/ and /t/ for sr. For speaker hp /d/, /n/, and /s/ showed significant volume effects, for speaker kh only /l/ and /n/ had significant volume level effects.

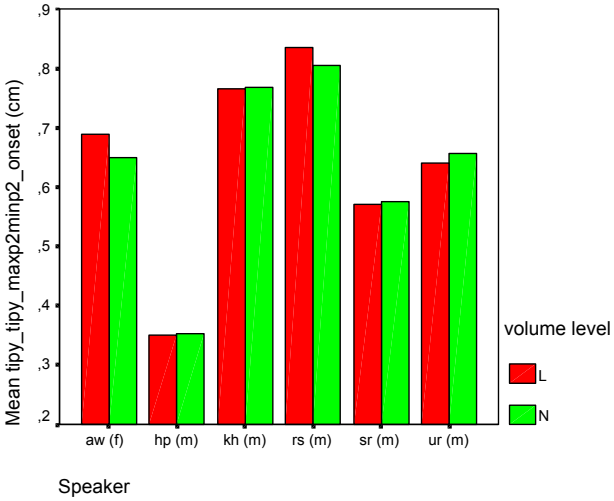


Figure 6-10a. Tongue tip at consonantal target position. Means for loud vs. normal volume.

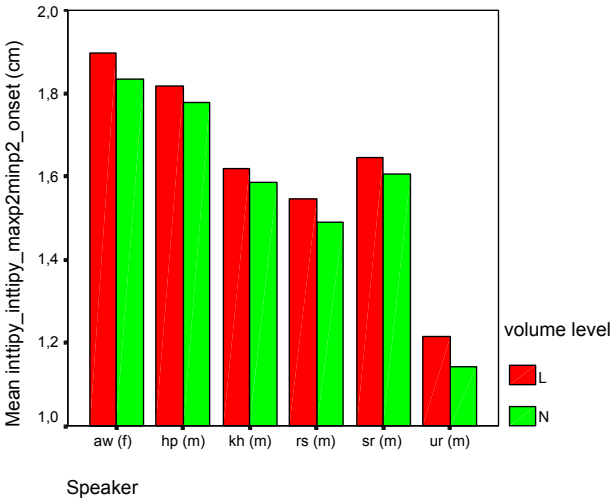


Figure 6-10b. Intrinsic tongue tip at consonantal target position. Means for loud vs. normal volume.

6.4.3 Tongue - jaw compensation in consonants

The general patterns for correlation between intrinsic tongue tip and jaw position are presented in Figure 6-11.

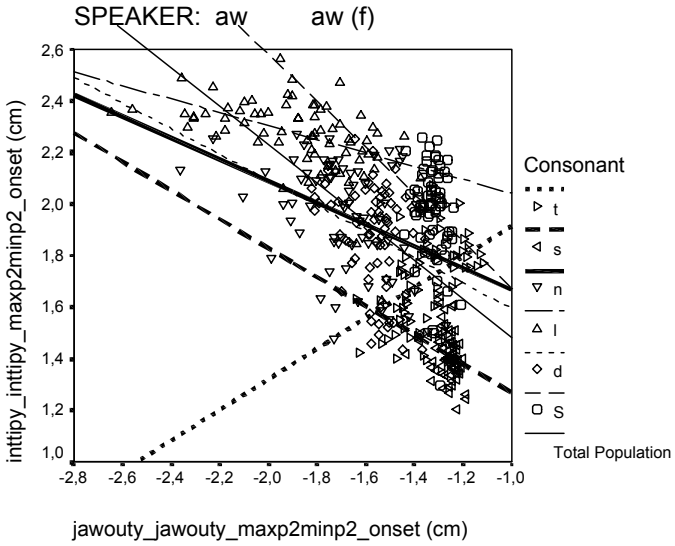


Figure 6-11a. Correlation plot and linear regression lines of intrinsic tip and jaw at consonantal target position. For speaker aw.

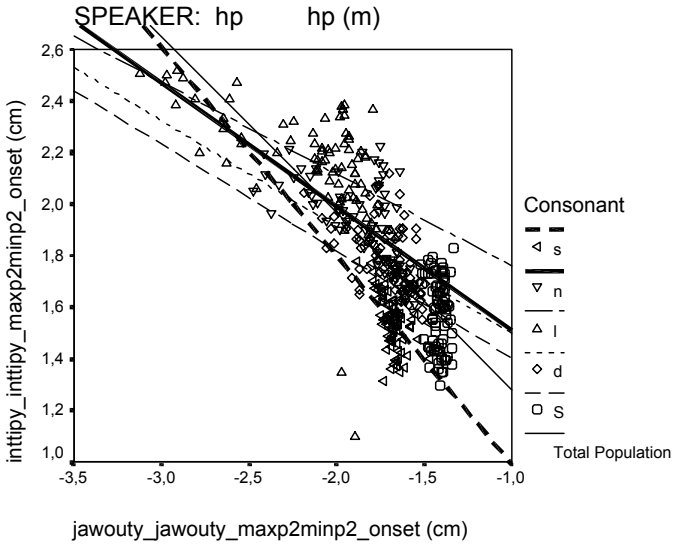


Figure 6-11b. Correlation plot and linear regression lines of intrinsic tip and jaw at consonantal target position. For speaker hp.

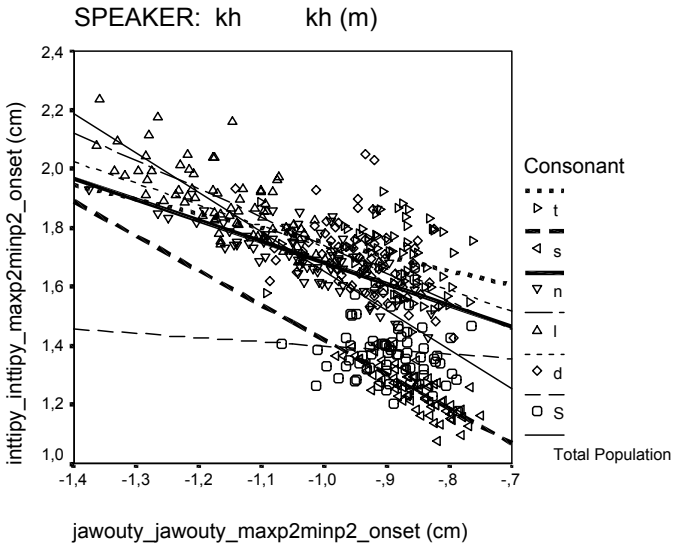


Figure 6-11c. Correlation plot and linear regression lines of intrinsic tip and jaw at consonantal target position. For speaker kh.

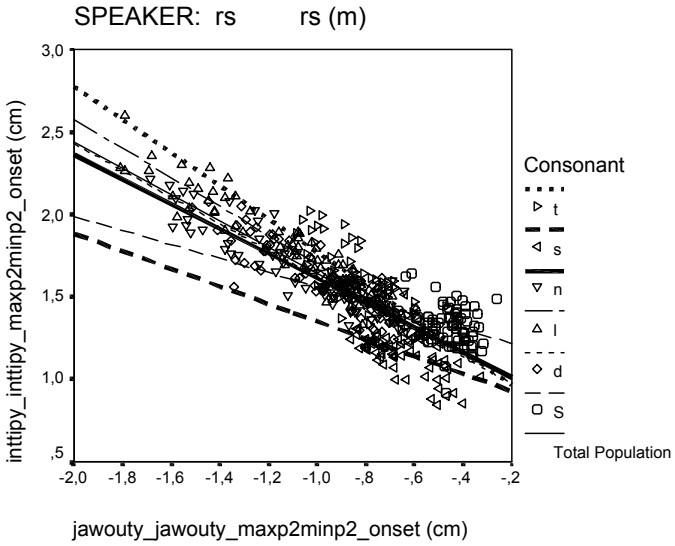


Figure 6-11d. Correlation plot and linear regression lines of intrinsic tip and jaw at consonantal target position. For speaker rs.

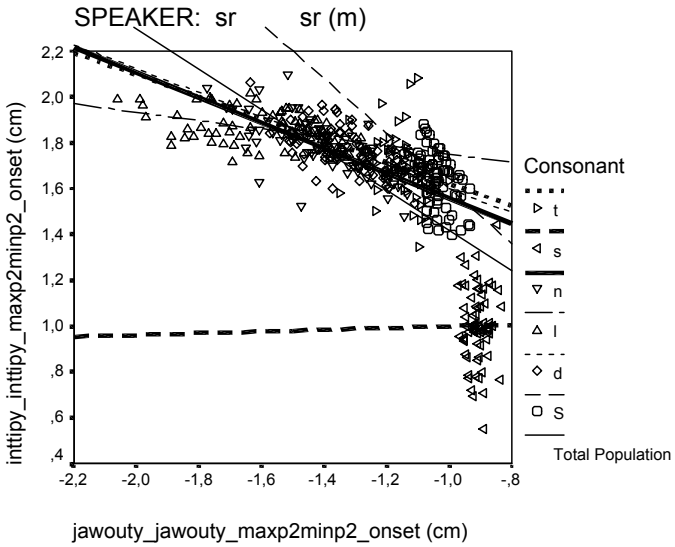


Figure 6-11e. Correlation plot and linear regression lines of intrinsic tip and jaw at consonantal target position. For speaker sr.

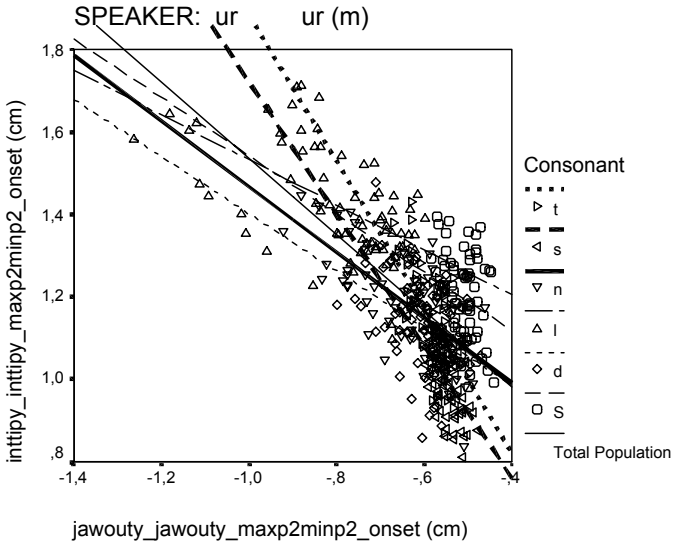


Figure 6-11f. Correlation plot and linear regression lines of intrinsic tip and jaw at consonantal target position. For speaker ur.

Figure 6-11 demonstrates the general tendency of compensatory behavior in the production of the apical consonant.

The correlations between jaw and intrinsic tip for consonantal peak position differ to some extent for individual speakers.

Speaker aw (Figure 6-11a): weak negative correlation for /n/, /l/, no correlation (Rsq < 0.1) for /ʃ/, /s/, /d/, even weak positive correlation for /t/.

Speaker hp (Figure 6-11b): fairly strong negative correlation for /n/, /l/, weak negative correlation for /s/, /d/, no correlation for /ʃ/ (/t/ is missing for this speaker).

Speaker kh (Figure 6-11c): strong negative correlation for /l/, /n/, /s/, weak negative correlation for /d/, no correlation for /ʃ/, /t/.

Speaker rs (Figure 6-11d): strong negative correlation for /d/, /t/, /l/, /n/, weak negative correlation for /s/, no correlation for /ʃ/.

Speaker sr (Figure 6-11e): stronger negative correlation for /d/, /n/, weak negative correlation for /ʃ/, /l/, no correlation for /s/, /t/.

Speaker ur (Figure 6-11f): stronger negative correlation for /l/, /n/, /s/, /t/, weak negative correlation for /d/, no correlation for /ʃ/.

The fact that there are no correlations for /ʃ/ can probably best be explained with the tongue tip position not representing the active articulator for the postalveolar fricative. This would explain why the tongue tip position does not have to compensate for jaw height differences.

Weak or missing correlations for /t/ might be due to the fact that /t/ has only small jaw variability see Figures 6-5, 6-6 and 6-8. Another explanation for a missing correlation fot /t/ and higher tongue tip variability is that /t/ could be produced with target overshoot, i.e. lower jaw height would still not cause loss of oral full closure, and has thus not to be compensated.

6.4.4 Durational aspects of consonant articulation

Articulatory durations for each articulator were determined via the distances between kinematic points (see Figure 6-12).

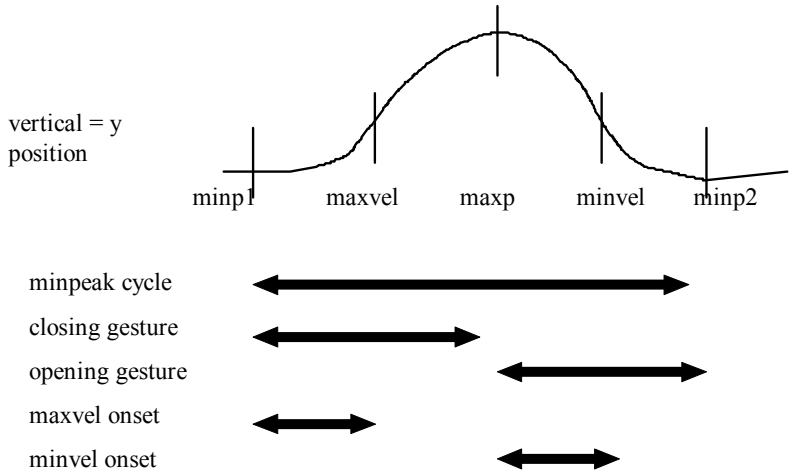


Figure 6-12. Kinematic durations analysed.

The consonantal articulation duration is represented in the upper figure with the closing gesture duration.

A multifactorial analysis of variance (general linear model, three fixed factors) in the jaw closing gesture duration was performed and showed highly significant effects for speaker, consonantal target and vowel context. The effect of volume level however was not significant. Interactions between speaker and consonantal target $F=1.843$ and ($p=0.008$) and speaker and volume level $F=3.339$, $p=0.005$ are significant at a $p<0.01$ level, volume*consonant interactions as well as interactions of all three factors are not significant.

In Figure 6-13a mean durations for the jaw closing gesture are shown, in Figure 6-13b the vowel context effect is shown. A multifactorial analysis (as above with V2 context as additional fixed factor) shows that the effect of the vowel context (similar results for preceding vowel context) is highly significant $p<0.001$, as are interactions between speaker and vowel context, vowel context and consonantal target, and the interaction between all three consonant, vowel context and speaker.

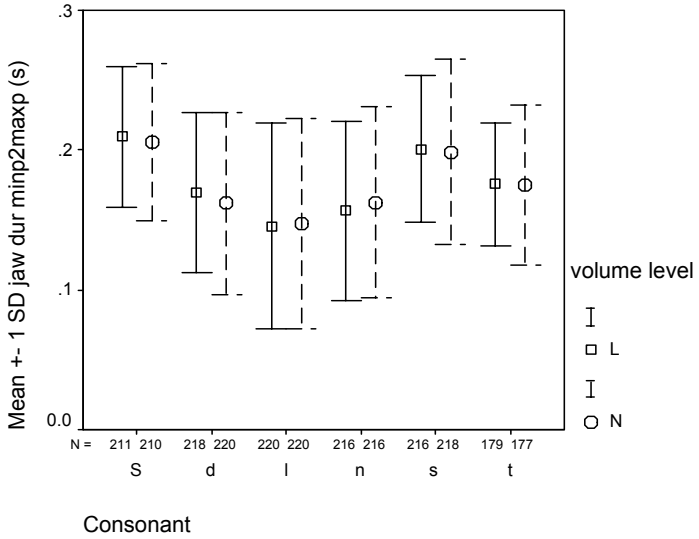


Figure 6-13a. Errorbars for duration of jaw closing gesture (C). Splitted by consonant qualities and volume levels.

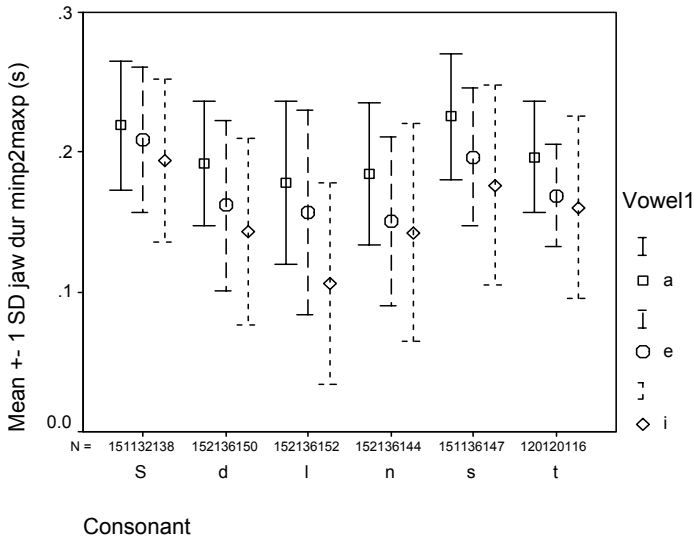


Figure 6-13b. Errorbars for duration of jaw closing gesture (C). Splitted by consonant qualities and preceding vowel qualities.

6.4.5 Articulatory amplitude vs. work for consonant production

As we have seen before heights of articulators vary between speakers significantly. One way to overcome effects simply resulting from individual articulatory strategies is looking rather at articulatory amplitudes than at exact height differences.

The difference between V1 minimal position and C maximal y position has been defined as amplitude. The „work“ has been defined earlier (chap. 6.3) as product of amplitude and minimum to maximum duration.

Figure 6-14 presents the different work patterns for individual speakers. Generally, the jaw work proportion is especially high for /s/ and /ʃ/. For /l/ and /n/ the intrinsic tip proportion is

especially high. The speakers sr and ur present extreme individual differences. While speaker sr shows up with a generally very high jaw work proportion, speaker ur has over all consonants a much smaller jaw work proportion.

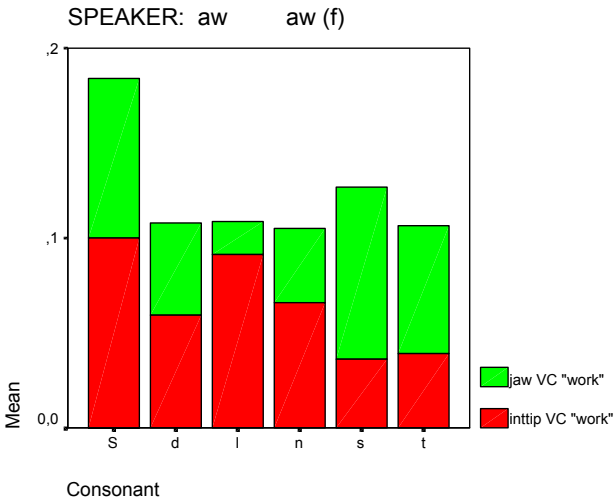


Figure 6-14a. Intrinsic tip - jaw VC „work“ for different consonants. Speaker aw.

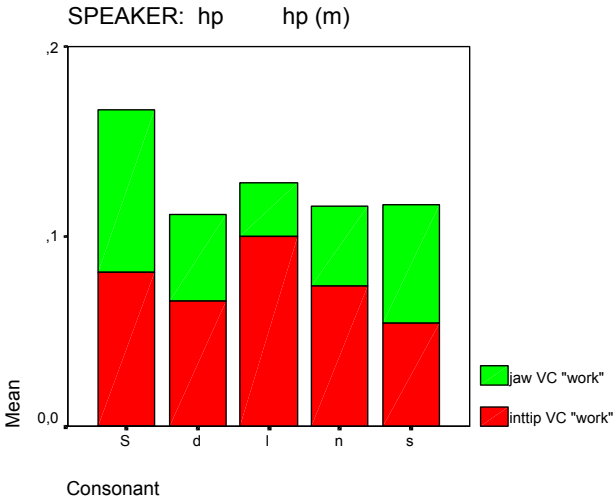


Figure 6-14b. Intrinsic tip - jaw VC „work“ for different consonants. Speaker hp.

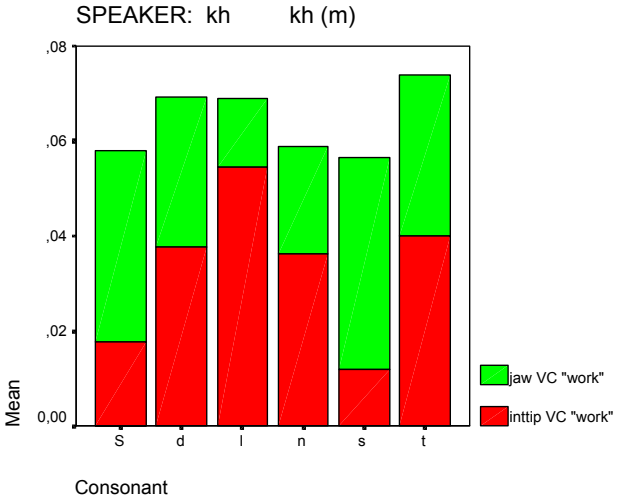


Figure 6-14c. Intrinsic tip - jaw VC „work“ for different consonants. Speaker kh.

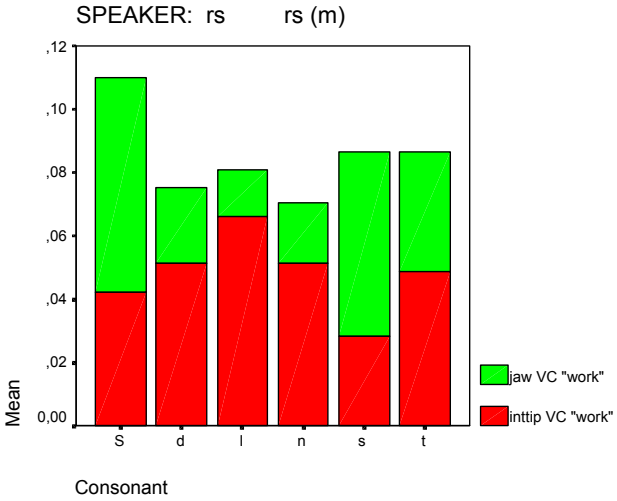


Figure 6-14d. Intrinsic tip - jaw VC „work“ for different consonants. Speaker rs.

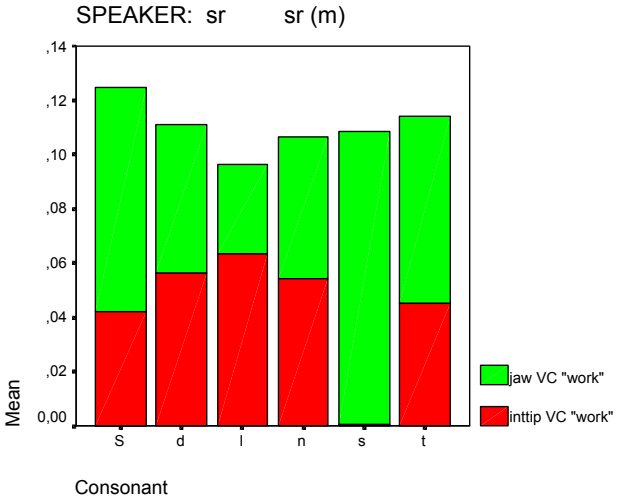


Figure 6-14e. Intrinsic tip - jaw VC „work“ for different consonants. Speaker sr. For /s/ even lowering movement of the tip was observed. This causes problems for the work algorithm. But the tendency of a very large proportion of jaw movement is still correct.

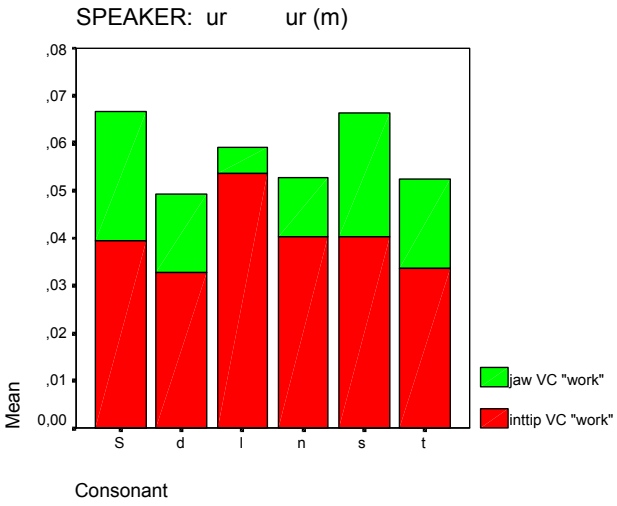


Figure 6-14f. Intrinsic tip - jaw VC „work“ for different consonants. Speaker ur.

6.4.5.1 Influence of volume level

The volume level effects on the intrinsic tip and jaw amplitude and work associated with the consonant gesture are described in Figure 6-15. A multifactorial analysis (general linear model, four fixed factors: consonant, V1, speaker, volume) shows that all factors have a highly significant effect on the work associated with the jaw VC gesture. Interactions between factors are highly significant as well, not significant is the interaction between all four factors and less significant are the interaction between speaker, consonant and volume (F=2.221 p=0.001) and volume, consonant and vowel (V1) (F=2.050, p=0.025)

The same multifactorial analysis with work associated with intrinsic tongue tip articulation as dependent variable shows similar results. All four factors are highly significant, interactions between speaker and volume, speaker and consonant, speaker and vowel and between speaker, consonant and vowel are highly significant.

A number of independent t-tests have been carried out in addition for all jaw, intrinsic tongue tip and tongue tip, to describe more in detail the effects for individual factors, additionally results for the jaw opening gesture are presented in Table 6-2. Results for the work associated with a gesture are compared with results for the amplitude of the gestures. Levels of significance are given in Table 6-2 broken down by speaker and 6-3 broken down by consonant type. The effect of volume level on intrinsic tip amplitude is significant only for two speakers (rs $p<0.05$; kh $p<0.01$). The effect on work associated with the intrinsic tip closing gesture is significant for three speakers (hp, kh $p<0.05$; rs $p<0.01$). The jaw closing gesture amplitude is for all speakers significantly higher under loud condition, jaw work is significantly higher in loud condition for five speakers, (speaker kh not significant).

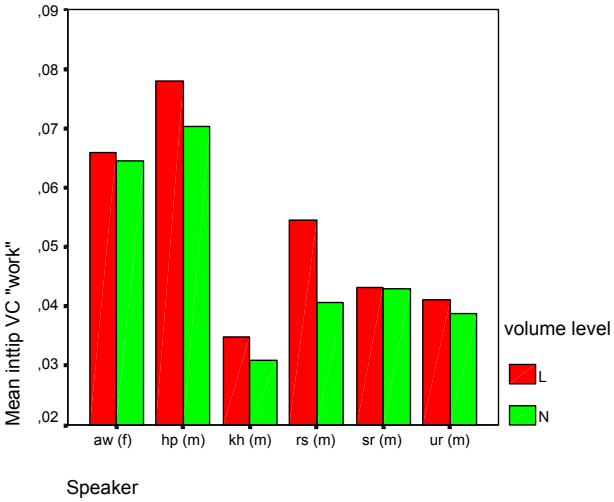


Figure 6-15a. Intrinsic tip VC „work“ for loud vs. normal volume.

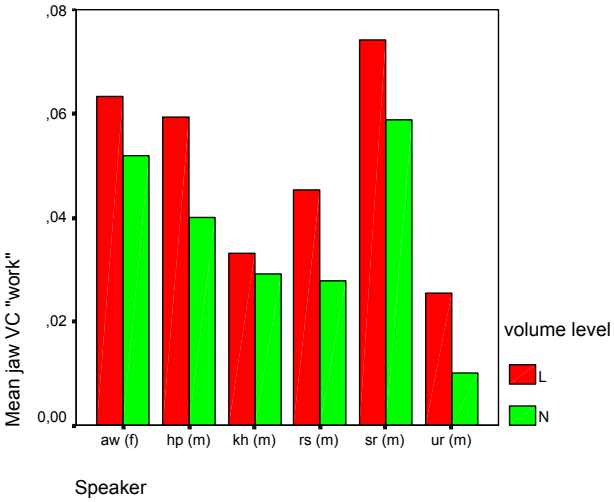


Figure 6-15b. Jaw VC „work“ for loud vs. normal volume.

n=max 216 two-tailed t-test for independent samples Loud vs. Normal Speech, t (p) level of significance	aw	hp	kh	rs	sr	ur
tip amplitude closing gesture	-3.121 p = .002	-2.696 p = .007	-3.373 p = .001	-3.660 p = .000	-1.752 p = .080	-3.100 p = .002
inttip amplitude closing gesture	-1.957 p = .051	-1.519 p = .130	-2.710 p = .007	-2.247 p = .025	-.129 p = .898	-1.619 p = .106
jaw amplitude closing gesture	-2.943 p = .003	-3.199 p = .002	-2.391 p = .017	-5.015 p = .000	-3.477 p = .001	-7.352 p = .000
tip amplitude opening gesture	-2.284 p = .023	-2.076 p = .039	-2.908 p = .004	-3.406 p = .001	-2.572 p = .010	-3.370 p = .001
inttip amplitude opening gesture	-1.368 p = .172	-1.669 p = .096	-1.598 p = .111	-1.965 p = .050	-1.193 p = .234	-2.426 p = .016
jaw amplitude opening gesture	-2.322 p = .021	-1.761 p = .079	-3.702 p = .000	-6.223 p = .000	-3.288 p = .001	-7.457 p = .000
„work“ associated with tip closing gesture	-1.909 p = .057	-3.062 p = .002	-2.663 p = .008	-4.591 p = .000	-1.356 p = .176	-2.564 p = .011
„work“ associated with inttip closing gesture	-.387 p = .699	-2.544 p = .011	-1.984 p = .048	-3.425 p = .001	-.025 p = .980	-.692 p = .489
„work“ associated with jaw closing gesture	-2.374 p = .018	-3.120 p = .002	-1.245 p = .214	-4.729 p = .000	-2.850 p = .005	-7.305 p = .000

Table 6-2. t-test for significance of the effect of volume level (L vs. N) on y amplitude and „work“ mean differences broken down by speaker. Variables were tested for normal distribution with the one-sample Kolmogorov-Smirnov test, and for homogeneity of variances with Levene’s test. The results for K-S and Levene’s test will not be reported. Where homogeneity of variances could not be assumed (Levene’s test $p< 0.05$) the t value for heterogeneous variances was taken.

n=max 220 two-tailed t-test for independent samples Loud vs. Normal Speech, t (p) level of significance	/ʃ/	/d/	/l/	/n/	/s/	/t/
inttip amplitude closing gesture	-1.220 p = .223	-1.762 p = .079	-1.924 p = .055	-2.058 p = .040	-.935 p = .350	-1.793 p = .074
jaw amplitude closing gesture	-3.736 p = .000	-3.475 p = .001	-2.429 p = .016	-3.054 p = .002	-4.099 p = .000	-4.104 p = .000
„work“ associated with inttip closing gesture	-.891 p = .373	-1.495 p = .136	-2.526 p = .012	-1.773 p = .077	-.220 p = .826	-1.105 p = .270
„work“ associated with jaw closing gesture	-3.766 p = .000	-3.236 p = .001	-2.058 p = .040	-2.457 p = .014	-3.386 p = .001	-3.815 p = .000

Table 6-3. t-test for significance of volume level (L vs. N) effect on y amplitude and „work“ mean differences over all speakers for different consonants. Variables were tested for normal distribution with the one-sample Kolmogorov-Smirnov test, and for homogeneity of variances with Levene’s test. The results for K-S and Levene’s tests will not be reported. Where homogeneity of variances could not be assumed (Levene’s test $p< 0.05$) the t value for heterogeneous variances was taken.

6.5 Results for vowel articulation

In the following subchapters the articulatory results for vowel production will be described.

6.5.1 Jaw articulation in vowels

Here results for jaw articulation in vowel production are described.

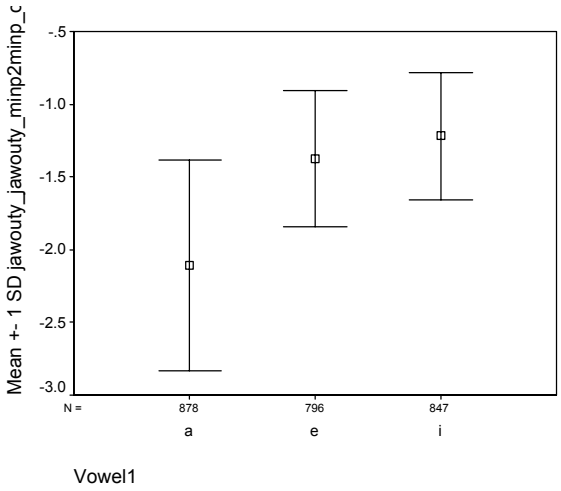


Figure 6-16. Over all speakers, mean jaw height at vowel (V1) target position, and standard deviation.

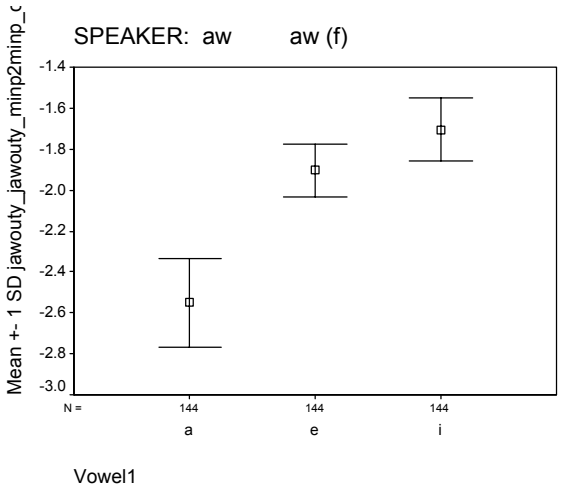


Figure 6-17a. Speaker aw, mean jaw height at vowel (V1) target position, and standard deviation in [cm].

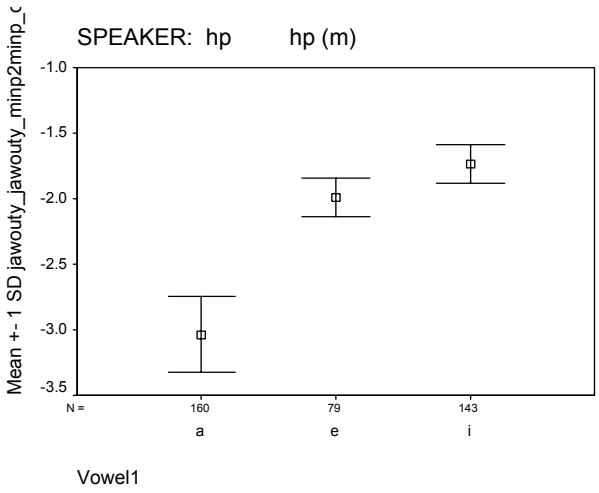


Figure 6-17b. Speaker hp, mean jaw height at vowel (V1) target position, and standard deviation in [cm].

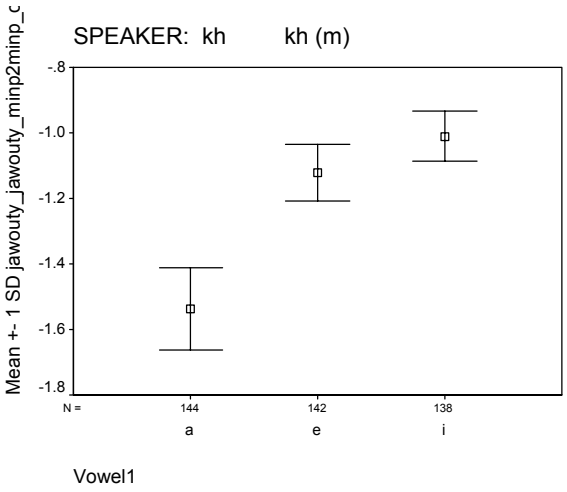


Figure 6-17c. Speaker kh, mean jaw height at vowel (V1) target position, and standard deviation in [cm].

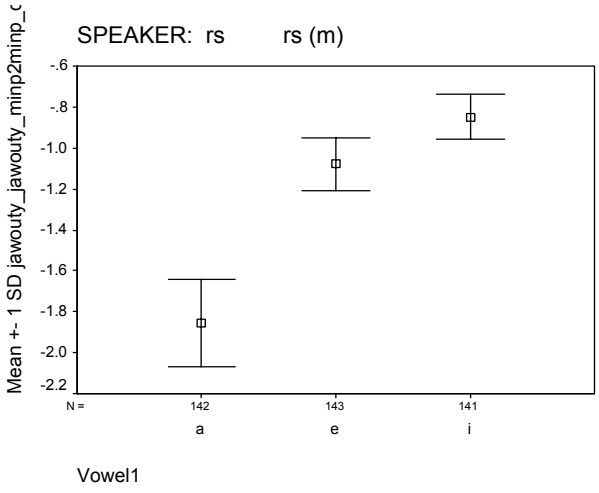


Figure 6-17d. Speaker rs, mean jaw height at vowel (V1) target position, and standard deviation in [cm].

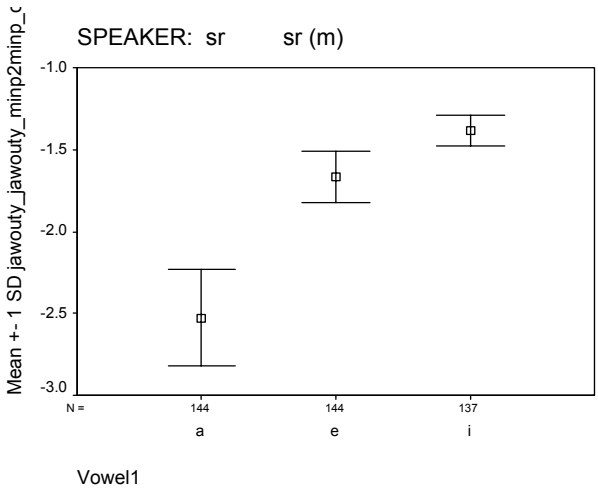


Figure 6-17e. Speaker sr, mean jaw height at vowel (V1) target position, and standard deviation in [cm].

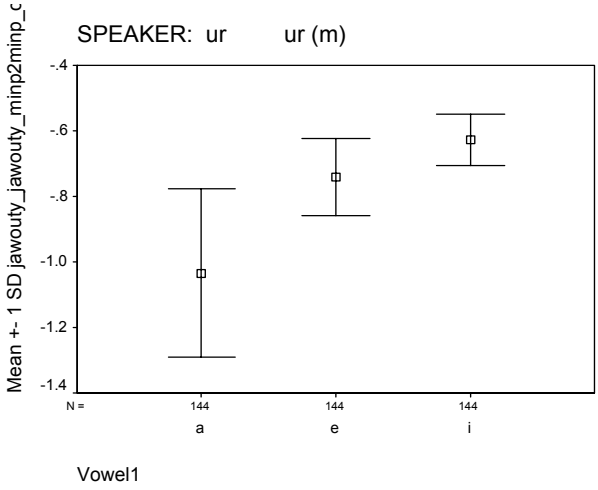


Figure 6-17f. Speaker ur, mean jaw height at vowel (V1) target position, and standard deviation in [cm].

Figure 6-16 and Figure 6-17 present jaw height differences at vowel (V1) target position. The general tendency can be descibed as the following:

⇒ Jaw height is by far lowest for /a:/, higher for /e:/ and highest for /i:/. This is in accordance with phonological vowel height definition.

⇒ Jaw variability is highest for /a:/, smaller for /e:/, very small for /i:/. The finding for /i:/ might result from physiological constraints as has been discussed in sub-chapter 6.4.3.1.

The effect of intervocalic consonant on V1 and V2 jaw position was tested with two unifactorial Anovas. In general the effect of anticipatory coarticulation (effect on V1) was weaker than that of perseverative coarticulation (effect on V2). The effect on V1 jaw height was significant for /i:/ (F 4.382 p= 0.001) and /a:/ (F 2.865 p=0.014) but not for /e:/ (F2.181 p=0.054). V2 jaw height was affected significantly by consonant context for all three vowels. /i:/ (F 4.589, $p<0.001$); /e:/ (F 3.296 p=0.006); /a:/ (F 12.810 $p<0.001$).

6.5.2 Tongue articulation in vowels

In this chapter tongue articulation here represented by tongue tip articulation is described. The arguments for choosing this sensor have been given above. As we can see from Figures 6-18 and 6-19 the height patterns for tongue tip and intrinsic tongue tip for V1 look very much alike. They are consistent for all speakers as can be seen in Figure 6-20. Data for V2 are not presented here but are very similar to those for V1.

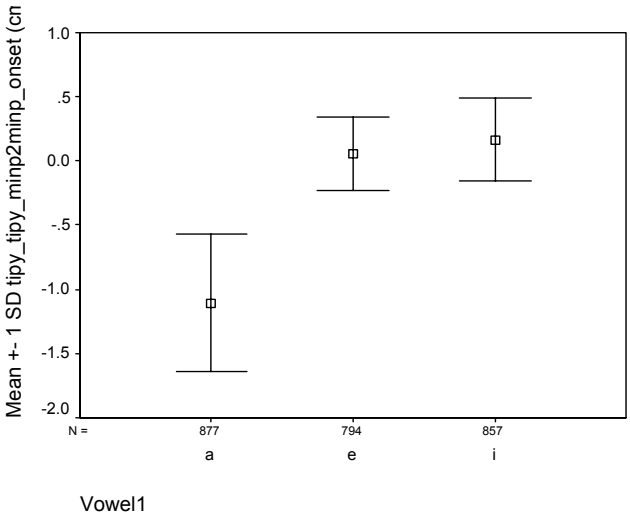


Figure 6-18. Over all speakers, mean tongue tip height at vowel (V1) target position, and standard deviation in [cm].

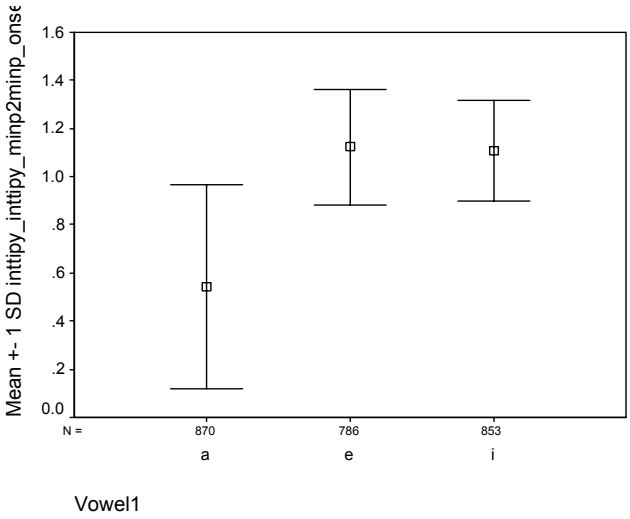


Figure 6-19. Over all speakers, mean intrinsic tongue tip height at vowel (V1) target position, and standard deviation in [cm].

A two factorial analysis of intrinsic tongue tip height with volume and consonant context as fixed factors showed for V2 for speakers aw, hp, sr a highly significant effect ($p<0.001$), for kh and rs at $p<0.01$ level, for preceding consonant context. Volume and volume consonant context interaction are not significant at all.

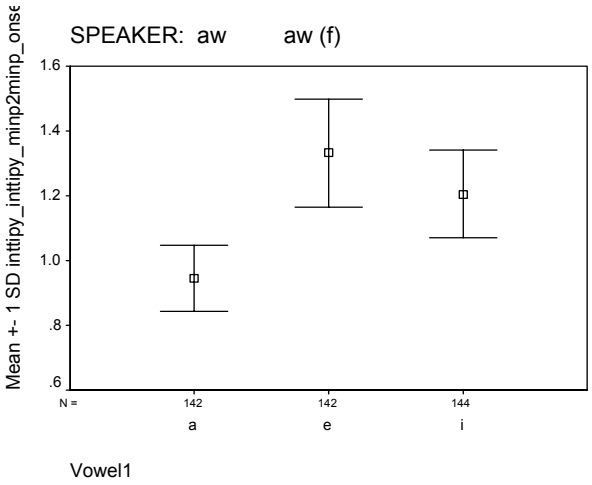


Figure 6-20a. Speaker aw, mean intrinsic tongue tip height at vowel (V1) target position, and standard deviation in [cm].

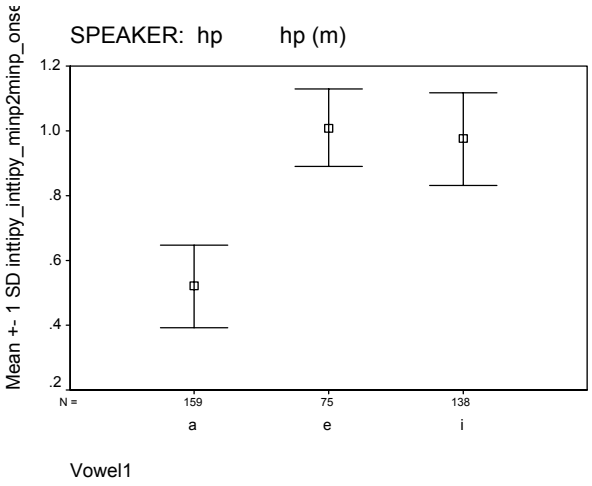


Figure 6-20b. Speaker hp, mean intrinsic tongue tip height at vowel (V1) target position, and standard deviation in [cm].

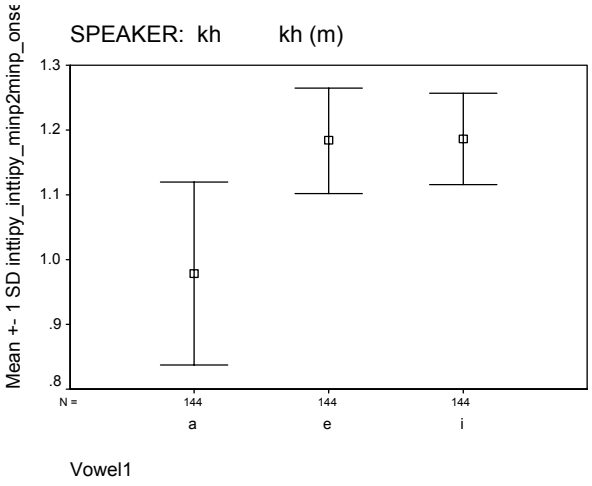


Figure 6-20c. Speaker kh, mean intrinsic tongue tip height at vowel (V1) target position, and standard deviation in [cm].

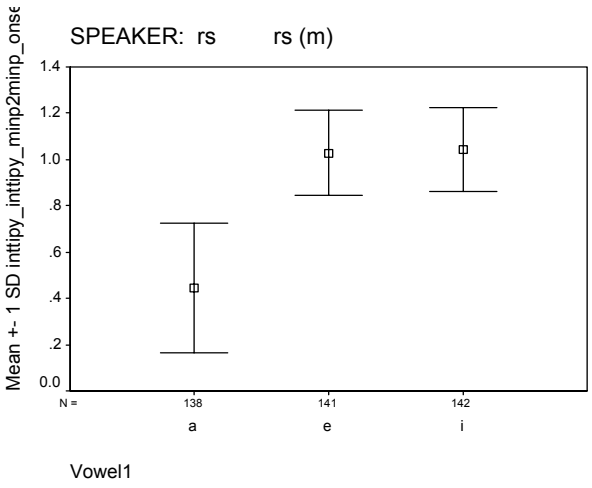


Figure 6-20d. Speaker rs, mean intrinsic tongue tip height at vowel (V1) target position, and standard deviation in [cm].

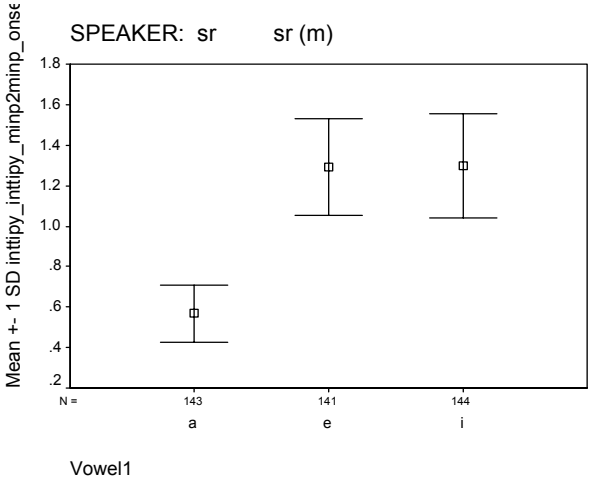


Figure 6-20e. Speaker sr, mean intrinsic tongue tip height at vowel (V1) target position, and standard deviation in [cm].

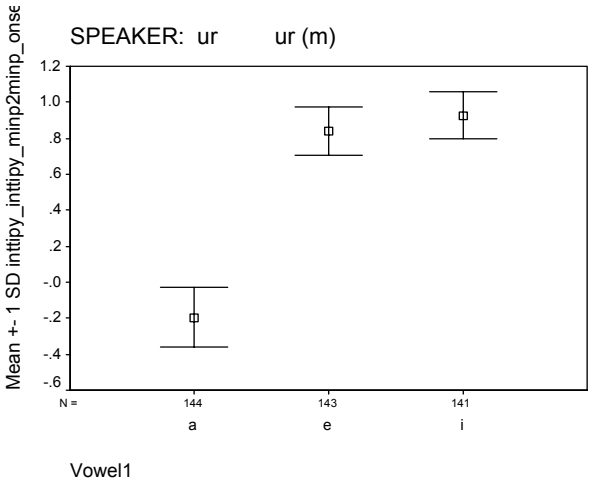


Figure 6-20f. Speaker ur, mean intrinsic tongue tip height at vowel (V1) target position, and standard deviation in [cm].

6.5.3 Tongue - jaw compensation in vowels

Compensation for jaw height with intrinsic tongue tip height is discussed in the following.

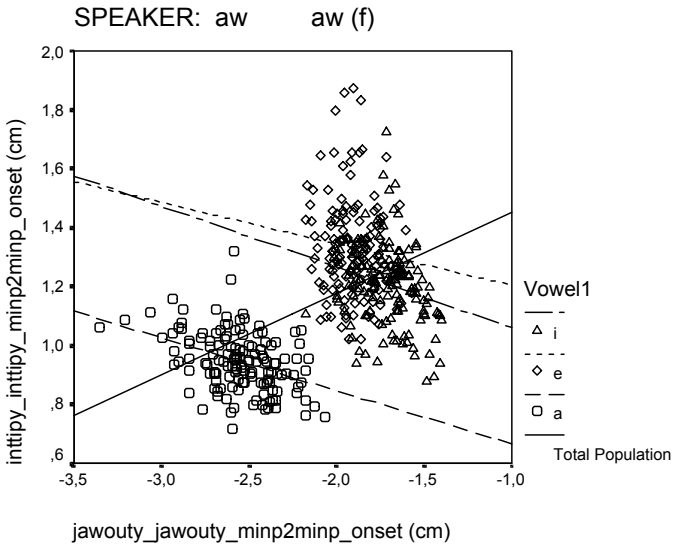


Figure 6-21a. Correlation plot and linear regression lines of intrinsic tip and jaw at vowel target position (V1). For speaker aw.

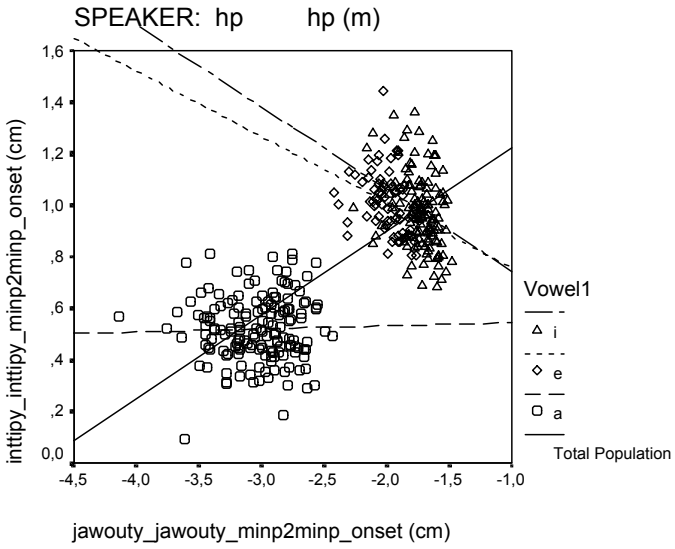


Figure 6-21b. Correlation plot and linear regression lines of intrinsic tip and jaw at vowel target position (V1). For speaker hp.

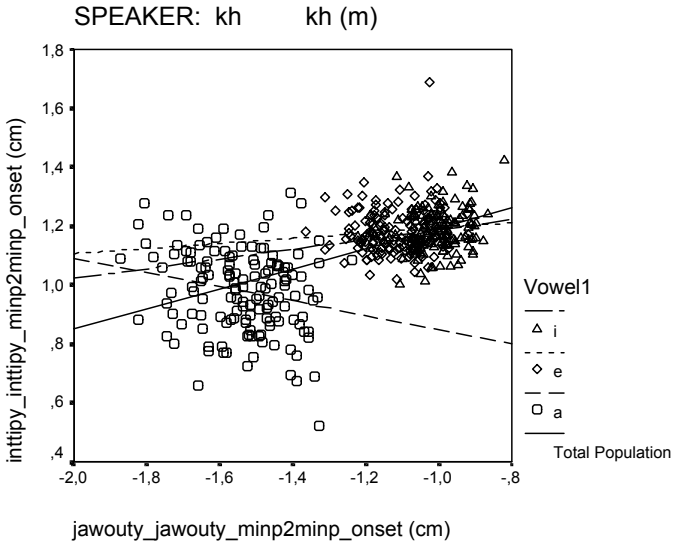


Figure 6-21c Correlation plot and linear regression lines of intrinsic tip and jaw at vowel target position (V1). For speaker kh.

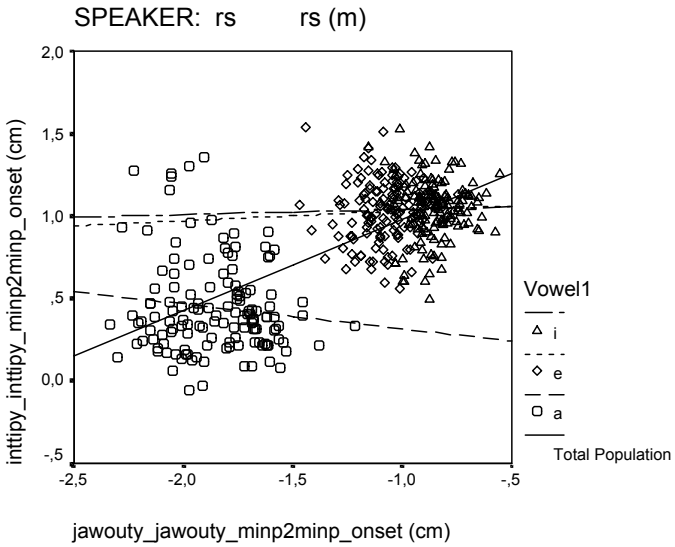


Figure 6-21d Correlation plot and linear regression lines of intrinsic tip and jaw at vowel target position (V1). For speaker rs.

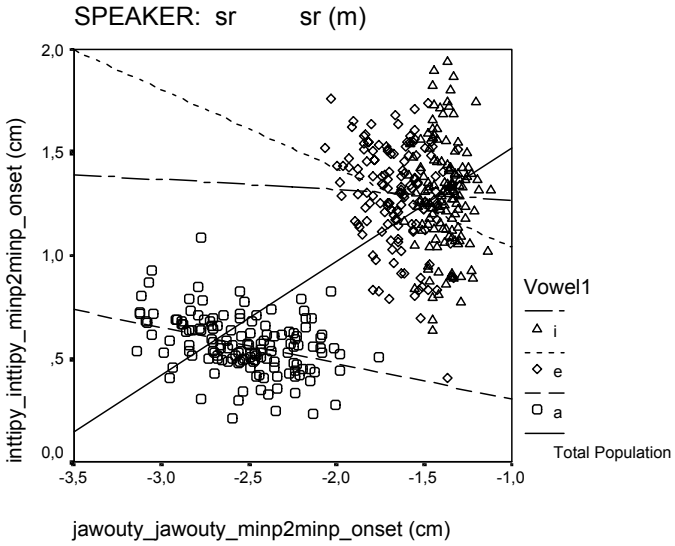


Figure 6-21e. Correlation plot and linear regression lines of intrinsic tip and jaw at vowel target position (V1). For speaker sr.

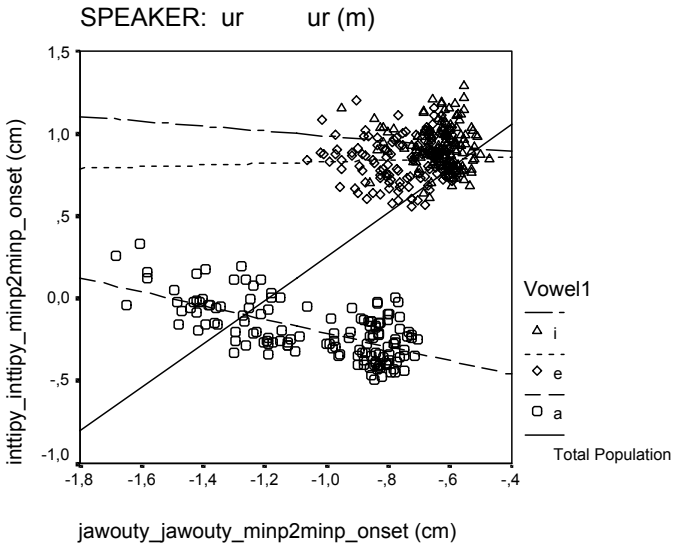


Figure 6-21f. Correlation plot and linear regression lines of intrinsic tip and jaw at vowel target position (V1). For speaker aw.

Figure 6-21 shows a pattern of two different clusters that represent the vowel qualities /i:/, /e:/ more or less bundled together and /a:/. The two higher vowel qualities /e:/ (left) and /i:/ (right) form the upper cluster and /a:/ the lower. The patterns for different speakers and vowels show a different behavior for most speakers with a weak negative correlation in /a:/ production but no correlation for /i:/ and /e:/. Some illustrations are given in Figure 6-22.

The difference in the correlations beween /a:/ and /e:/, /i:/ could be explained by assuming that precise tip positioning is more relevant for /a:/ production, which, on first glance, does not seem to be very obvious. A potential explanation, considering the different muscles involved for /i:/, /e:/ vs. /a:/ production, will be presented in the discussion part in chapter 8.

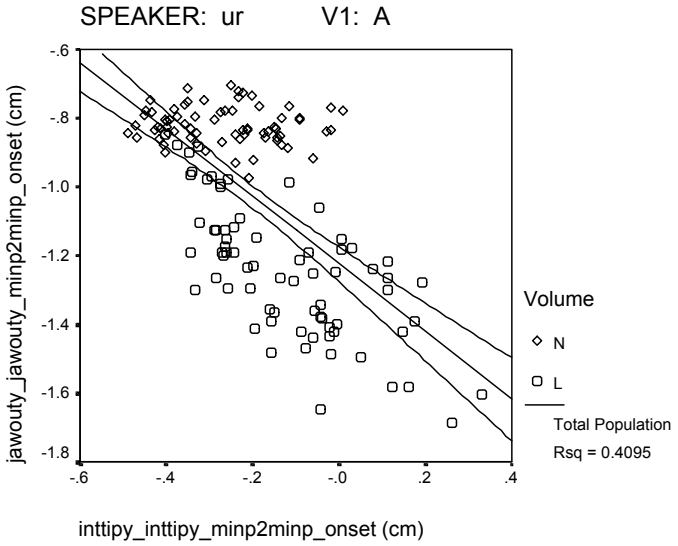


Figure 6-22a. Correlation of y position of jaw minimal peak with intrinsic tip minimal peak for vowel /a/ for speaker ur.

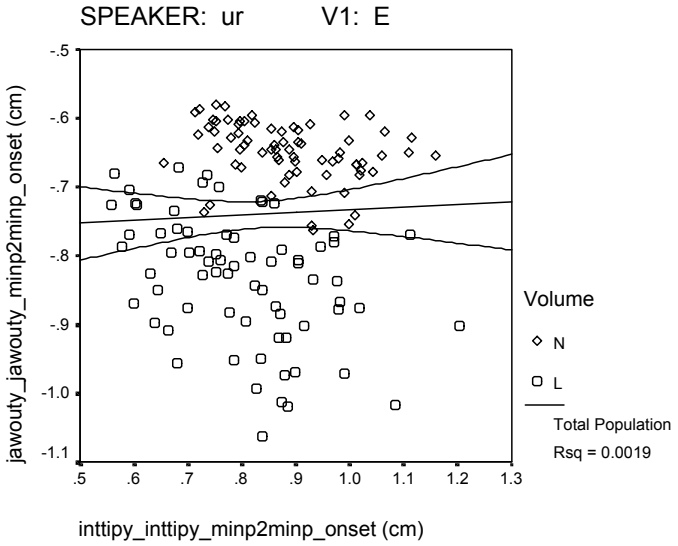


Figure 6-22b. Correlation of y position of jaw minimal peak with intrinsic tip minimal peak for vowel /e/ for speaker ur.

In Figure 6-23 intrinsic tip and jaw vertical position values at the vowel (V1) target position are presented. A multifactorial analysis (general linear model, four fixed factors: vowel, consonant context, volume, speaker) for the jaw position showed highly significant effects for vowel, consonant context and speaker and a lesser but still highly significant effect with F=7.859, p=0.005 for volume. For intrinsic tongue tip the same test showed highly significant effects for vowel target, consonant context speaker and volume. In the following and Table 6-4 further significance test results are presented.

Jaw position (Figure 6-23b) at V1 target position is for all speakers highly significantly ($p<0.001$) lowered in loud condition.

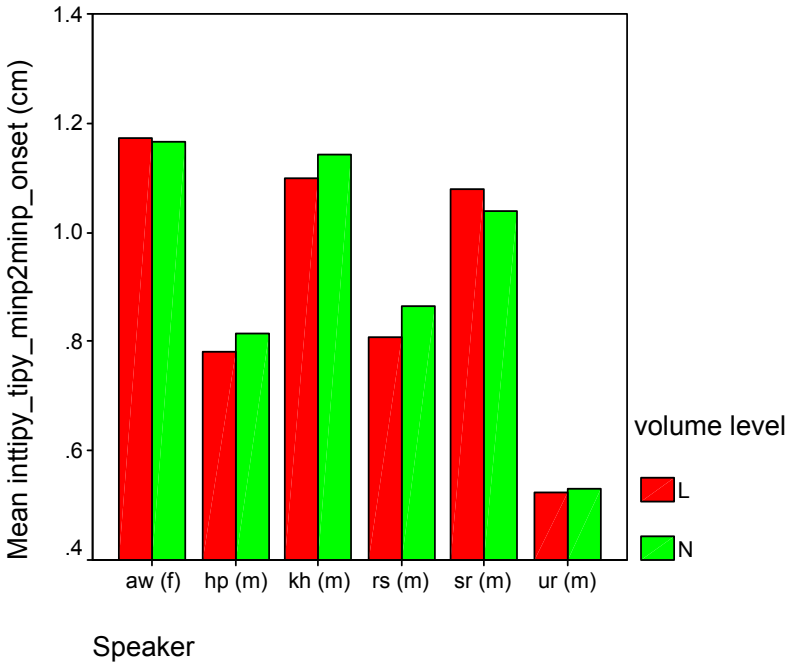


Figure 6-23a. Intrinsic tip y minimal position (V1) means for loud vs. normal volume.

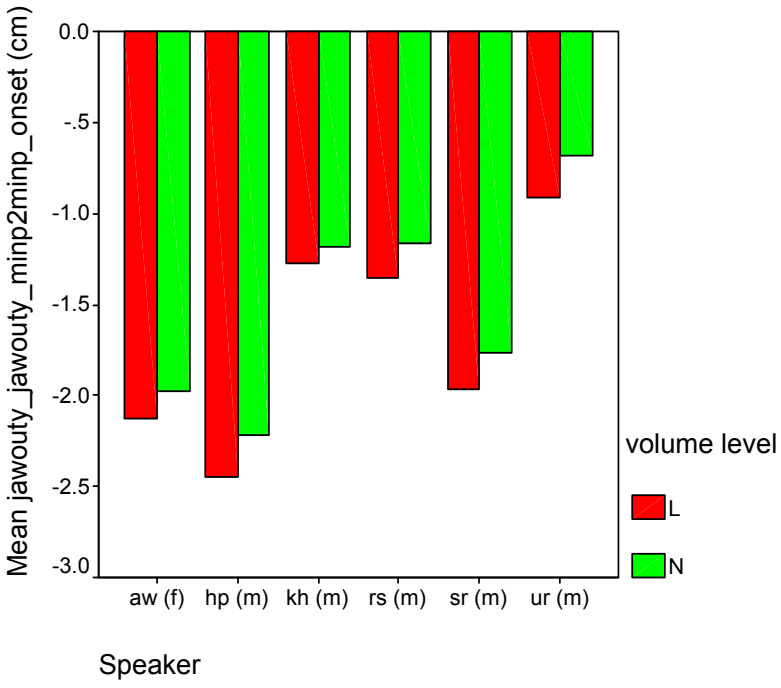


Figure 6-23b. Jaw y minimal position (V1) means for loud vs. normal volume.

The significances of the effect of volume level on y positions of single articulators are presented in Table 6-4.

n=max 216 two-tailed t-test for independent samples Loud vs. Normal Speech, t (p) level of significance	aw	hp	kh	rs	sr	ur
tip minp (V1)	2.541 p = .011	2.696 p = .007	4.333 p = .000	3.173 p = .002	1.702 p = .089	3.074 p = .002
tip maxp (C)	-1.804 p = .072	.112 p = .911	.093 p = .926	-2.187 p = .029	.294 p = .769	1.260 p = .208
tip minp (V2)	1.180 p = .239	2.091 p = .037	3.548 p = .000	2.937 p = .003	2.395 p = .017	3.325 p = .001
inttip minp (V1)	.031 p = .975	1.179 p = .239	3.457 p = .001	1.408 p = .160	-.797 p = .426	.145 p = .885
inttip maxp (C)	-2.133 p = .033	-1.409 p = .159	-1.323 p = .186	-2.108 p = .036	-1.286 p = .199	-4.592 p = .000
inttip minp (V2)	-.729 p = .467	1.097 p = .273	1.001 p = .317	1.310 p = .191	.232 p = .817	.762 p = .447
jaw minp (V1)	3.921 p = .000	3.593 p = .000	3.782 p = .000	4.233 p = .000	4.136 p = .000	11.260 p = .000
jaw maxp (C)	1.092 p = .275	.792 p = .429	2.675 p = .008	.294 p = .769	1.696 p = .091	7.738 p = .000
jaw minp (V2)	3.795 p = .000	2.294 p = .022	5.128 p = .000	5.015 p = .000	5.507 p = .000	12.910 p = .000

Table 6-4. t-test for significance of the effect of volume level (L vs. N) on y position mean differences in vowels and consonants broken down by speaker. Variables were tested for normal distribution with the one-sample Kolmogorov-Smirnov test, and for homogeneity of variances with Levene's test. The results for K-S and Levene's tests will not be reported. Where homogeneity of variances could not be assumed (Levene's test $p < 0.05$) the t value for heterogeneous variances was taken.

6.5.4 Durational aspects of vowel articulation

Articulatory durations for each articulator were determined via the distances between kinematic points (see Figure 6-24).

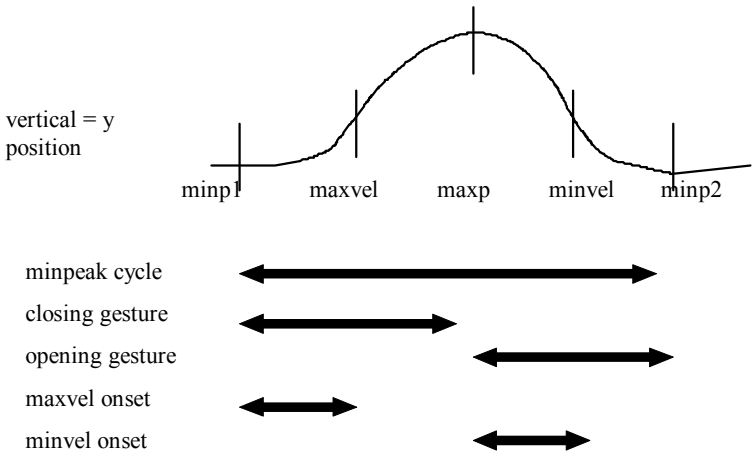


Figure 6-24. Kinematic durations analysed.

The vowel (V2) articulation duration is represented in the Figure 6-24 with the opening gesture duration.

A multifactorial analysis of variance in the jaw opening gesture duration was performed and showed highly significant effects for speaker, consonantal target and vowel context. The effect of volume level was significant p=0.016 at a five percent level. Figures 6-25a and 6-25b present errorbars for the different vowel targets split up by volume level and preceding consonant context.

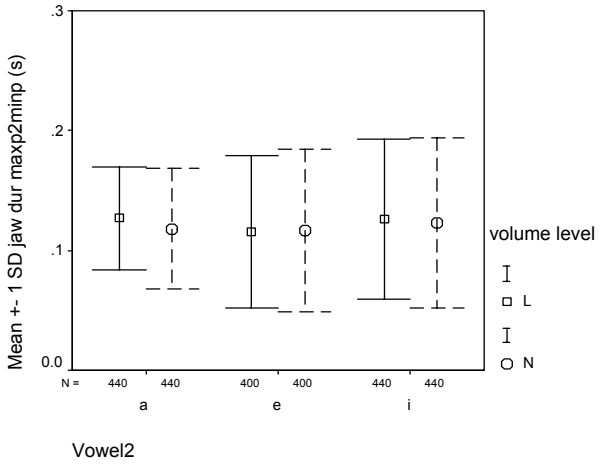


Figure 6-25a. Errorbars for duration of jaw opening gesture (V2). Splitted by vowel qualities and volume levels.

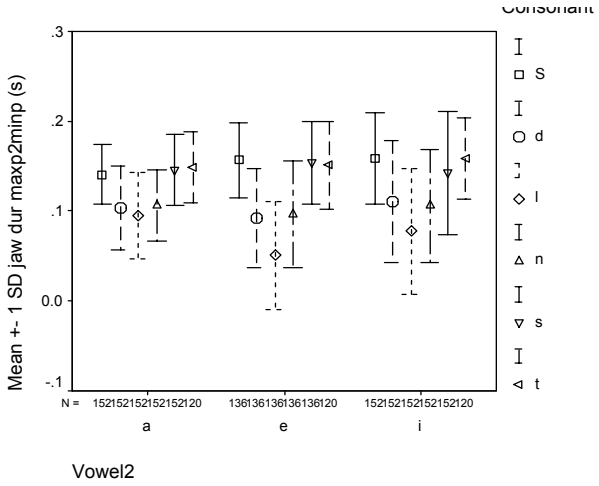


Figure 6-25b. Errorbars for duration of jaw opening gesture (V2). Splitted by vowel qualities and preceding consonants.

6.6 Interspeaker differences in articulatory activity

It was already noted that speakers differ considerably in their use of jaw and tongue as articulators. In Figure 6-15b the sizes of jaw work of the consonantal closing gesture averaged over all vowel contexts are given. They show relatively small amplitudes for speakers ur and to a certain degree kh and rs. Speaker kh is on the other hand showing only small intrinsic tip amplitudes, to make the contribution of individual articulators moreeasily comparable they are displayed conjointly in Figure 6-23 for vowel context /a:/.

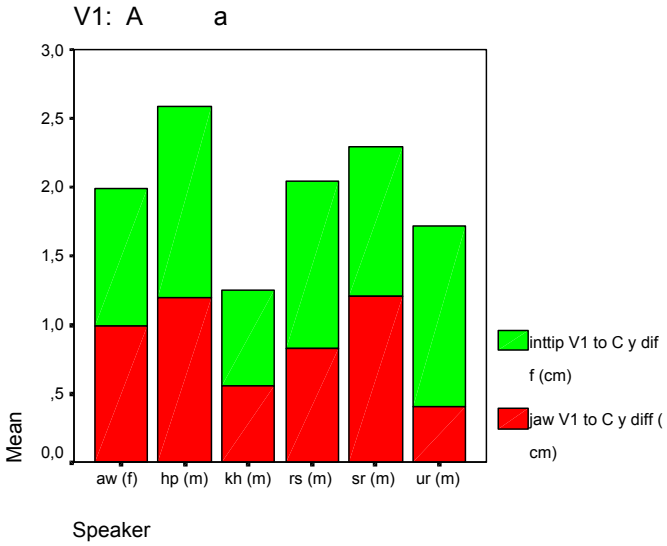


Figure 6-26. Intrinsic tip and jaw amplitudes of the VC closing gesture over all consonants in the /a:/ context.

Figure 6-26 shows clearly that speaker kh has an overall relatively small extension of his articulatory movement. Speaker ur has a relatively small jaw amplitude contribution to his

closing gesture articulation, while speaker sr shows an above-average proportion of the jaw amplitude. In Figure 6-24 the jaw variability in the vowel target position as well as the consonant target position is shown. Again there are noticable differences between speakers.

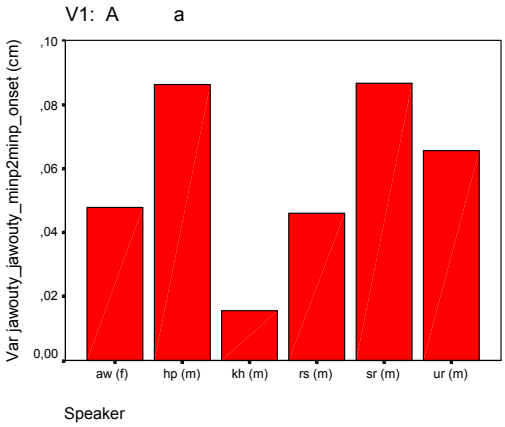


Figure 6-27a.Vertical jaw variability at vowel (V1) /a:/ target position. Splitted for individual speakers.

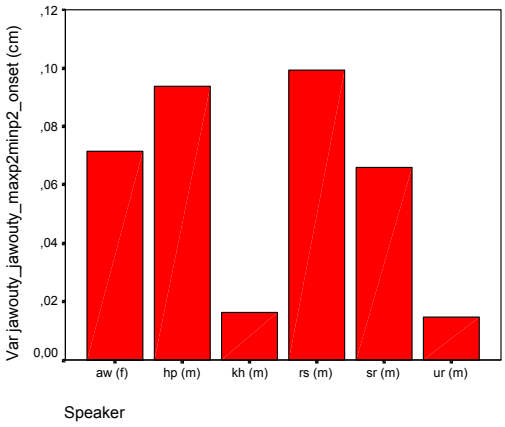


Figure 6-27b.Vertical jaw variability at consonant target position. Averaged over all consonant qualities, splitted for individual speakers.

6.6.1 Palate contours of speakers

It has been argued (see for a discussion of this Johnson/Ladefoged/Lindau (1993) and Lee (1994)) that the shape of the palate can affect the amount of jaw variability. The tendency then is described in the following way: a highly domed palate especially in the dorsal region can account for lower jaw variability in vowels (Ladefoged et al. 1972). Higher jaw variability at vowel midpoint was discussed to be associated with shallowly domed palate. In Figure 6-25 the midsagittal palate contours (ascertained by speaker's dental cast) are documented. A classification of the individual palate shapes is presented in Table 6-5.

It has been mentioned before that for consonant production general jaw activity is especially large for speaker sr and especially small for speaker ur. Based on Ladefoged et al. (1971).

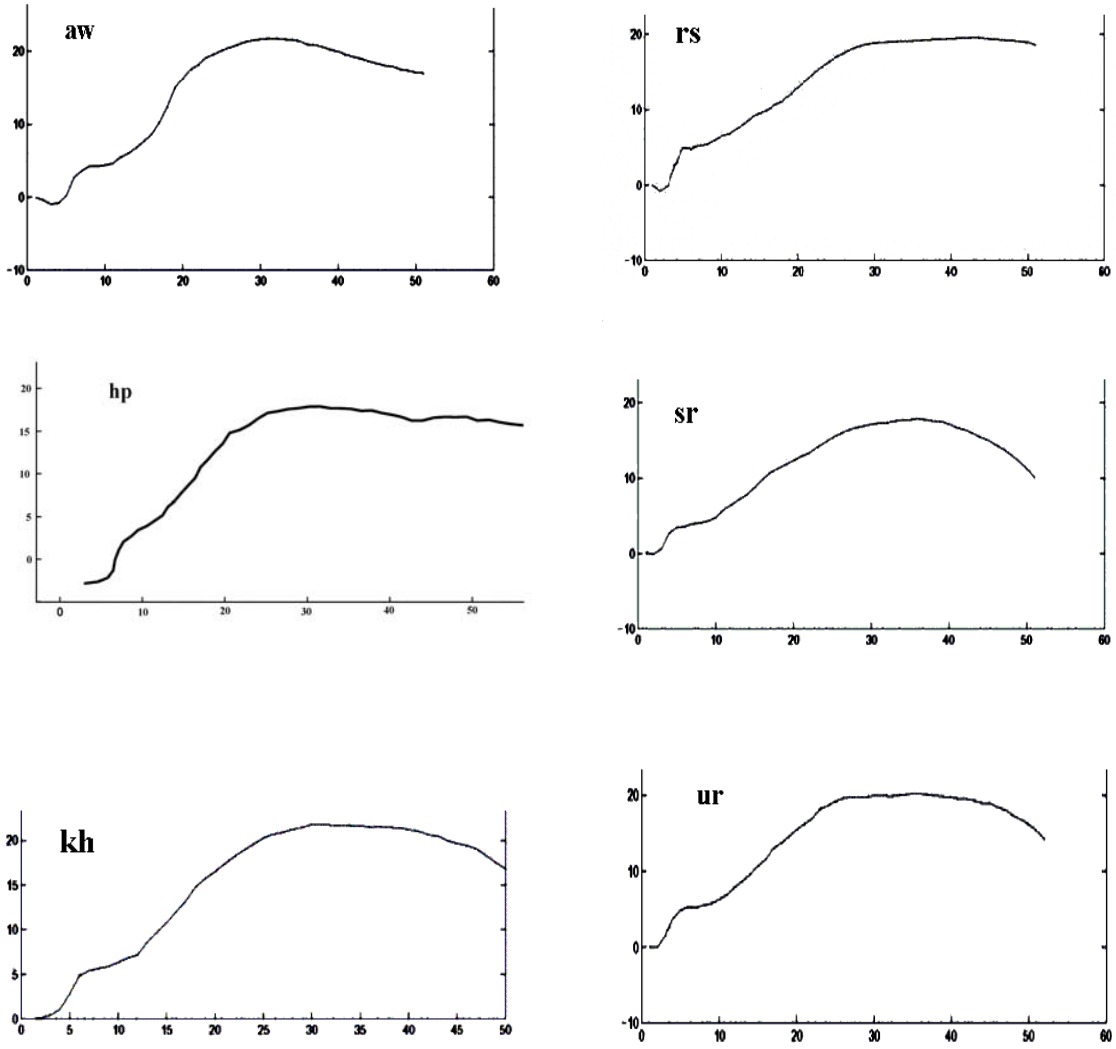


Figure 6-28 Palate contours for all speakers. Scaling in [mm].

Johnson et al. (1993) discussed if differences for vowel production in jaw activity might have to do with back palate vault shape. Ladefoged et al. (1971) argued that a flat back part of the palate vault might be correlated with higher jaw activity in the production of sounds with a constriction in that region. A constriction within a highly domed palate would best be produced with higher tongue activity. Palate contours are given in Figure 6-28. Speaker ur has a somewhat flat back palate contour but relatively small consonantal jaw activity, speaker sr has a domed contour and stronger jaw activity. The dome factor (see Table 6-5) of both is very similar.

Only speaker rs showed a stronger deviating dome factor and a flat back part of the palate. However, his jaw activity patterns are comparable to the mean of other speakers.

Interestingly, speakers ur and sr have the strongest differences in translational jaw movement as indicated in Table 6-1. During /a:/ production speaker sr shows the largest proportion of

translational jaw movement, while ur shows virtually no jaw translation in the speech tasks at all. It was already said, that this translational effect is only noticable for lowest jaw position during /a:/ production, otherwise one could have assumed implications for our estimation of intrinsic tongue tip activity proportion especially for consonant production.

Speaker	dome factor (depth of palate vault/length alveolar ridge to dorsum)	informal description
aw	0.21	highly domed back part of palate
hp	0.21	domed back part of palate
kh	0.20	domed back part of palate
rs	0.11	flat back part of palate
sr	0.18	domed back part of palate
ur	0.17	flat back part of palate

Table 6-5. Typology of palate contours. The dome factor was determined by the midsagittal palate contour extracted from dental casts for each speaker, the distance from alveolar ridge to the point after the second molars (end of dental cast) was determined. The highest position in the palate vault from this base line was then determined.

7. Acoustic Results

Acoustic data for all speakers were segmented as described in chapter 5.2. The segmentation was based on the acoustic time signal (oscillogramm) and was performed with the Mtnew program (written by Phil Hoole) in Matlab. The further analysis of the 24kHz/16bit data was performed with the Praat program package (version 3.9.00 for linux), developed by Paul Boersma and David Weenink, University of Amsterdam. Our analysis focuses, beside intensity, on durational patterns and acoustic analysis of the vowels (fundamental frequency and formants).

7.1 Intensity

Intensity analysis was performed for the following types of segments:

- Whole sentence
- VCV sequence, in the following called (pseudo) word
- Intervocalic target consonants
- target vowels (V1 and V2)

The algorithm is described in the following. Values in the sound are squared and convolved with a Kaiser-20 window (sidelobes below -190dB). The effective length of this window is 3.2/minimum pitch, which guarantees that a periodic signal is analysed with a pitch-synchronous intensity ripple <0.00001dB. Minimum pitch was set to 100 Hz for male speakers and 130 Hz for the female speaker. Time steps were computed as 1/minimum pitch.

A multifactorial analysis of variance (general linear model, univariate, three fixed factors) revealed that vowel type, vowel position and speaker have all a highly significant effect on vowel intensity, interactions between vowel type and speaker and vowel position and speaker are highly significant as well ($p<0.001$).

In Figure 7-1 to Figure 7-4 results for volume level differences are presented. All data were tested with independent samples t-tests. Normal distribution could be assumed on a sufficient level (one sample Kolmogorov-Smirnov tests were performed), homogeneity of variances was tested with Levene's test.

The significant effect is certainly not surprising. More interesting are individual differences. Speaker ur showed in general the highest intensity for loud speech, and largest difference between two volume levels. This is in accordance with the auditory impression of the investigator, that this speaker's loud volume came close to shouting.

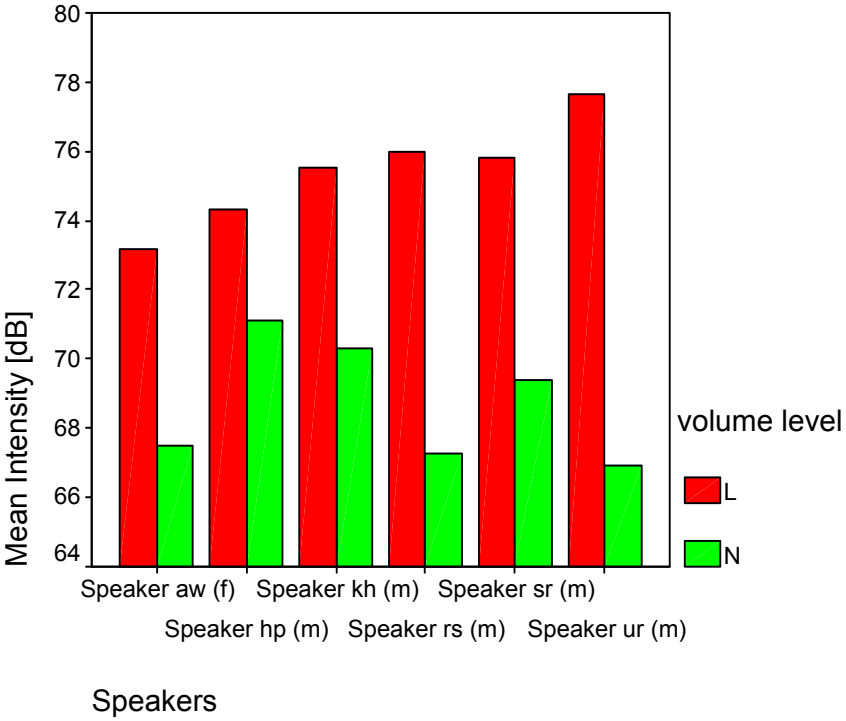


Figure 7-1a. Intensity of vowel segments (V1 and V2). T-test was performed and revealed for all speakers a highly significant effect of volume level.

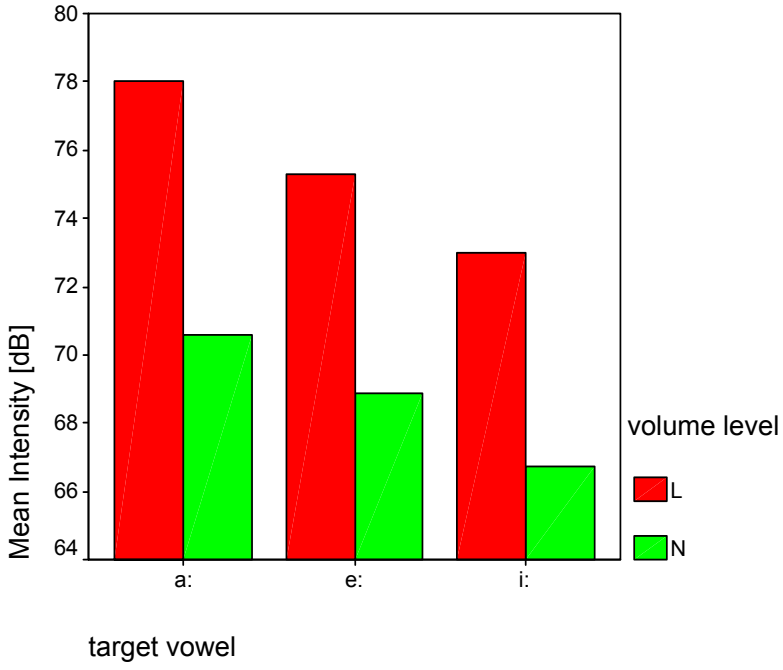


Figure 7-1b. Intensity of different target vowels. T-test was performed and revealed for all vowels a highly significant effect of volume level.

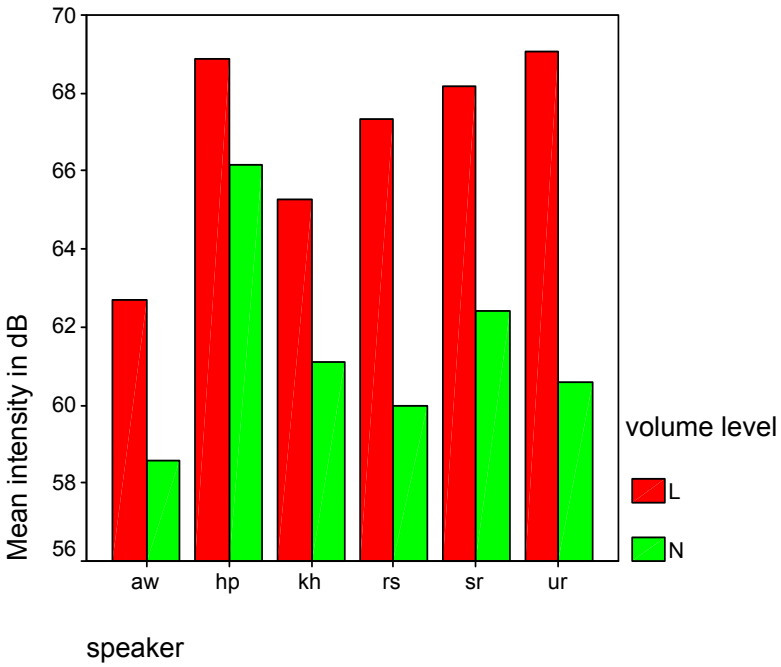


Figure 7-2a. Intensity of consonant segments (target C). T-test was performed and revealed for all speakers a highly significant effect of volume level.

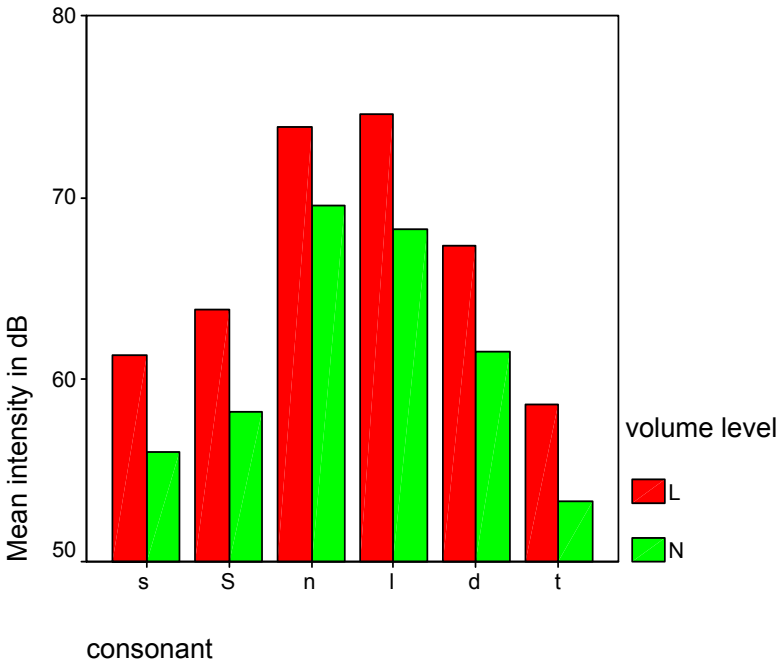


Figure 7-2b. Intensity of different target consonants. T-test was performed and revealed for all consonants a highly significant effect of volume level. Intensity of /t/ is based on the burst portion.

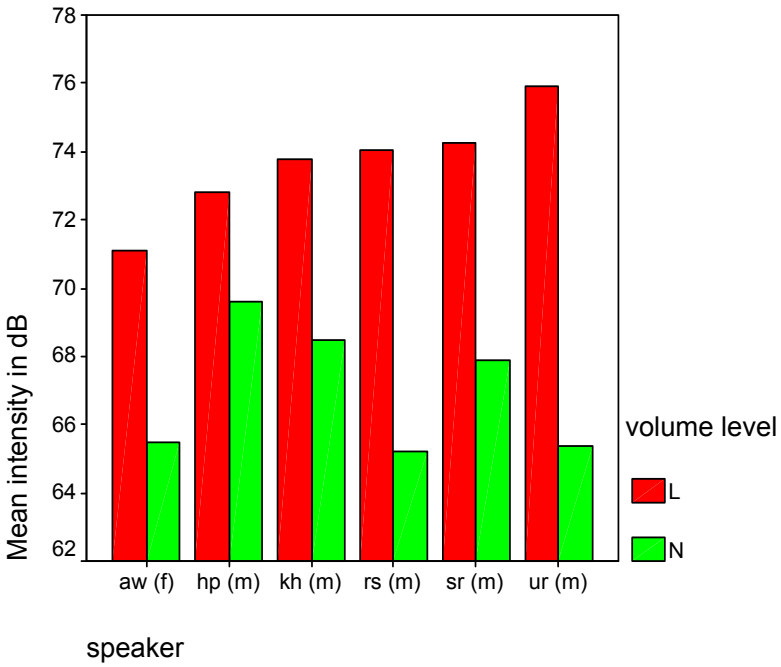


Figure 7-3. Intensity of word segments. T-test was performed and revealed for all speakers a highly significant effect of volume level.

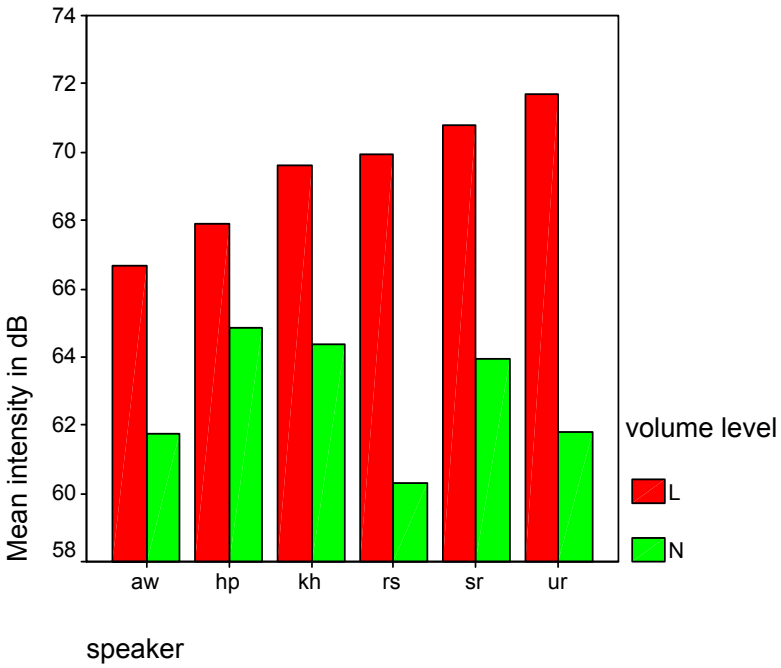


Figure 7-4. Intensity of whole sentences. T-test was performed and revealed for all speakers a highly significant effect of volume level.

7.2 Duration

Durations were determined for the same segment types as for intensity in 7.1. Figures 7-5 to Figure 7-8 present the effect of volume level on duration. A multifactorial analysis (general linear model, univariate, three fixed factors) revealed that vowel type, vowel position and speaker have all a highly significant effect on vowel duration, all interactions between factors are highly significant as well ($p<0.001$).

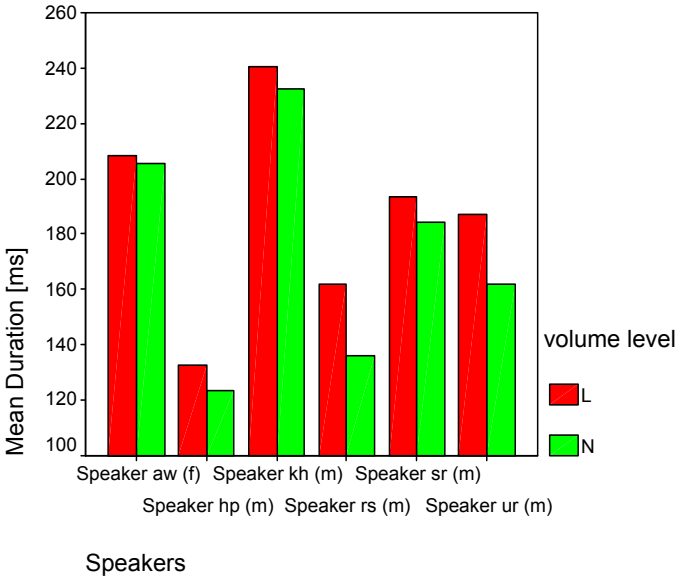


Figure 7-5a. Duration of vowel segments (V1 and V2). T-test was performed and revealed for speakers rs and ur a highly significant ($p<0.001$) effect of volume level.

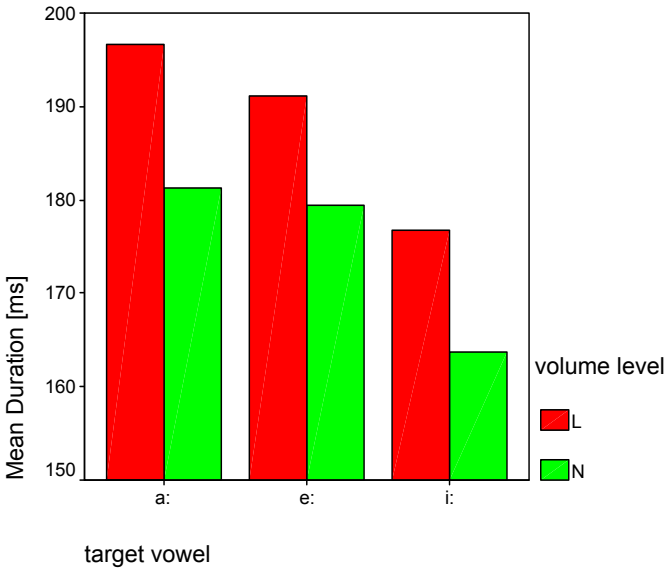


Figure 7-5b. Duration of different vowel targets. T-test was performed and revealed for all vowels a highly significant ($p<0.001$) effect of volume level.

Vowel duration (see Figure 7-5) is for two speakers significantly higher in loud volume condition. Further inquiry for individual speakers revealed that speakers hp and sr tended to reduce the unstressed /e:/ substantially, to almost Schwa like quality. This had also a strong effect on duration, independent of volume level. This might explain why means for these two speakers are not different on a statistically significant level.

Intrinsic vowel duration differs substantially (the higher - the shorter); intrinsic vowel duration differences are kept at the higher volume level.

Figure 7-6 shows durational effects of volume on consonantal segments. The tendency is quite clear recognizable in Figure 7-6b. Loud voice condition shows for all consonants except /t/ a highly significant shorter duration. The results in Figure 7-6a for individual speakers over all consonants are blured by the very large durational differences for different consonant types.

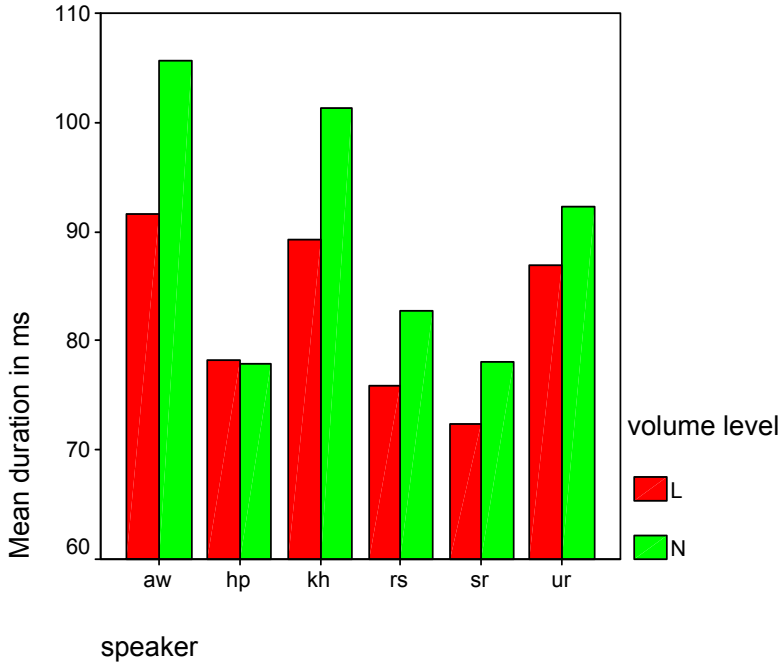


Figure 7-6a. Duration of consonantal segments (target C). T-test was performed and revealed a highly significant effect of volume level for aw and kh ($p<0.001$), a significant effect for rs ($p<0.05$).

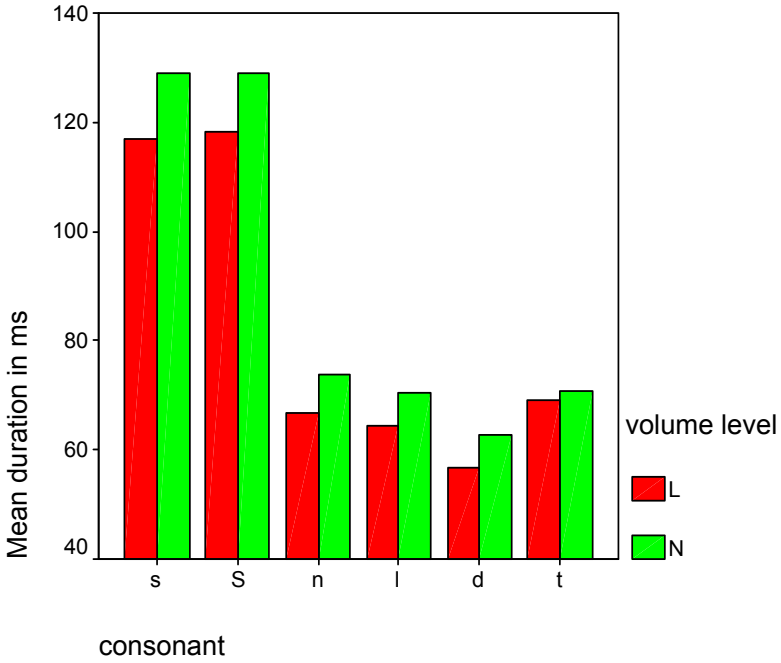


Figure 7-6b. Duration of consonantal segments (target C). T-test was performed and revealed for /s/, /ʃ/, /n/, /l/, /d/ a highly significant effect ($p<0.001$, for /d/: $p<0.01$) of volume level. /t/ non-significant.

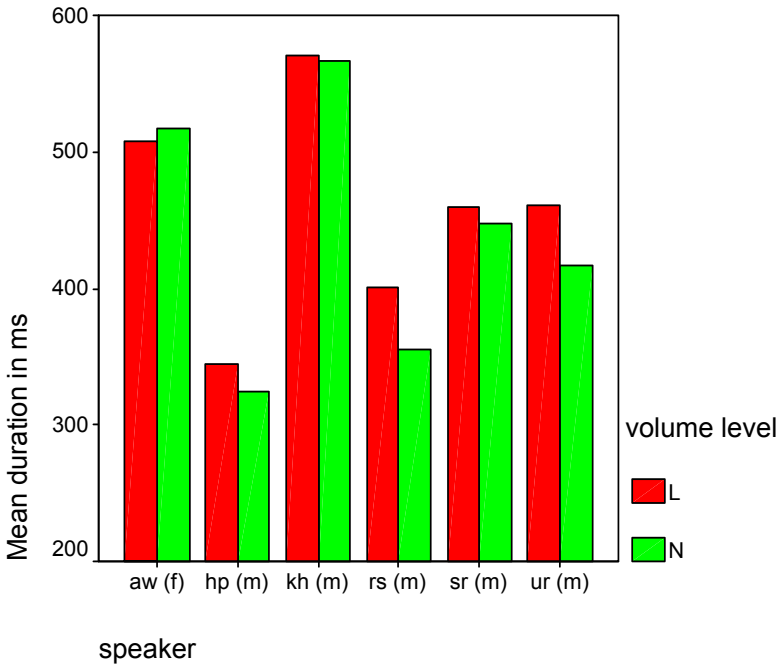


Figure 7-7. Duration of word segments. T-test was performed and revealed for speakers hp, rs and ur a highly significant ($p<0.001$) effect of volume level.

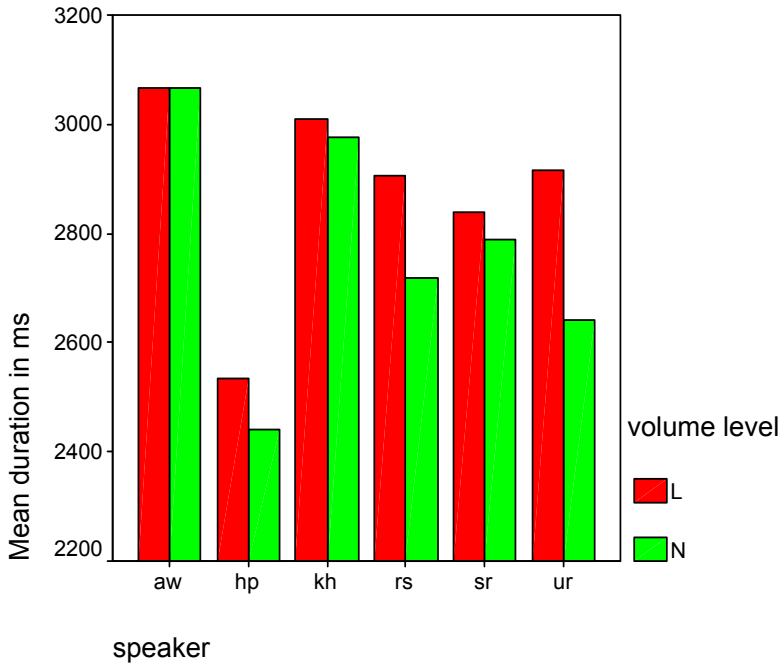


Figure 7-8. Duration of whole sentences. T-test was performed and revealed for speakers hp, rs and ur a highly significant ($p<0.001$) effect and for speakers kh and sr a significant ($p<0.05$) effect of volume level.

Word duration and sentence duration tends to be longer for loud volume level (see Figure 7-7 and 7-8). The effect for words which consist of two vowels and one consonant is of course most probably a reflection of longer vowel duration. For sentence duration we determined the number of consonantal vs. vowel phonemes. The phoneme count of single word citation forms for the whole sentence (not counting three presumably vocalized /r/ phonemes) is 20 consonants vs. 12 vowels (4 long vowels). Since the long vowels are considerably longer than consonants, they might still affect the sentence result, although this seems to be less probable. There might be additional effects accounting for longer sentence duration in loud speech, like more pauses. The auditive inspection of all sentences, performed in the process of acoustic segmentation, though, did not reveal noticable breathing pauses, during single sentence production.

7.2.1 Correlation of intensity and duration

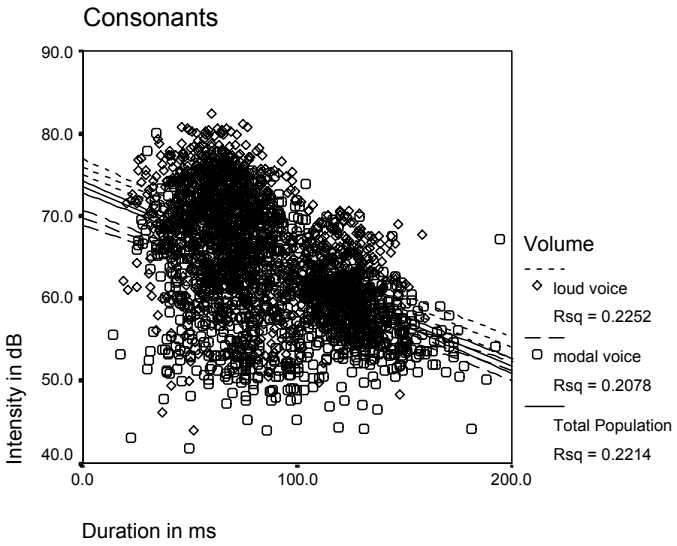


Figure 7-9. Correlation of duration and intensity for consonants.

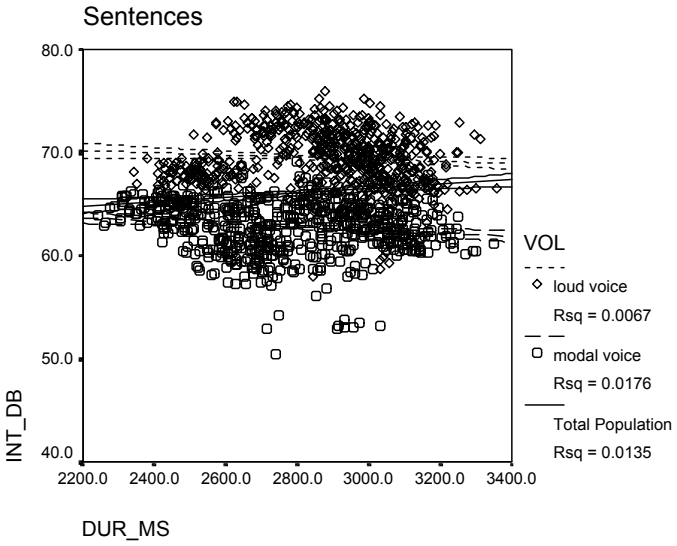


Figure 7-10. Correlation of duration and intensity for sentences.

Duration and intensity in consonants show a clear negative correlation (Figure 7-9), which was expected from the results presented in the preceding subchapters. This pattern is consistent if the data are split up for individual speakers and consonant types, data in 7-9 are given across speakers and consonant types since they are robust enough to show the effect.

Scatterplots of duration and intensity for whole sentences (Figure 7-10) and over all vowels reveal no correlation. Slightly distinct patterns can be found for stressed and unstressed vowels (Figure 7-11). This might be caused by greater dispersion of intensity in unstressed vowels. There is even a very slight tendency of negative correlation within volume level for unstressed vowels, Pearson coefficients are highly significant and vary between approx. -2 and -3.5 (one exception unstressed

loud /a:/ Pearson correlation coefficient -.119, significant on a $p<0.05$ level). The tendency of a slight positive correlation for stressed vowels, but negative correlation for unstressed vowels is even stronger for data of individual speakers, pooling vowel qualities. It must be noted that speakers differ considerably in the height of the correlation coefficient, speaker hp shows even for unstressed vowels a slight positive correlation. As an example data for speaker ur are given in Figure 7-12.

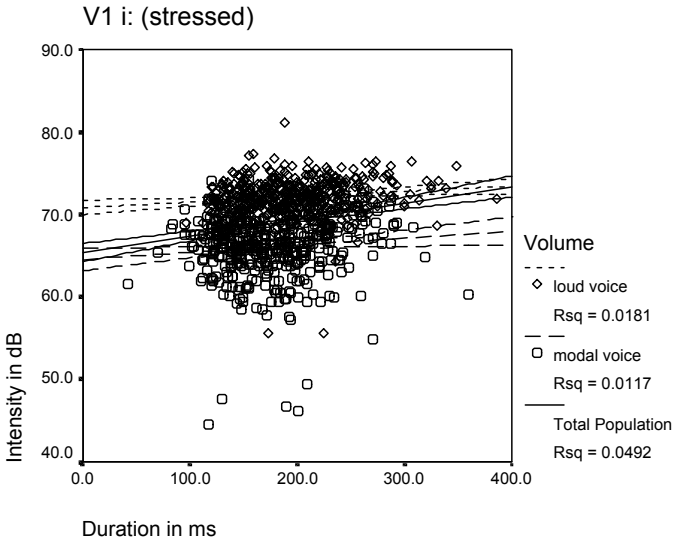


Figure 7-11a. Correlation of duration and intensity for stressed vowel /i:/.

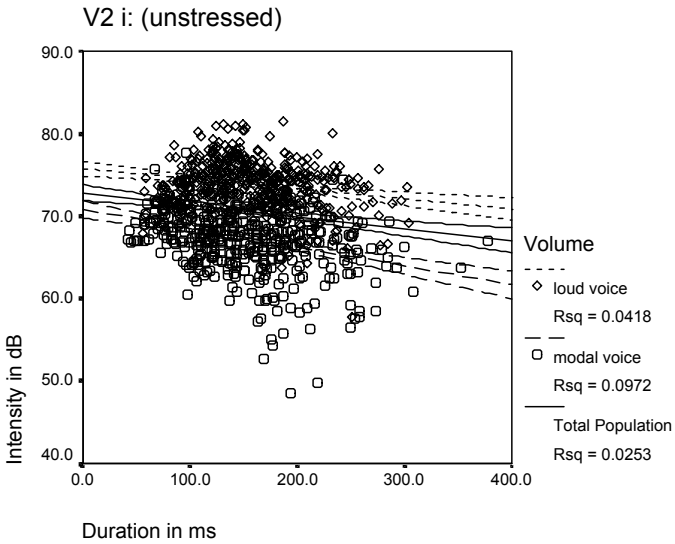


Figure 7-11b. Correlation of duration and intensity for unstressed vowel /i:/.

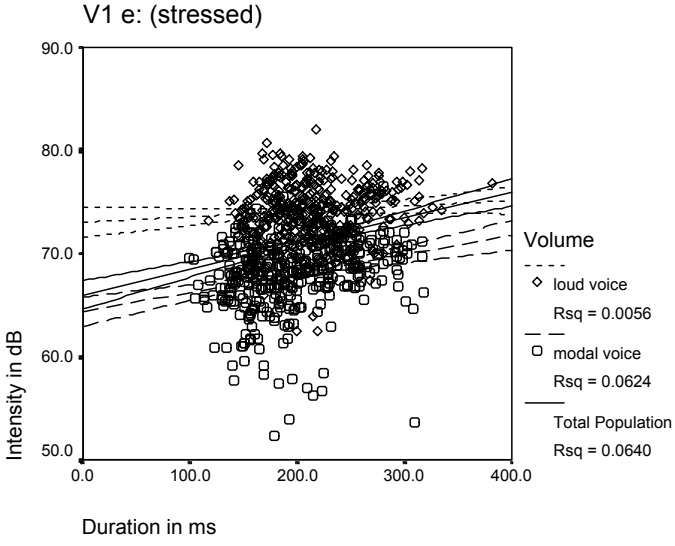


Figure 7-11c. Correlation of duration and intensity for stressed vowel /e:/.

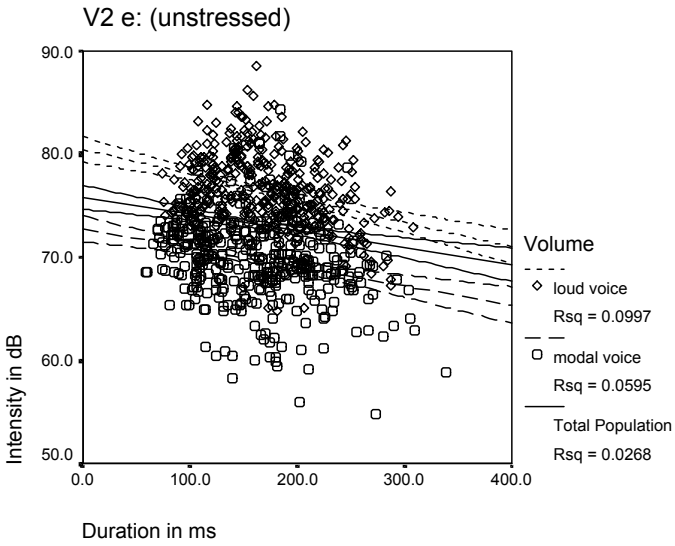


Figure 7-11d. Correlation of duration and intensity for unstressed vowel /e:/.

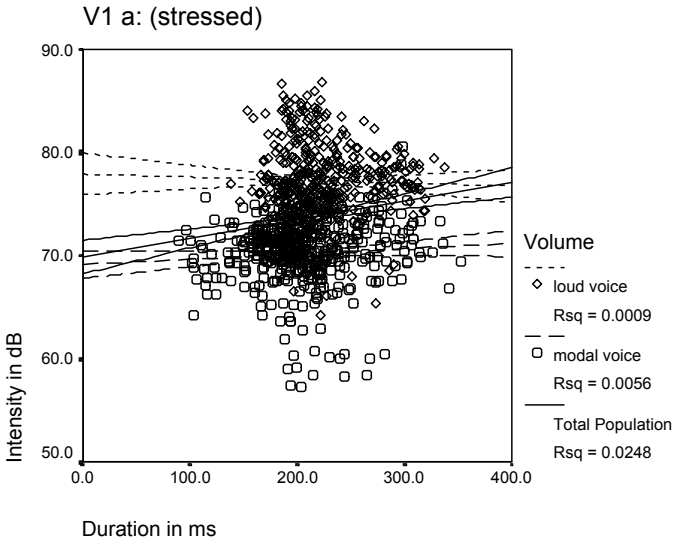


Figure 7-11e. Correlation of duration and intensity for stressed vowel /a:/.

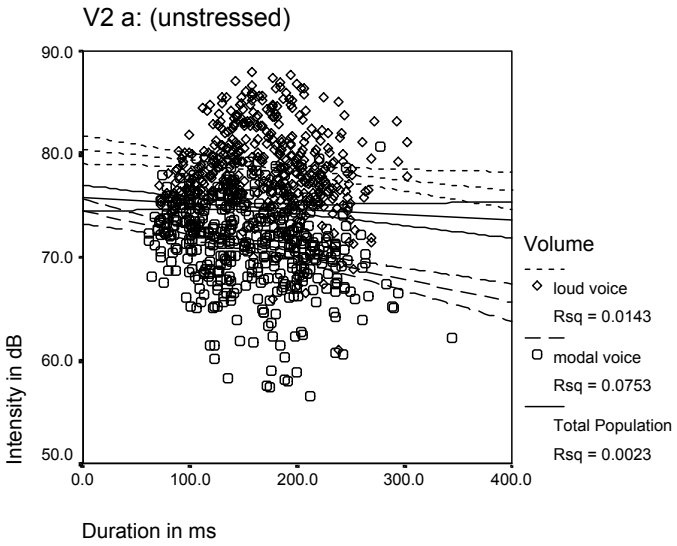


Figure 7-11f. Correlation of duration and intensity for unstressed vowel /a:/.

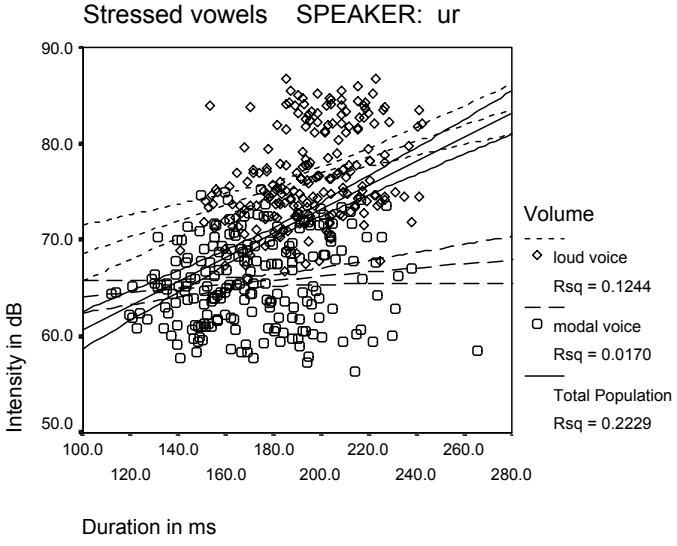


Figure 7-12a. Correlation of duration and intensity for stressed vowels of speaker ur.

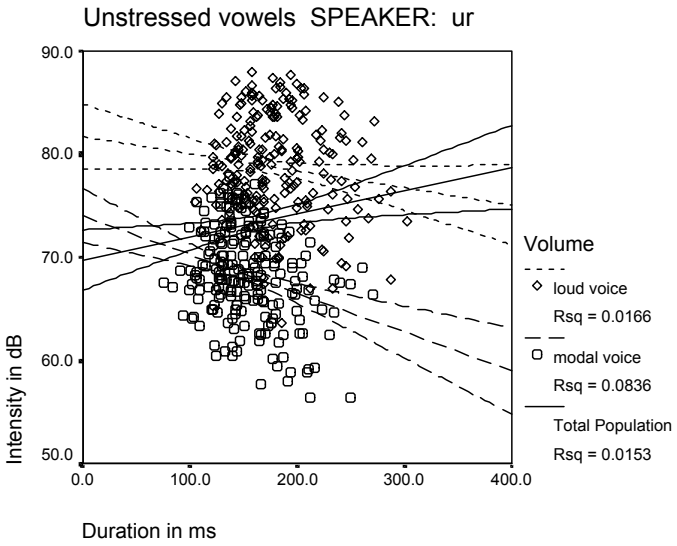


Figure 7-12b. Correlation of duration and intensity for unstressed vowels of speaker ur.

7.3 Fundamental frequency

The automatic detection of fundamental frequency is in general problematic. Nevertheless, since the corpus is relatively large (5120 vowels), an at least partially automated analysis was highly preferable. Tests with the then used algorithm showed very reliable results, which might certainly be due to the fact that sound quality was very good, and only long vowels were analysed.

Fundamental frequency analysis was performed with Praat using a method based on autocorrelation, described in detail in Boersma (1993). Minimum pitch was set to 75Hz and maximum pitch set to 600Hz, with 10ms time steps. Pitch contours were inspected randomly, and showed very few errors. An average value of vowel fundamental frequency was then determined

by the median of all values in the time frame of the vowel. The obtained results were then checked for clear outliers. Only for some speakers were a very few values removed.

The general tendency of a clearly higher fundamental frequency for loud voice is shown in Figure 7-13. In Figure 7-14 fundamental frequencies (mean of all vowel qualities) for individual speakers are presented. An independent samples t-test for each speaker was performed to test the significance of the volume effect on fundamental frequency. The effect is for all speakers highly significant on a $p<0.001$ level. Nevertheless there are obviously large differences between speakers. Especially speaker ur shows very large mean differences for loud vs. normal voice. In the literature large f0 effects have been noted for shouted vs normal speech (Rostolland 1982). This would be in accordance with the especially high vowel intensity differences observed for this speaker (see Figure 7-1a). The other extreme is speaker hp, who shows only moderate, though significant, changes in fundamental frequency. He showed as well only moderate, though significant, differences in vowel intensity.

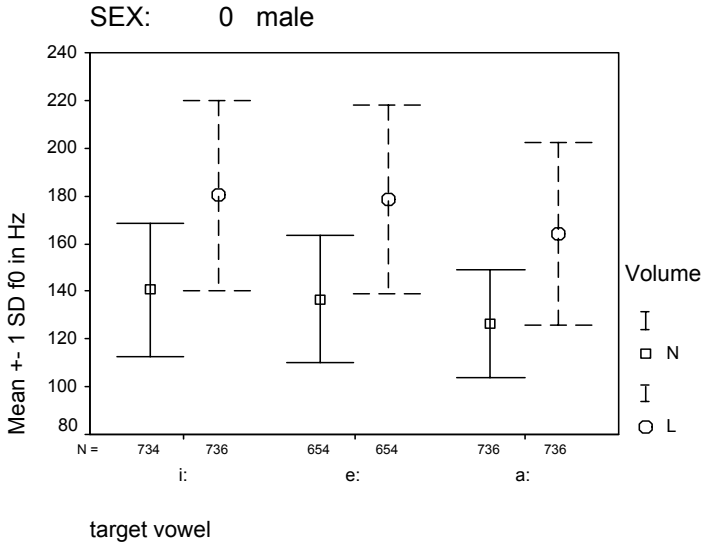


Figure 7-13a. Effect of volume level on average fundamental frequency in long vowels for male speakers, mean ± standard deviation.

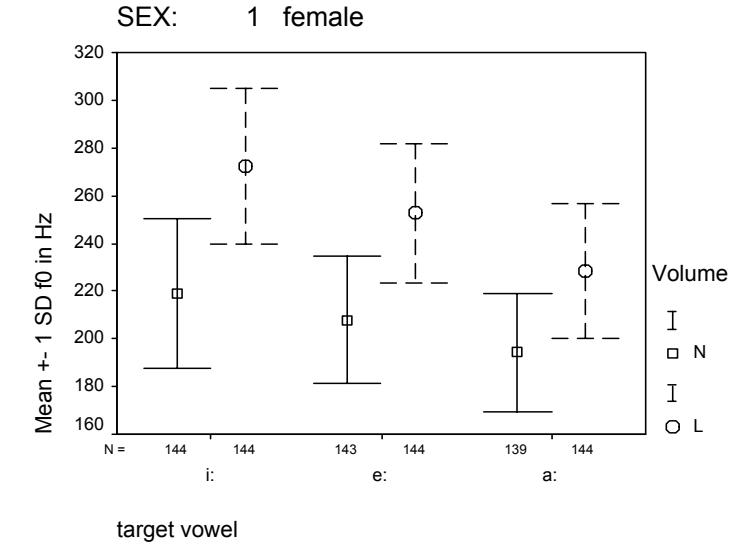


Figure 7-13b. Effect of volume level on average fundamental frequency in long vowels for one female speaker, mean ± standard deviation.

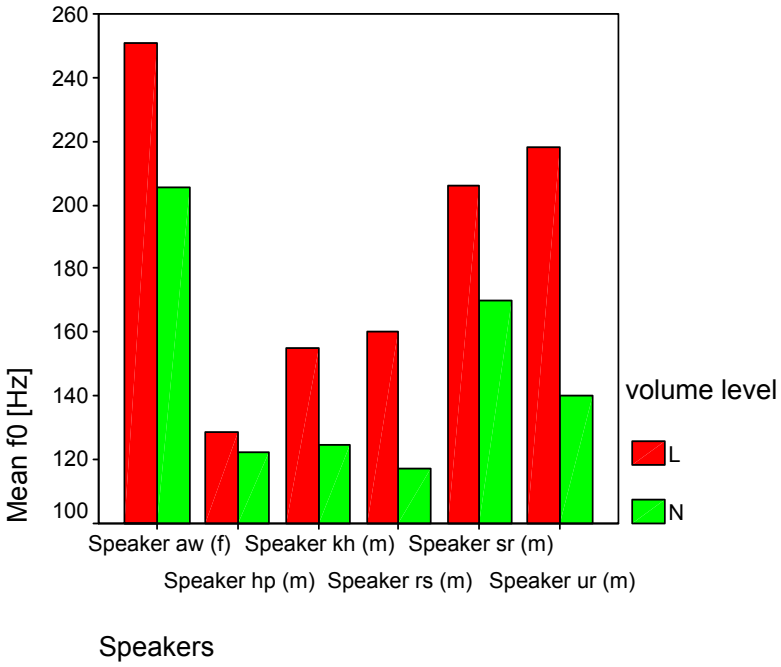


Figure 7-14. Effect of volume level on average fundamental frequency in long vowels for individual speakers. Independent t-test comparison of means showed that effect of volume level is for all speakers highly significant ($p<0.001$).

7.3.1 Correlation of f0 with intensity

We have seen that a clear volume level effect on fundamental frequency can be oberved. We will now look more deeply into correlations between intensity and fundamental frequency in the vowel

segments. Data for five male speakers are presented together in Figure 7-15. Figure 7-16 presents scatterplots for two individual speakers aw and rs for each target vowel separately .

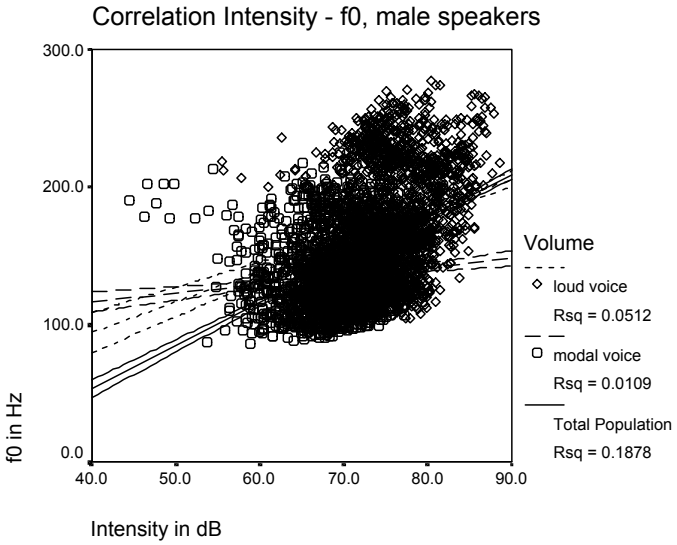


Figure 7-15. Scatterplot of intensity vs fundamental frequency for five male speakers.

The pooled data of five speakers in Figure 7-15 show a general tendency of positive correlation beween intensity and fundamental frequency, as well as a positive correlation within the two volume levels. For individual speakers the patterns within loud volume level show partially even a slight negative correlation, this effect, though is very weak.

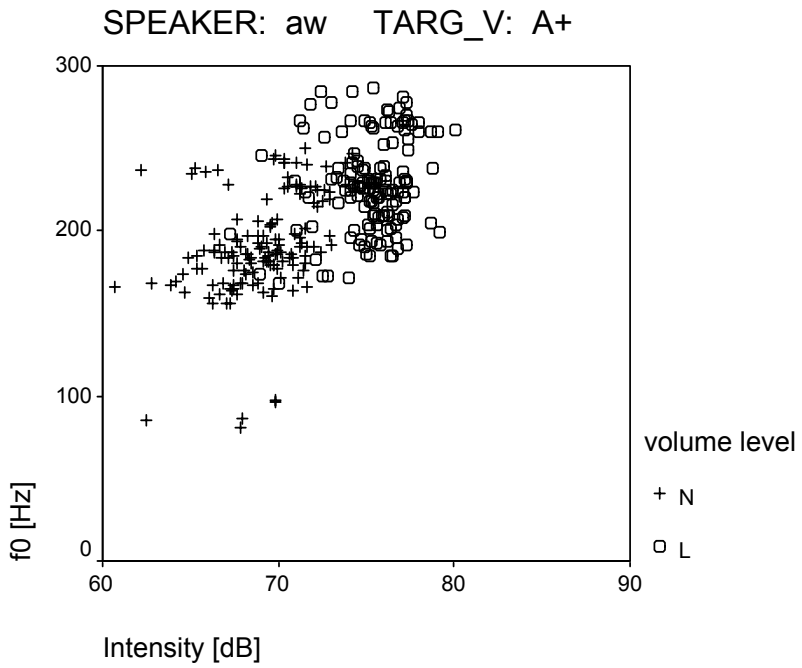


Figure 7-16a. Scatterplot of intensity vs fundamental frequency for vowel /a:/ produced by the female speaker aw.

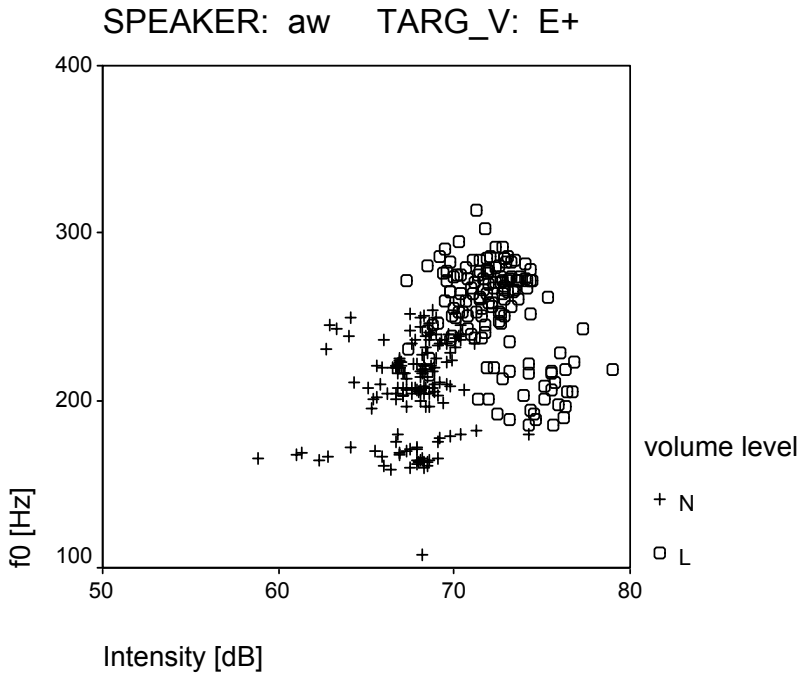


Figure 7-16b. Scatterplot of intensity vs fundamental frequency for vowel /e:/ produced by the female speaker aw.

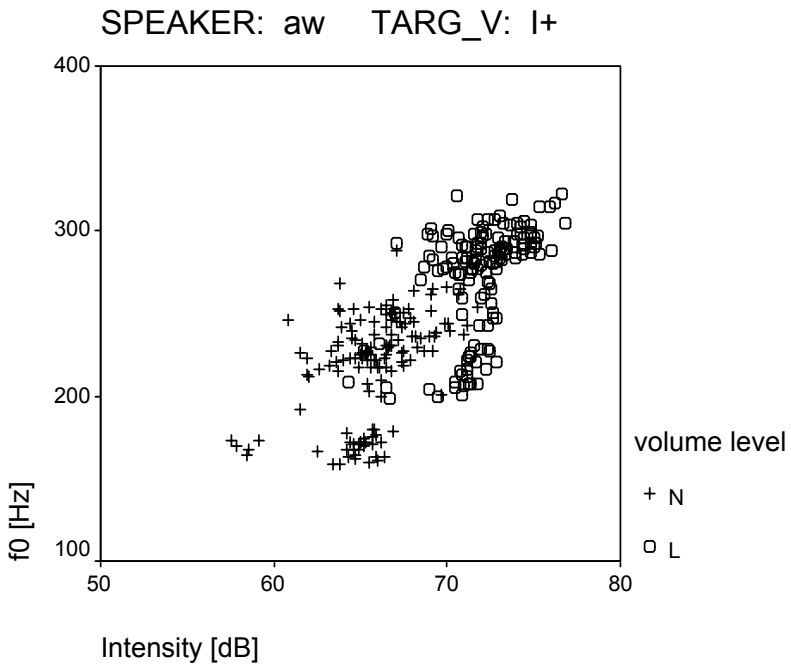


Figure 7-16c. Scatterplot of intensity vs fundamental frequency for vowel /i:/ produced by the female speaker aw.

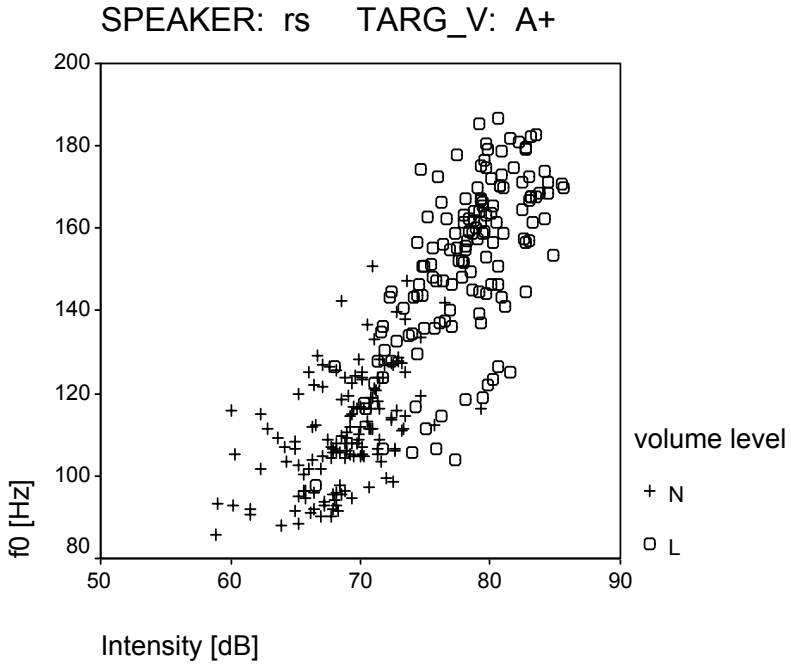


Figure 7-16d. Scatterplot of intensity vs fundamental frequency for vowel /a:/ produced by the male speaker rs.

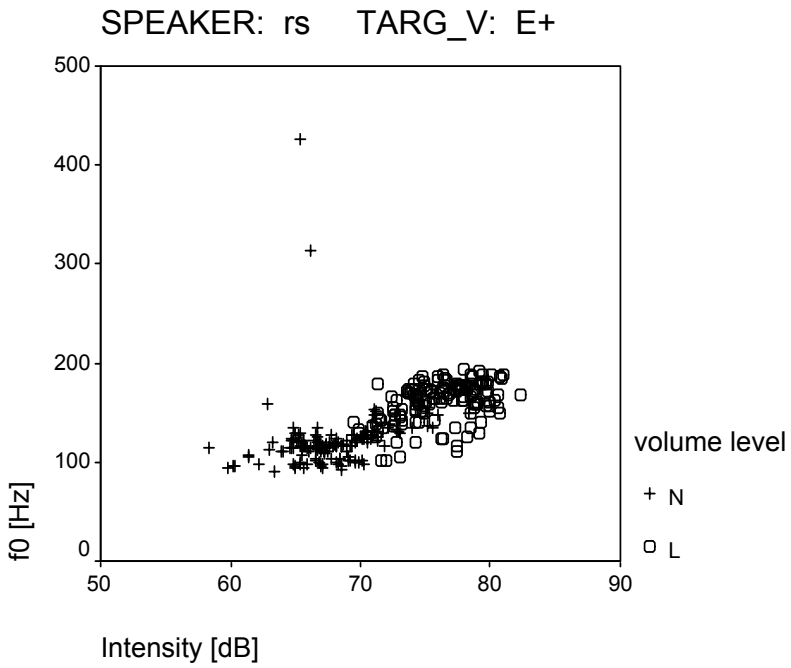


Figure 7-16e. Scatterplot of intensity vs fundamental frequency for vowel /e:/ produced by the male speaker rs.

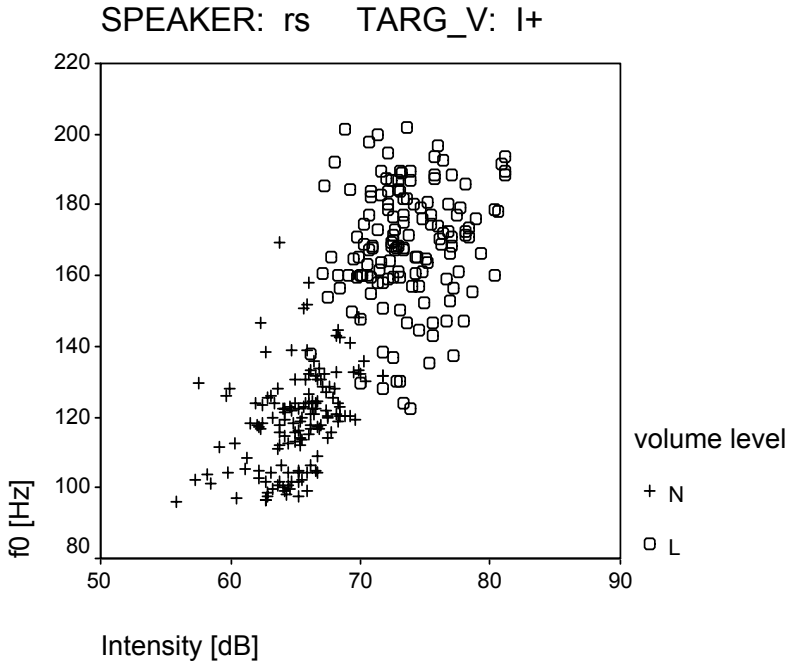


Figure 7-16f. Scatterplot of intensity vs fundamental frequency for vowel /i:/ produced by the male speaker rs.

7.3.2 Coarticulatory effect of consonantal context on f0

Analyses of vowel fundamental frequency effects of consonantal context were performed for different subgroups but no significant effects were found.

A multifactorial analysis of variance of the fundamental frequency (general linear model, univariate, two fixed factors: volume and consonant context, split up for speaker and vowel type) revealed no significant effect of consonant context and no significant interaction between volume level and consonantal context.

As one example data of the female speaker aw for fundamental frequency in V1 for different following consonants are presented in Figure 7-17.

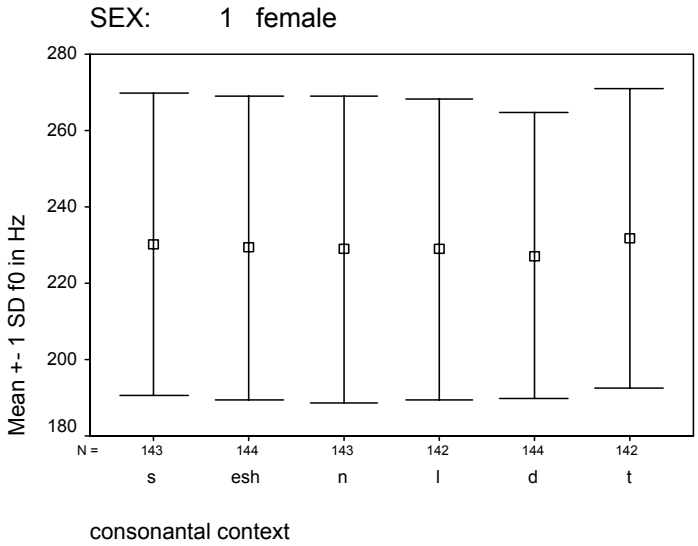


Figure 7-17. Fundamental frequency differences in V1 (all vowel qualities pooled) as effect of following consonant type. Data for female speaker aw. A unianova was performed and showed no significant effect, as for other speakers here not reported.

7.4 Formant analysis

Formant analysis was performed with the Praat implementation of the Burg LPC coefficient algorithm.

- Frame steps were 10ms
- F5 formants were extracted
- 5 formants were extracted for male speakers below 5000 Hz, for the female speaker below 5500Hz
- Window length 25ms
- Pre-emphasis from 50 Hz +6dB/Octave

Averaged formant values (F1, F2, F3) for each vowel were then determined as the median of all values for each formant in the time frame of the vowel. This method seems to be applicable since all vowels were rather long. With short vowels or in more spontaneaous speech, the actual target formant value more in the middle of the segment, would probably have been blured. A method of

using instead a single formant value in the middle of the vowel, led to more inconsistent results, the median is more robust against errors.

The analysis of formants is for this study certainly of special interest, since formants can be extracted from the acoustic signal, but contain as well information about shape and resonances of the vocal tract (e.g Stevens 1998, Tillmann 1980). The first formant is best associated with information about the vowel height, roughly speaking the higher the vowel, the lower is the first formant frequncy. The second formant is associated with the front vs back dimension of the vowels, i.e. a front vowel has a higher second formant. Formant analysis then allows us to obtain more information on tongue configuration, an important supply, since our lingual articulatory analyses are restricted to the tongue tip.

In some analyses formant values are presented in Bark instead of Hertz. It is a well discussed fact (e.g. Stevens 1998, Heid 1998) that interactions between fundamental frequency and formant frequency exist. Further it is known, that the spectral resolution of the auditory system is not linear as the Hertz frequency scale. A method, that takes both facts into account is presented by Syrdal and Gopal (1986), and uses the logarithmic Bark-scale (as proposed by Zwicker and Terhardt 1980).

Formant frequencies in Hertz were post processed to Bark-rate with the following formula:

$$7 \ln (x/650 + \sqrt{(1 + (x/650)^2)})$$

The method is discussed in Boersma (1998: 104f). We have not evaluated differences to the Zwicker/Terhardt (1980) formula in detail, but our Bark values are very similar to those presented by Heid (1998) for German vowels. Differences of bark-scaled fundamental frequency and formant frequencies were calculated afterwards.

7.4.1 Effects on F1

In Figure 7-18 the effect of loud voice on the first formant of different vowel qualities is shown. For all vowel qualities the first formant is significantly higher in loud speech. This is consistent and highly significant in all speakers. In articulatory terms this means that all vowels are significantly lowered, which fits with the data for jaw articulation. Data are presented in Hertz (7-18a) and Bark (7-18b).

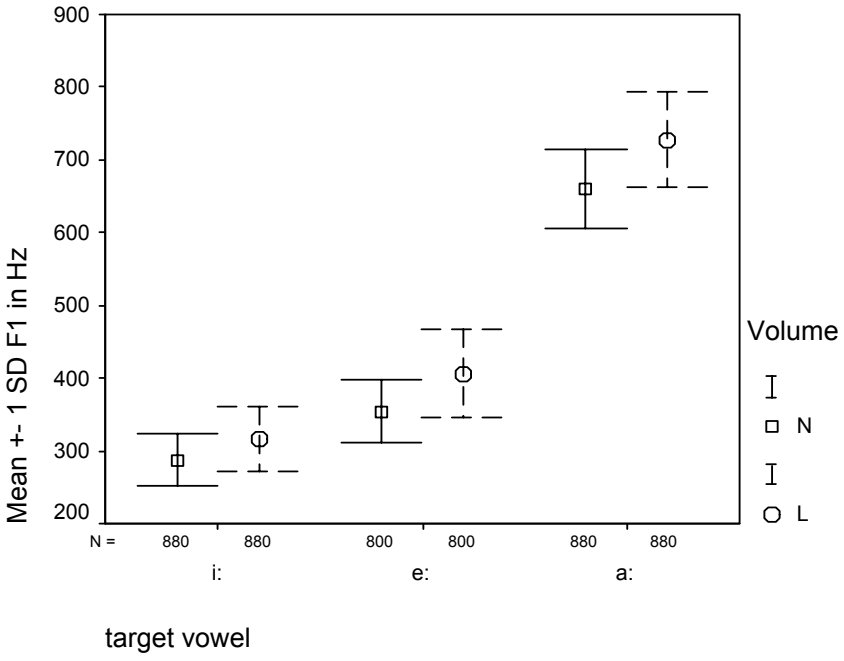


Figure 7-18a. Effect of volume level on first formant. Data are given in Hertz, error bars indicating standard deviations. Data pooled for all speakers and V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F1 ($p<0.001$) for all vowels.

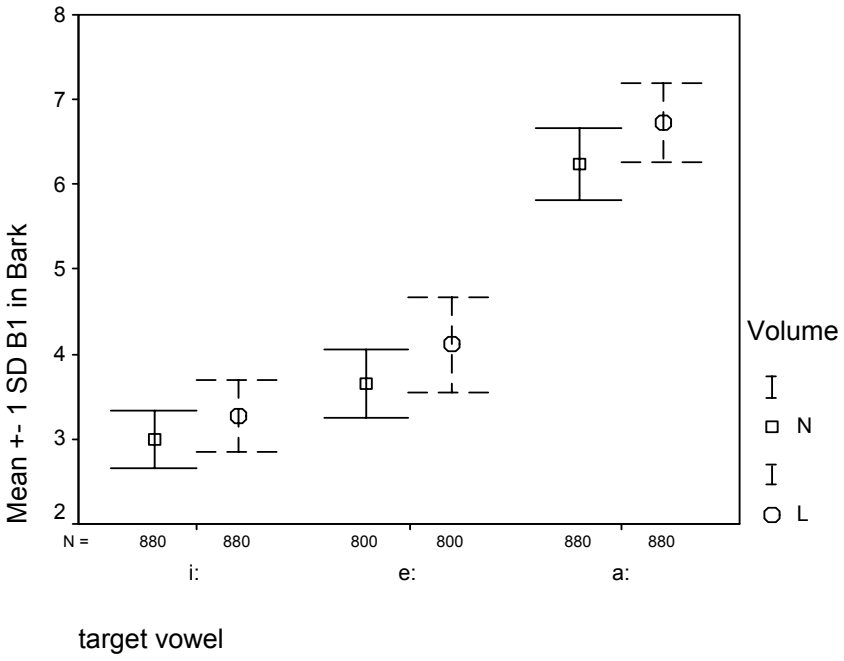


Figure 7-18b. Effect of volume level on first formant. Data are given in Bark, error bars indicating standard deviations. Data pooled for all speakers and V1 and V2.

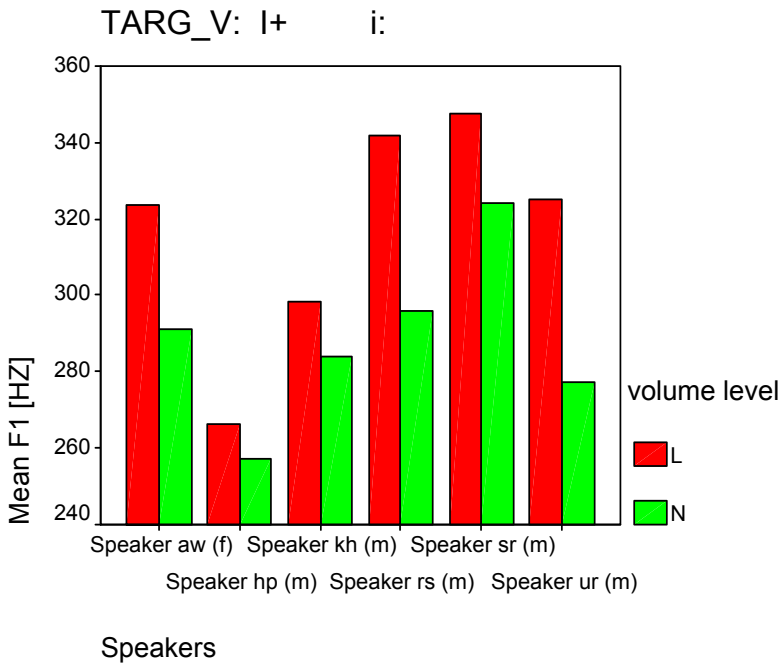


Figure 7-19a. Effect of volume level on first formant of /i:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F1 ($p<0.001$) for all speakers.

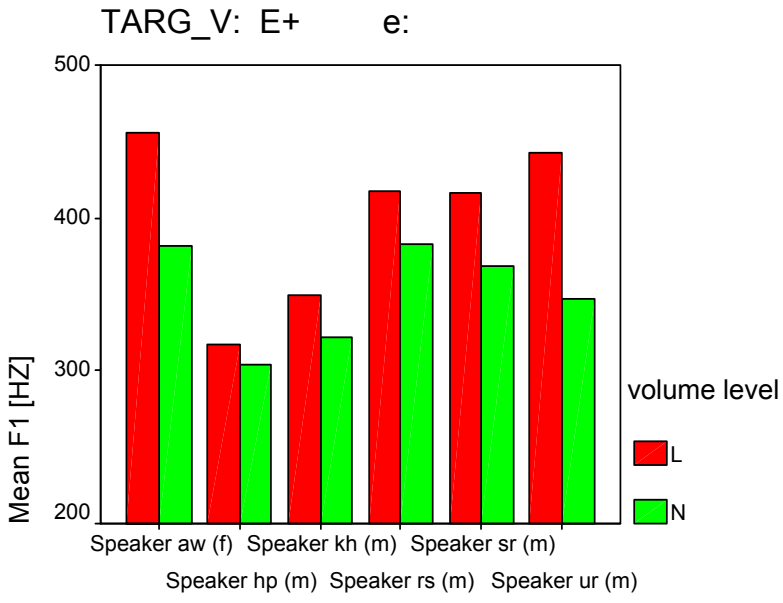


Figure 7-19b. Effect of volume level on first formant of /e:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F1 for all speakers. Speaker hp $p<0.01$; others $p<0.001$.

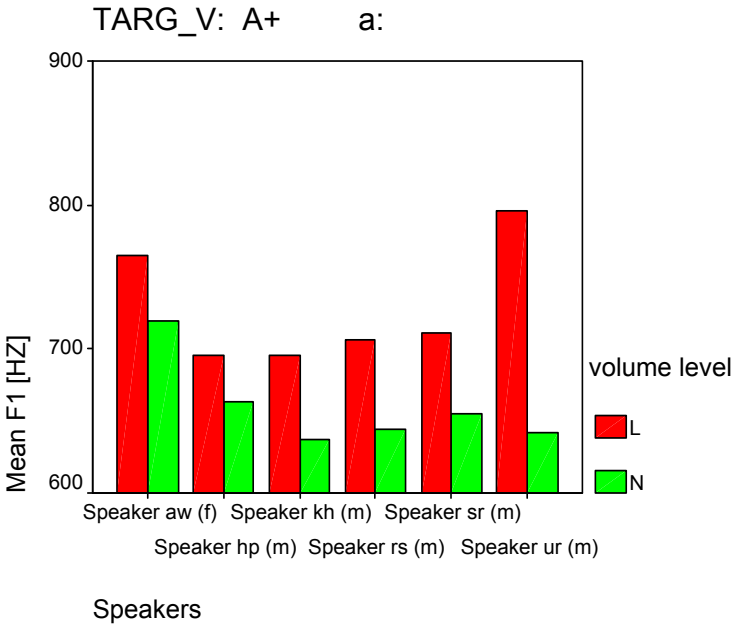


Figure 7-19c. Effect of volume level on first formant of /a:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F1 ($p<0.001$) for all speakers.

7.4.2 Effects on F2

The effect of volume level on the second formant is presented in Figures 7-20 and 7-21. Over all speakers only vowel /a:/ shows a highly significant higher second formant for loud voice. A slight tendency of higher second formant can be found as well for /i:/ and /e:/, but mean differences are only for /i:/ significant at the $p<0.05$ level.

The data for individual speakers (Figure 7-21) differ substantially. For vowel /i:/ three subjects (aw, kh, rs) show a significantly higher second formant for loud voice condition (Figure 7-21a). In Figure 7-21b data for vowel /e:/ are presented. They show for three speakers (hp, rs, ur) a significantly higher second formant for loud voice. Interestingly, speaker aw, who has a very high second formant for /e:/ as well as /i:/ has in loud voice condition a highly significant lowered second formant for /e:/. An explanation might be the fact that a higher second formant for /e:/ would overlap with /i:/.

The effect for vowel /a:/ is highly significant for all speakers. What might be of some interest are the much lower second formants of speaker kh and to a certain extent speaker ur. Since both speakers origin from Southeastern Germany (Bavaria), this is probably due to regional more retracted /ɑ:/ like quality, while for Standard High German some more central /a̠/ quality is assumed, which is here usually transcribed as /a:/.

The articulatory interpretation of the rise of the second formant in loud speech /a:/ could be interpreted as fronting of the central or even back vowel /a:/. (We will further use the symbol of the front cardinal vowel as commonly used

phoneme symbol, neglecting the actual individual vowel quality). The front vowels /i:/ and /e:/ might still show a minimal effect of fronting, but tendentially only with speakers whose front vowels are not completely front (speaker rs).

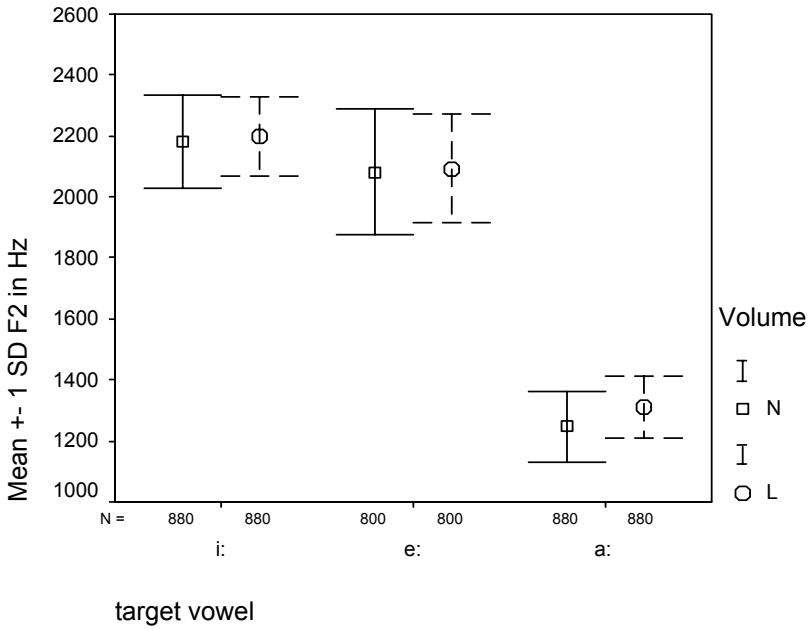


Figure 7-20a. Effect of volume level on second formant. Data pooled for all speakers and V1 and V2 in Hertz with standard deviations. An independent samples t-test was performed and showed a highly significant effect of volume level on F2 ($p<0.001$) for /a:/, significant effect ($p<0.05$) on /i:/, no significant effect on /e:/.

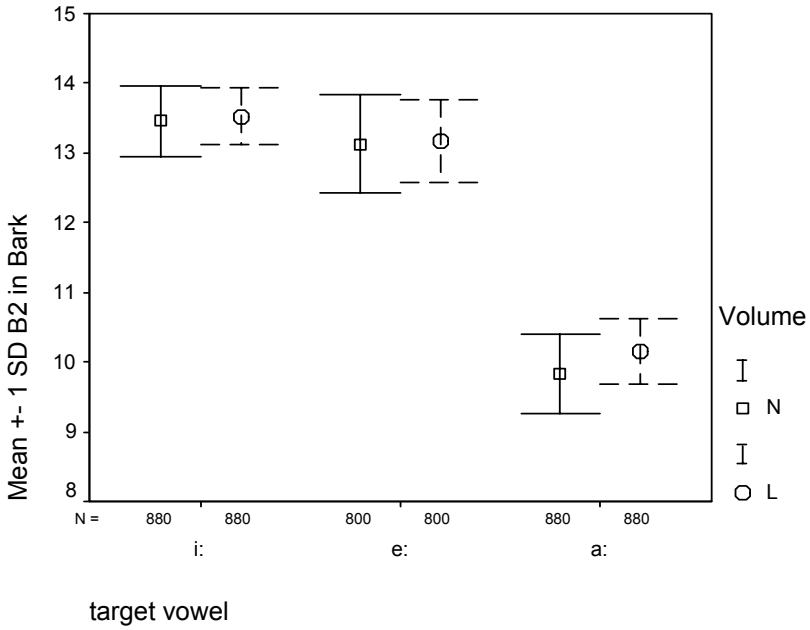


Figure 7-20b. Effect of volume level on second formant. Data pooled for all speakers and V1 and V2 in Bark with standard deviations.

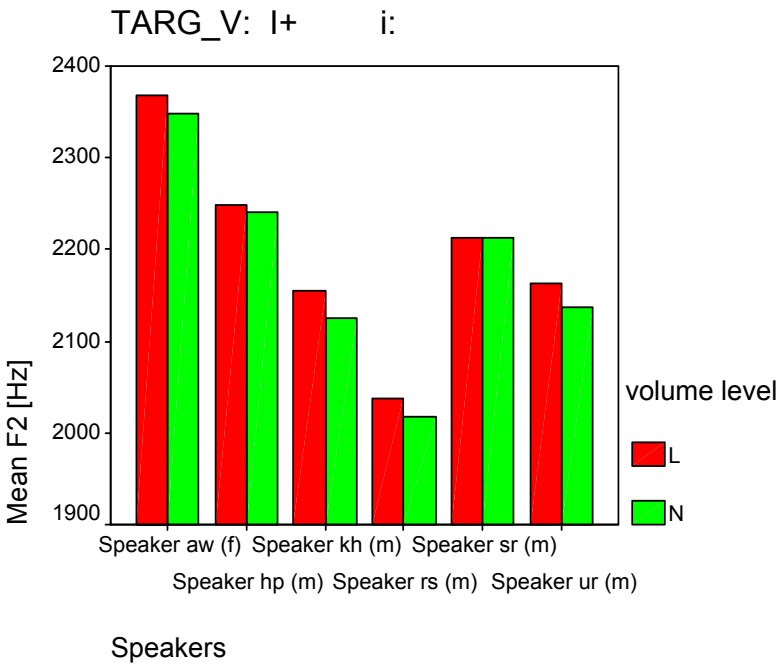


Figure 7-21a. Effect of volume level on second formant of /i:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F2 ($p<0.001$) for speaker kh, a significant effect for speakers aw and rs ($p<0.05$), no significant effect for speakers hp, sr, ur.

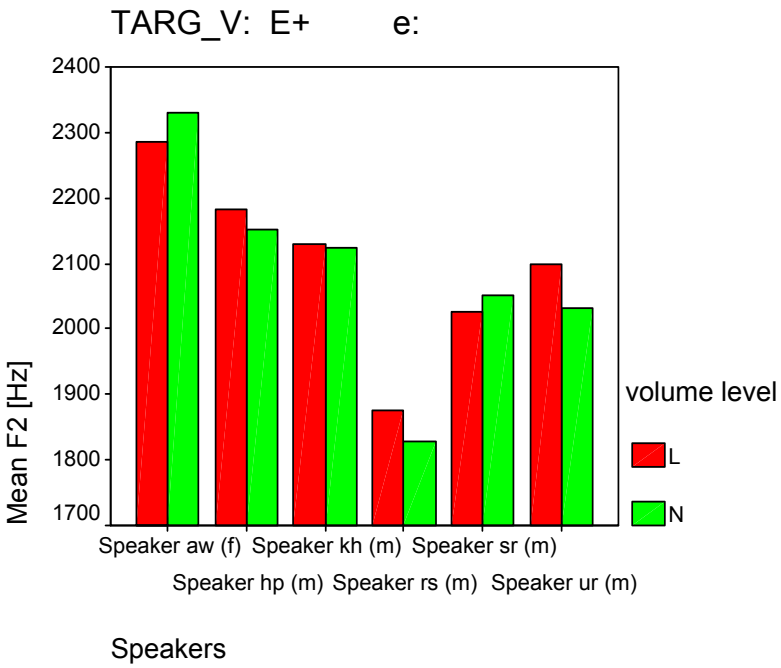


Figure 7-21b. Effect of volume level on second formant of /e:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F2 ($p<0.001$) for speakers aw and ur, a significant effect for speakers hp and rs ($p<0.05$), no significant effect for speakers kh and sr. It should be noted that loud voice effect is for speakers aw and sr opposite to the other speakers.

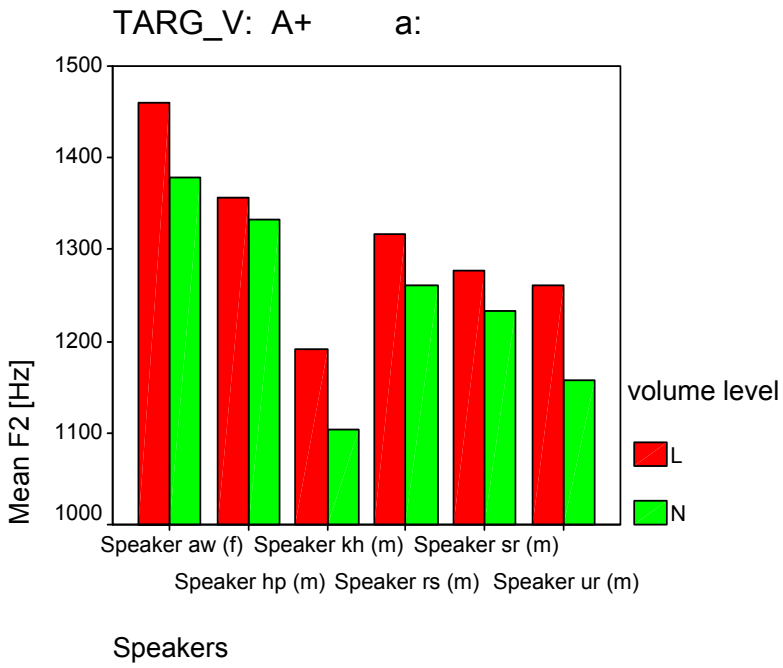


Figure 7-21c. Effect of volume level on second formant of /a:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F2 ($p<0.001$) for all speakers.

7.4.3 Effects on F3

The patterns for loud vowels third formant are presented in Figure 7-22 and 7-23. They can be best described as centralizing. Vowels with high third formant (/i:/ and /e:/) have in loud condition a lowered F3, the low F3 of vowel /a:/ is heightened. Differences for individual speakers differ substantially, for results and respective significance levels see Figure 7-23a-c and the legend.

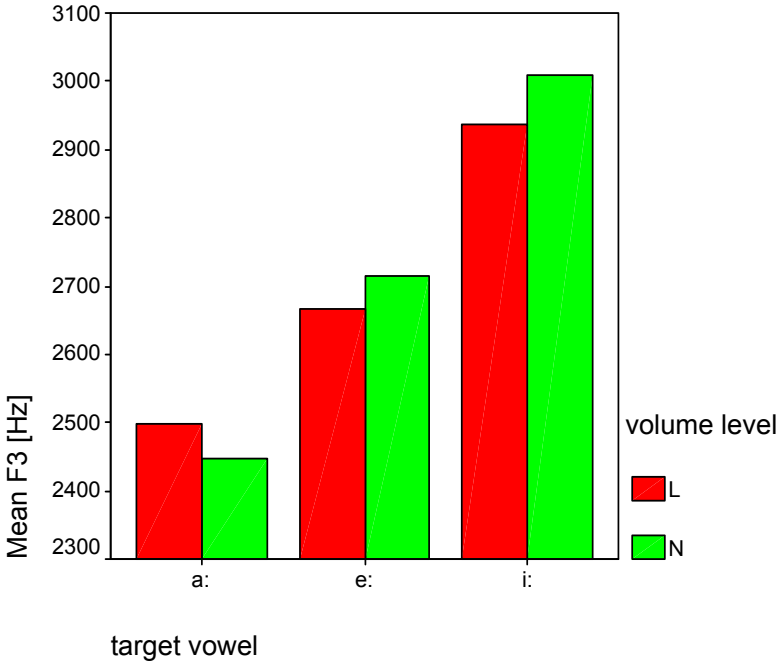


Figure 7-22. Effect of volume level on third formant. Data pooled for all speakers and V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F3. For /a:/ and /i:/ $p<0.001$, for /e:/ $p<0.01$.

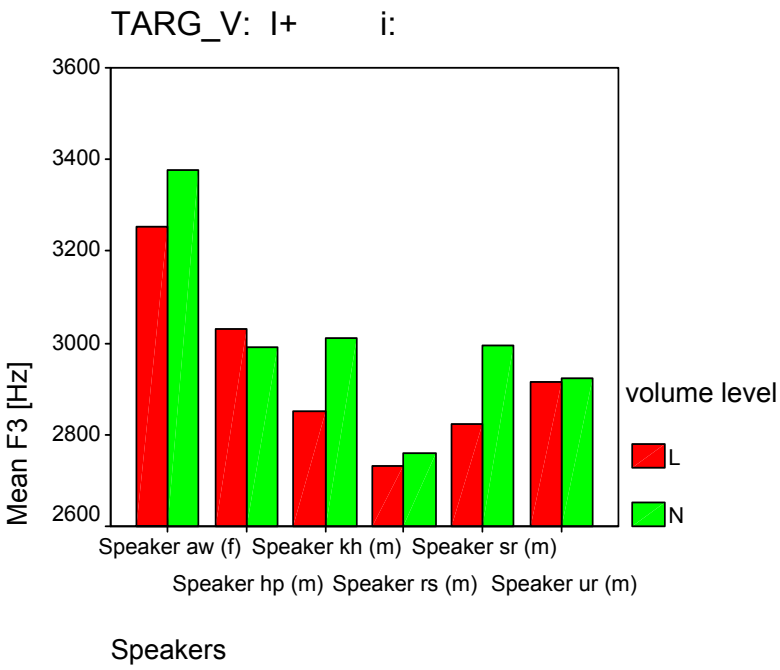


Figure 7-23a. Effect of volume level on third formant of /i:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F3 for speakers aw, kh, and sr ($p<0.001$), significant ($p<0.05$) effect for speaker hp, no significant effect for speakers rs and ur.

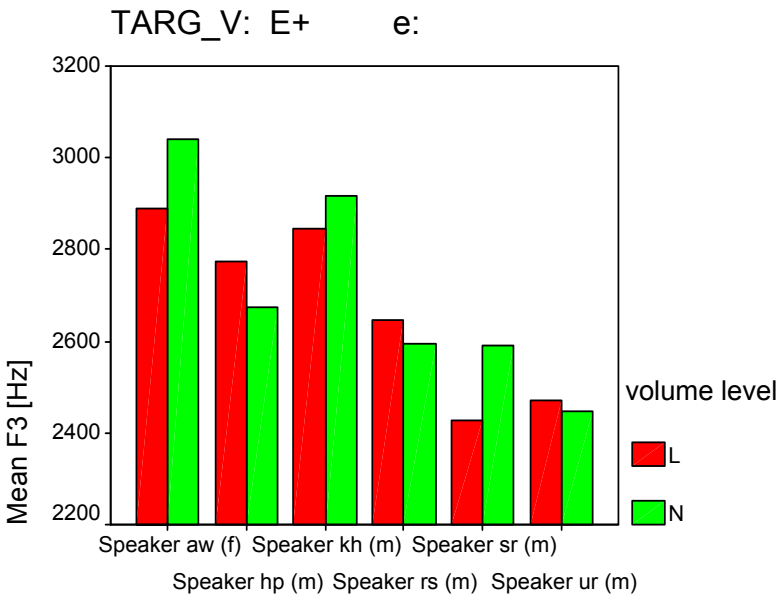


Figure 7-23b. Effect of volume level on third formant of /e:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F3 for speakers aw, hp, kh, sr ($p<0.001$) and rs ($p<0.01$), no significant effect for speaker ur.

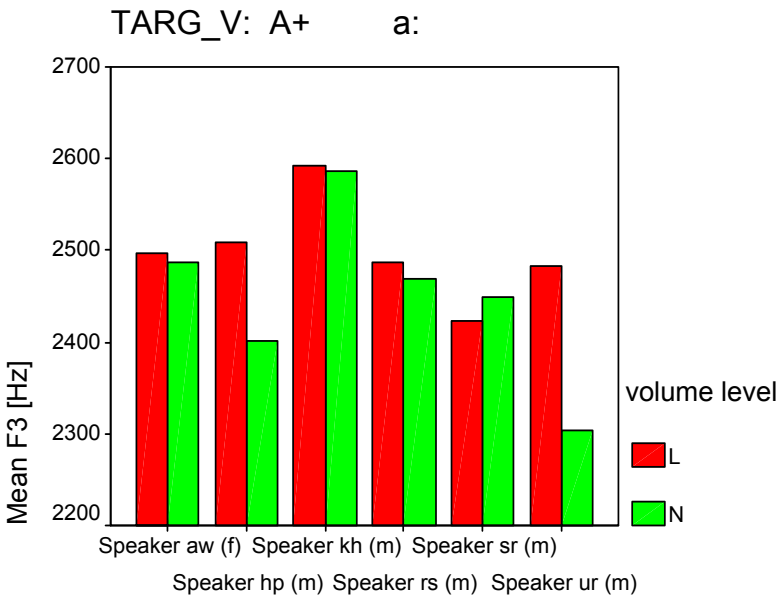


Figure 7-23c. Effect of volume level on third formant of /a:/. Data pooled for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F3 for speakers hp and ur ($p<0.001$), no significant effect for speakers aw, kh, rs, sr.

7.4.4 Formant charts in Hz and Bark

It has been already noted above that vowel perception is better described in a logarithmic Bark-scale (Syrdal and Gopal 1986). In the following figures formant patterns will be explored. Figure 7-24 presents scatterplots of all vowels for individual speakers. The formant values are given in bark and B2 is corrected by subtracting B1, and B1 corrected by the fundamental frequency in bark b0, as was suggested by Syrdal and Gopal (1986). The orientation of the axes is changed (by using the negative values of Bark values), as it seems to be more intuitive, using (corrected) first formant values along the vowel height dimension and (corrected) second formant values indicating the front-back dimension.

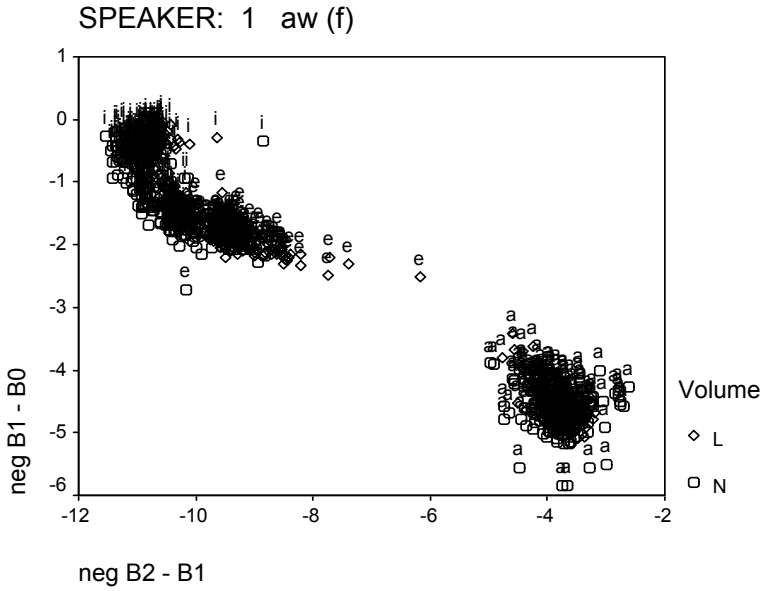


Figure 7-24a. Scatterplot of all vowels of speaker aw. Bark-scaled. Vowel qualities are marked with letters, volume level with different symbols.

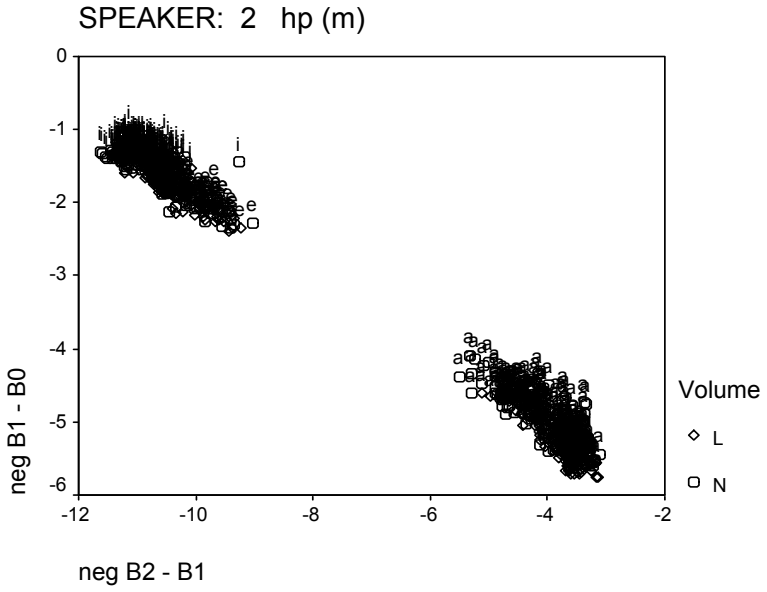


Figure 7-24b. Scatterplot of all vowels of speaker hp. Bark-scaled. Vowel qualities are marked with letters, volume level with different symbols.

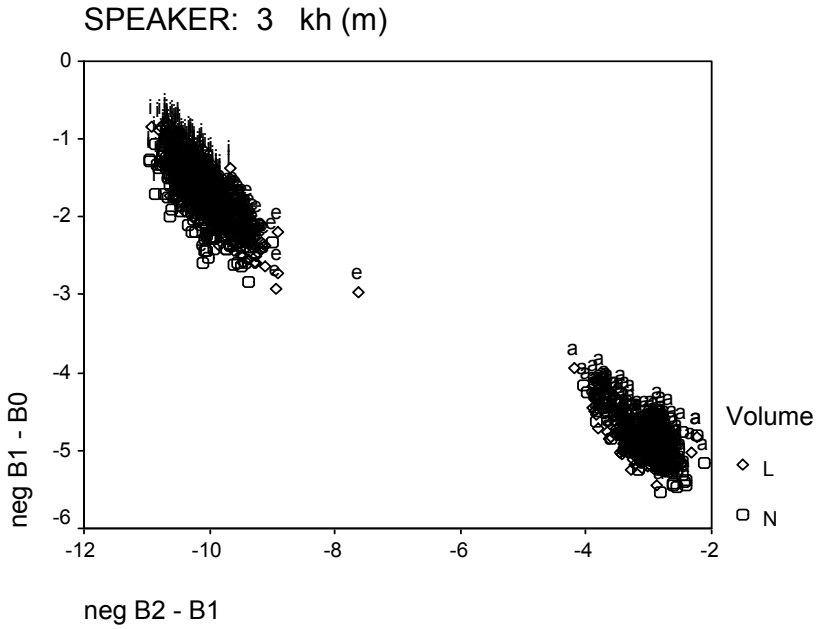


Figure 7-24c. Scatterplot of all vowels of speaker kh. Bark-scaled. Vowel qualities are marked with letters, volume level with different symbols.

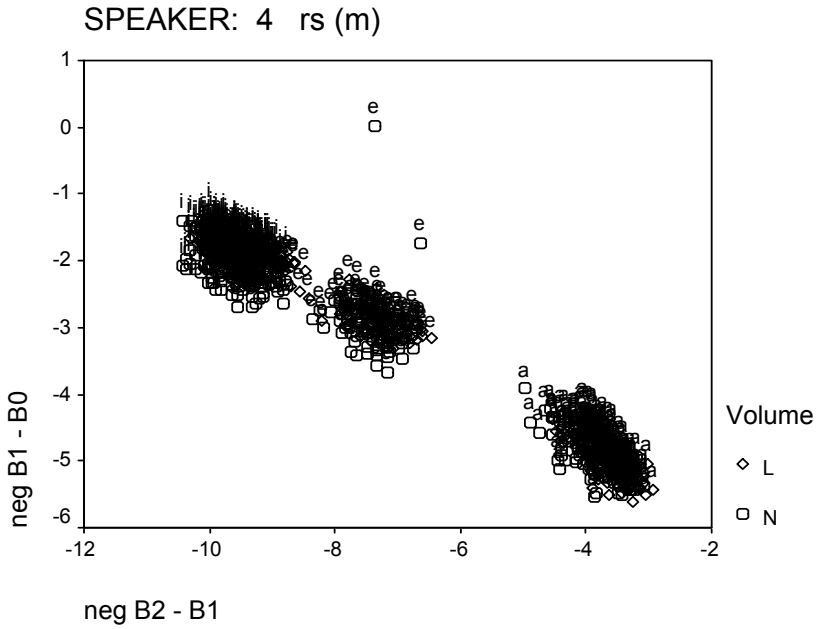


Figure 7-24d. Scatterplot of all vowels of speaker rs. Bark-scaled. Vowel qualities are marked with letters, volume level with different symbols.

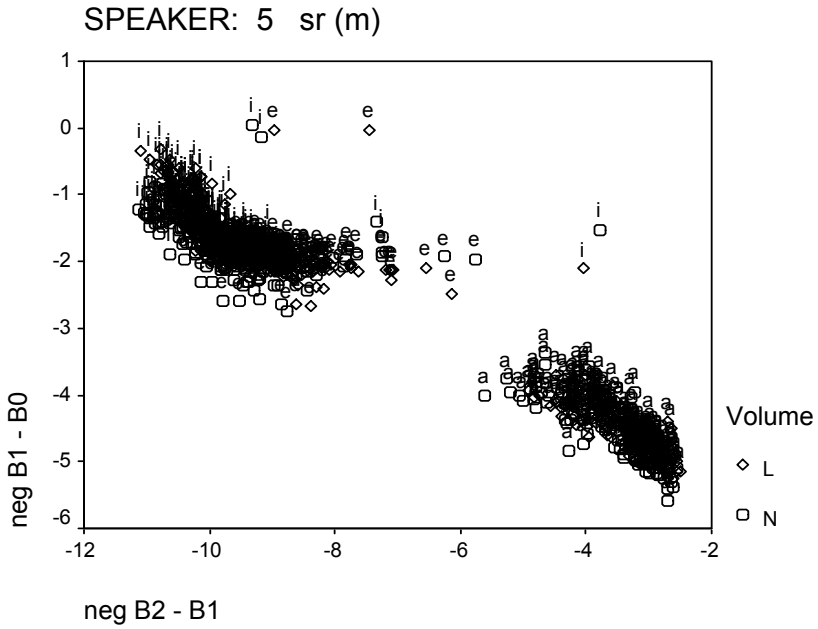


Figure 7-24e. Scatterplot of all vowels of speaker sr. Bark-scaled. Vowel qualities are marked with letters, volume level with different symbols.

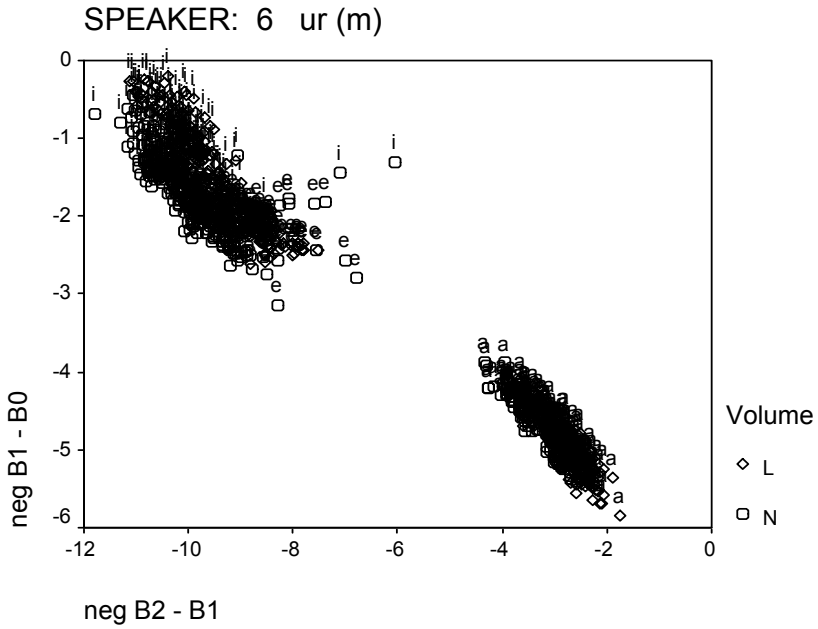


Figure 7-24f. Scatterplot of all vowels of speaker ur. Bark-scaled. Vowel qualities are marked with letters, volume level with different symbols.

In Figure 7-24 the two higher vowels overlap for most speakers to a certain extent, but it is somewhat difficult to decide, by which factors the overlap is caused.

In Figure 7-25 V1 vowels of speaker aw are plotted, separated for vowel quality, indicating volume level. Different symbols indicate different volume levels. These plots indicate that vowels cluster in a certain region, further front for normal volume and further bach for loud volume, for the two higher vowels. If one compares the scale values of the different plots one can assume that the overlap is only minimal for the two higher vowel qualities. To get a better survey only formant value means are presented in Figures 7-26 to 7-31.

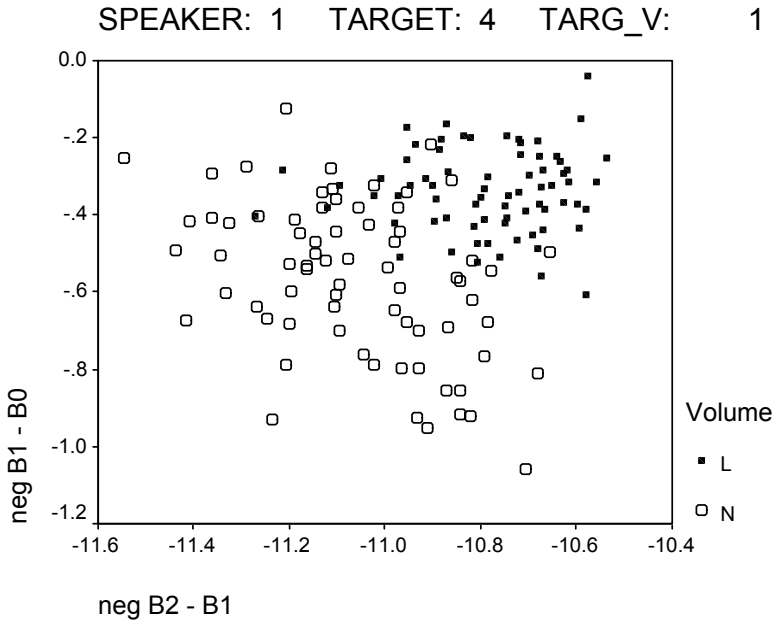


Figure 7-25a. Scatterplot for speaker aw, V1 /i:/. Bark-scaled. Volume level is marked with different symbols.

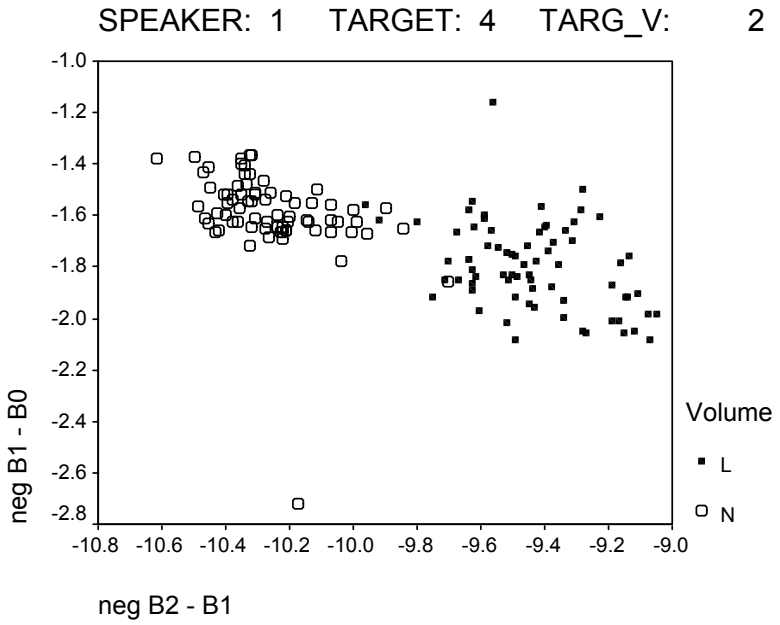


Figure 7-25b. Scatterplot for speaker aw, V1 /e:/. Bark-scaled. Volume level is marked with different symbols.

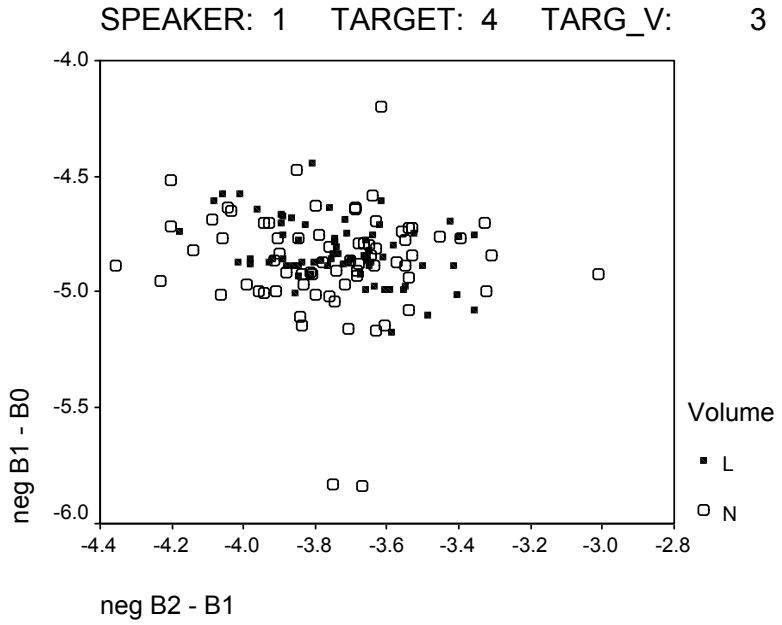


Figure 7-25c. Scatterplot for speaker aw, V1 /a:/. Bark-scaled. Volume level is marked with different symbols.

The formant plots in the Figure 7-26 to 7-31 present formant means for individual speakers. As said before, the orientation of the axes is changed (by using the negative values of Hertz and Bark values), as it seems to be more intuitive, using first formant values along the vowel height dimension and second formant values indicating the front-back dimension from left-to-right. In all following figures (a) represents Bark-scaled formant values with normalization for fundamental frequency (b1-b0). Figure (b) shows the same non-normalized formant values on a Hertz-scale. It can be stated that different vowel qualities appear clearer separated in Bark-scaled plots. We can further see that the effect of volume level differences is not responsible for some potential overlap. Higher dispersion as seen in 7-24 is more likely to be induced by differences between stressed V1 and less stressed V2 vowels.

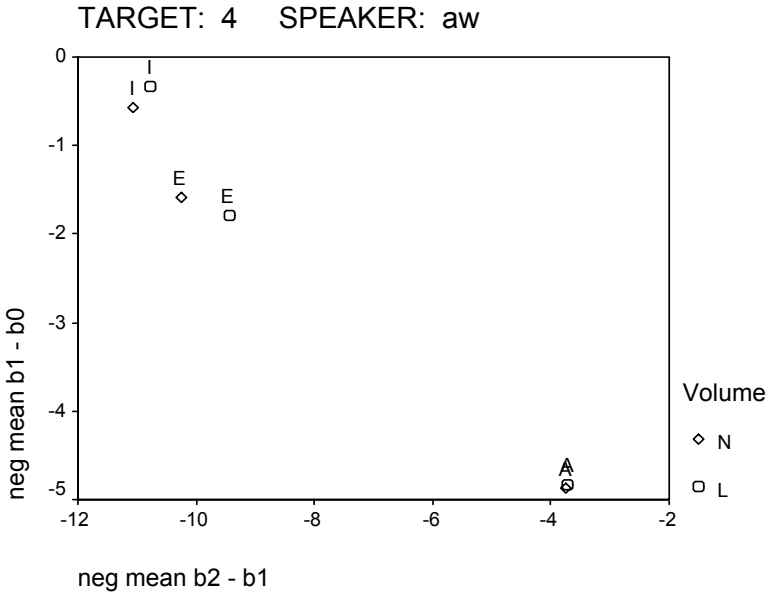


Figure 7-26a. Scatterplot of V1 means for speaker aw. Bark-scaled b1-b0 vs b2-b1. Volume level is marked with different symbols.

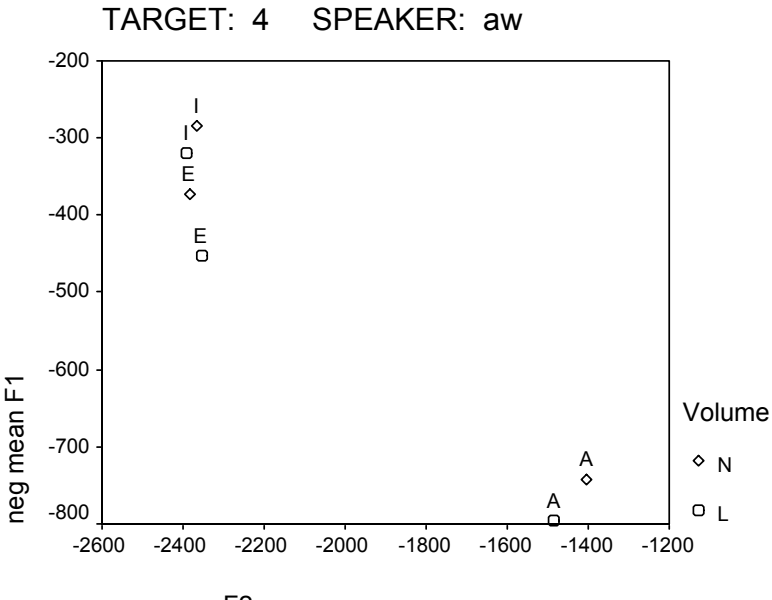


Figure 7-26b. Scatterplot of V1 means for speaker aw. Hertz-scaled F1 vs F2. Volume level is marked with different symbols.

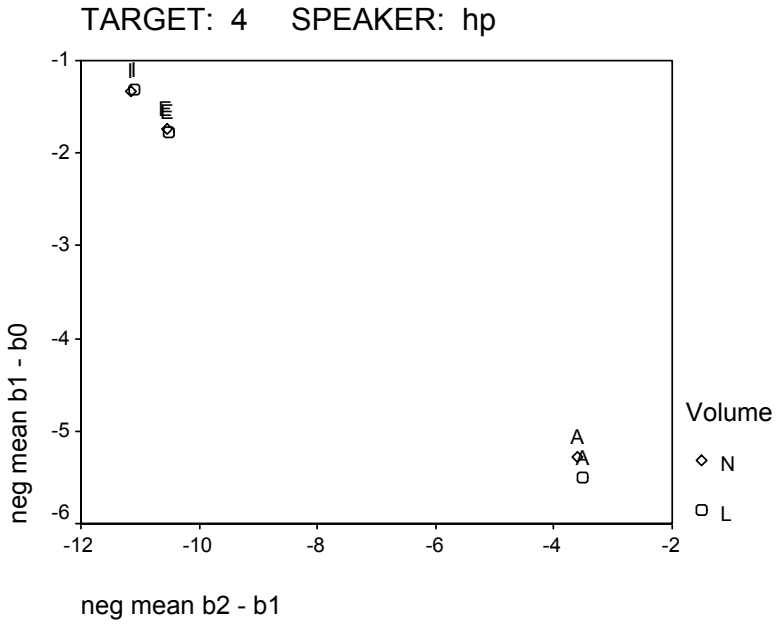


Figure 7-27a. Scatterplot of V1 means for speaker hp. Bark-scaled b1-b0 vs b2-b1. Volume level is marked with different symbols.

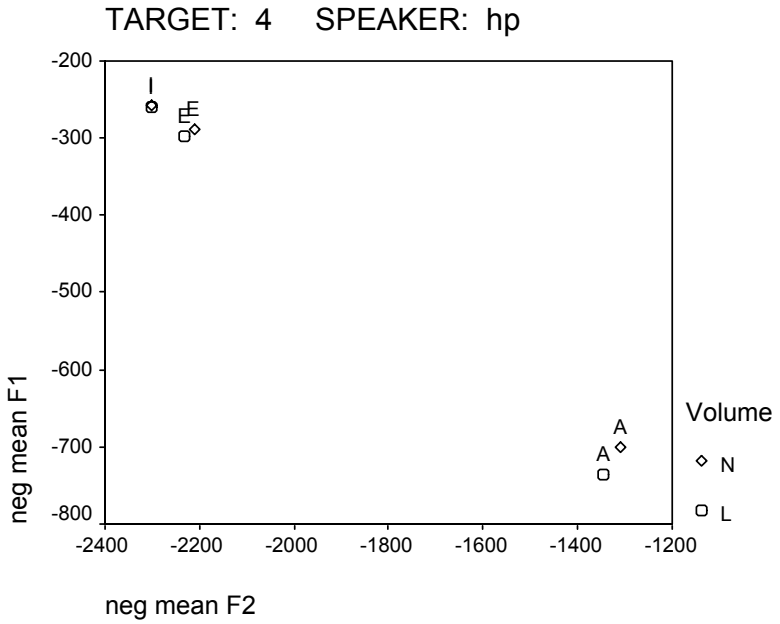


Figure 7-27b. Scatterplot of V1 means for speaker hp. Hertz-scaled F1 vs F2. Volume level is marked with different symbols.

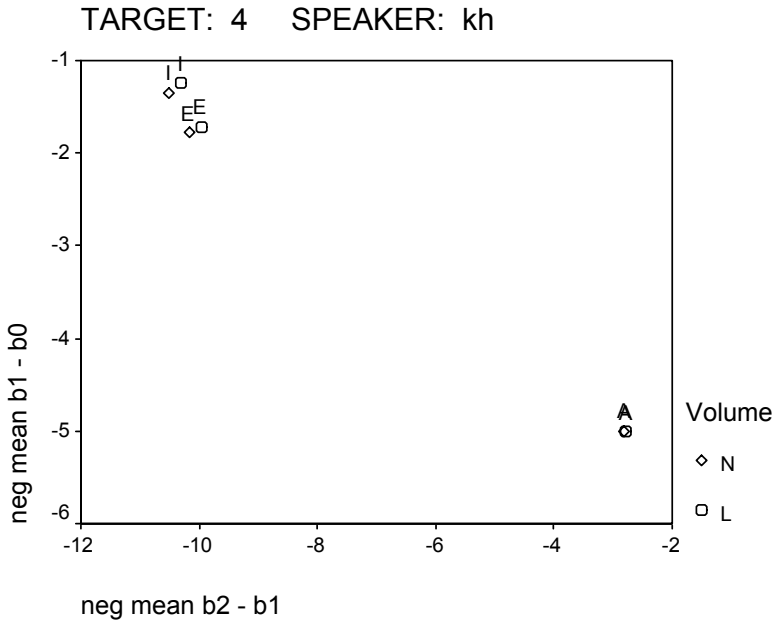


Figure 7-28a. Scatterplot of V1 means for speaker kh. Bark-scaled b1-b0 vs b2-b1. Volume level is marked with different symbols.

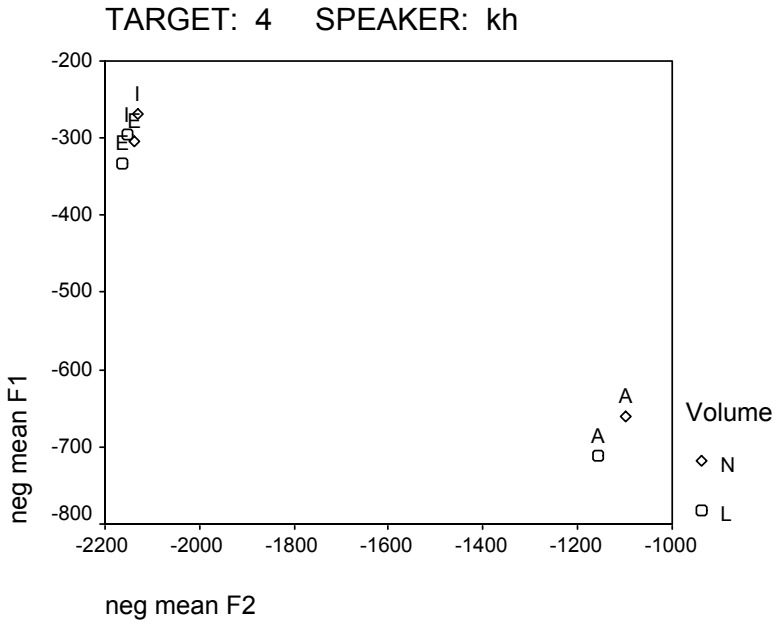


Figure 7-28b. Scatterplot of V1 means for speaker kh. Hertz-scaled F1 vs F2. Volume level is marked with different symbols.

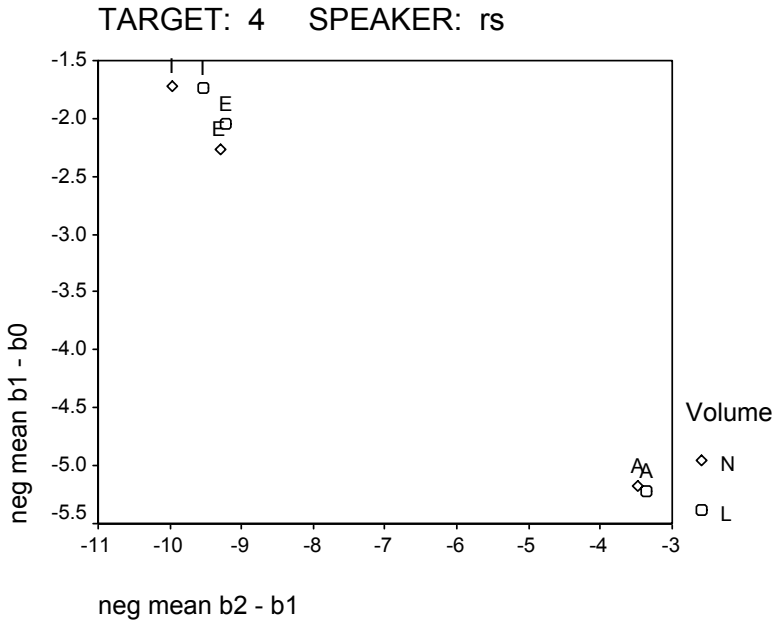


Figure 7-29a. Scatterplot of V1 means for speaker rs. Bark-scaled b1-b0 vs b2-b1. Volume level is marked with different symbols.

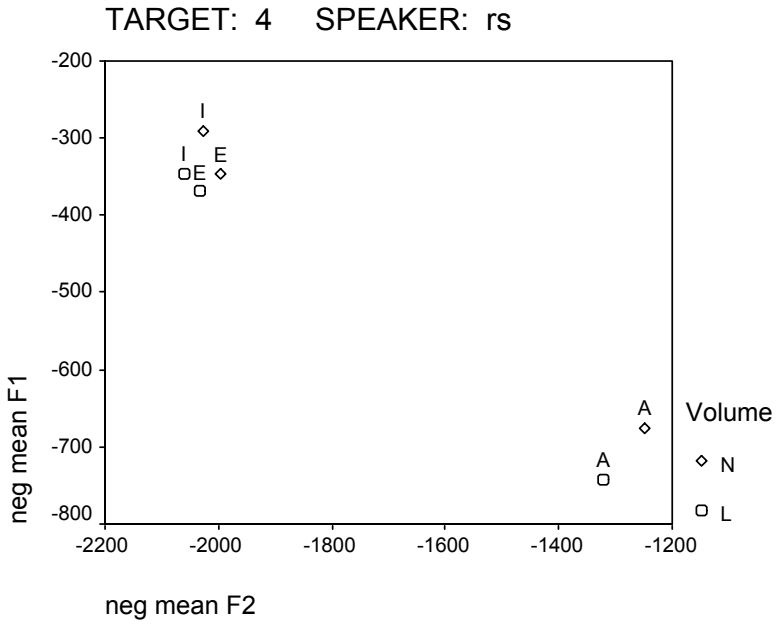


Figure 7-29b. Scatterplot of V1 means for speaker rs. Hertz-scaled F1 vs F2. Volume level is marked with different symbols.

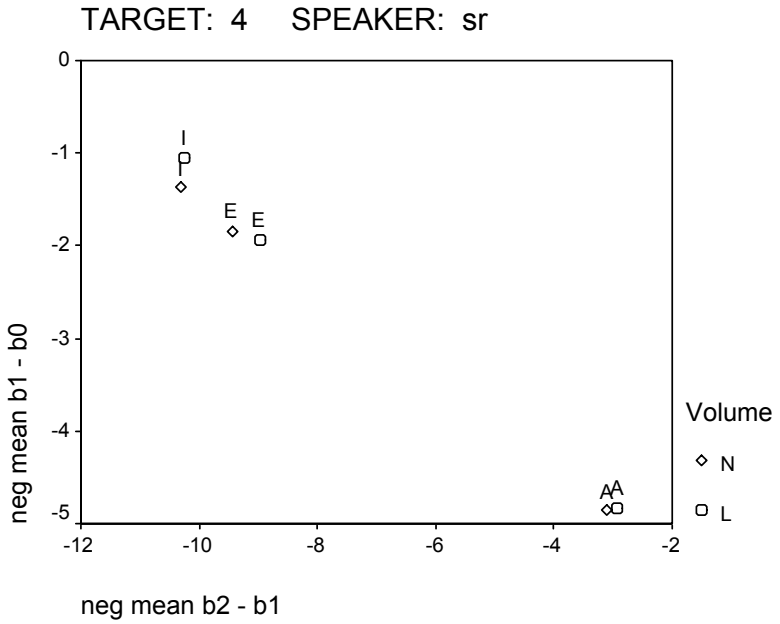


Figure 7-30a. Scatterplot of V1 means for speaker sr. Bark-scaled b1-b0 vs b2-b1. Volume level is marked with different symbols.

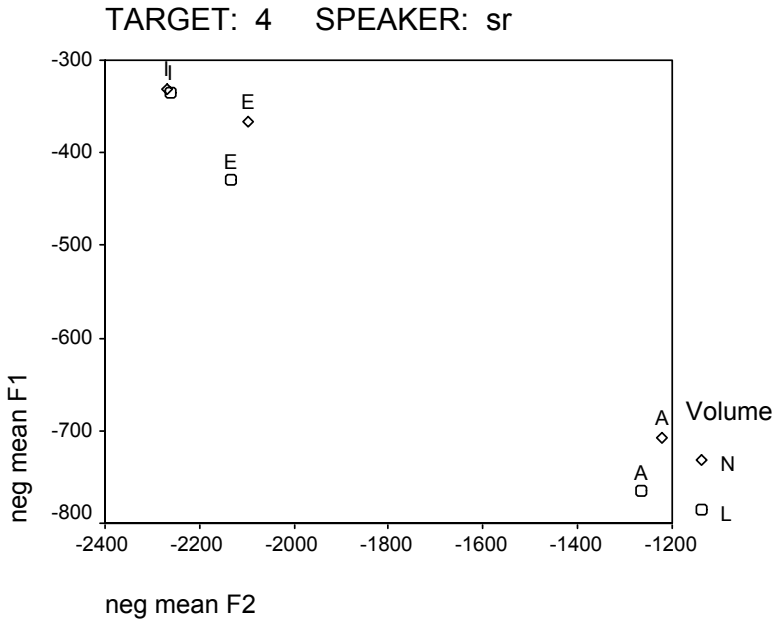


Figure 7-30b. Scatterplot of V1 means for speaker sr. Hertz-scaled F1 vs F2. Volume level is marked with different symbols.

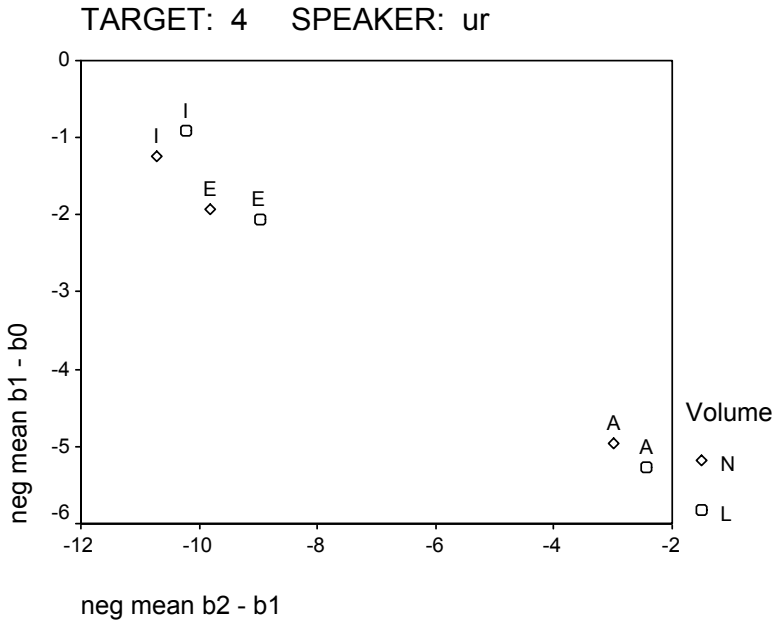


Figure 7-31a. Scatterplot of V1 means for speaker ur. Bark-scaled b1-b0 vs b2-b1. Volume level is marked with different symbols.

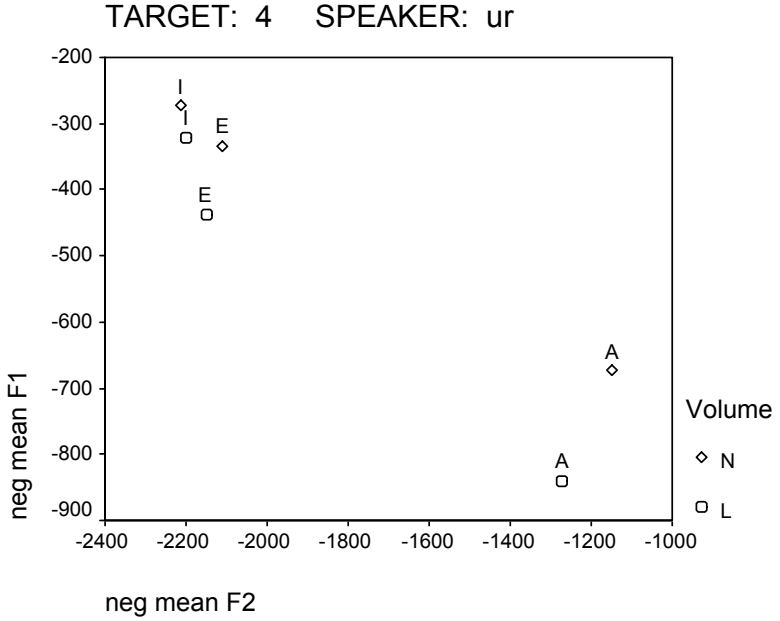


Figure 7-31b. Scatterplot of a means for speaker ur. Hertz-scaled F1 vs F2. Volume level is marked with different symbols.

In Figure 7-32 means of V1 and all speakers are plotted in one chart. While 7-32a might be a bit difficult to survey, different symbols are used in 7-32b and 7-32c to light up different factors. It emerges that overlap between /i:/ and /e:/ even across all speakers is not very high and that volume

level does not strongly influence mean deviation that would lead to categorial vowel confusion. The overlap in 7-32d on the Hertz scaled chart the overlap between F1 and F2 is much stronger.

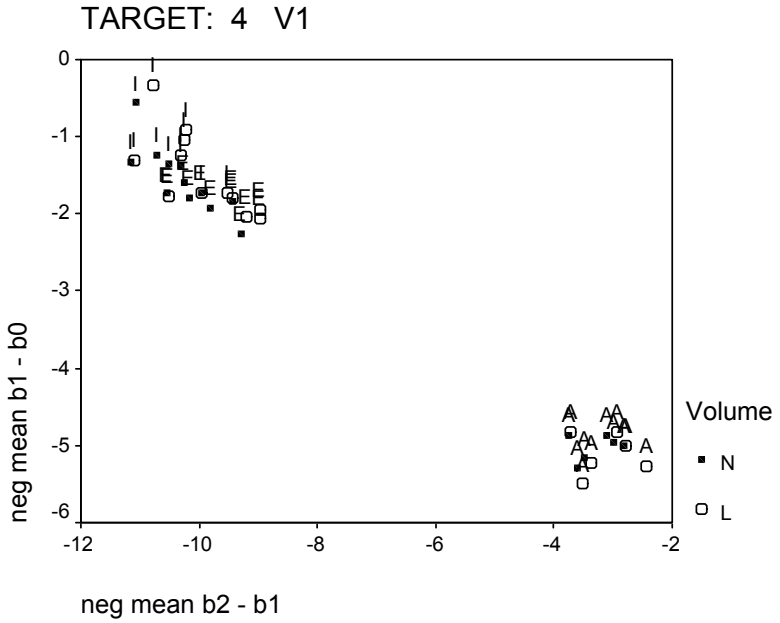


Figure 7-32a. Scatterplot of V1 means for all speakers. Bark-scaled b1-b0 vs b2-b1. Vowel quality is given in the figure with different letters, volume level is marked with different symbols.

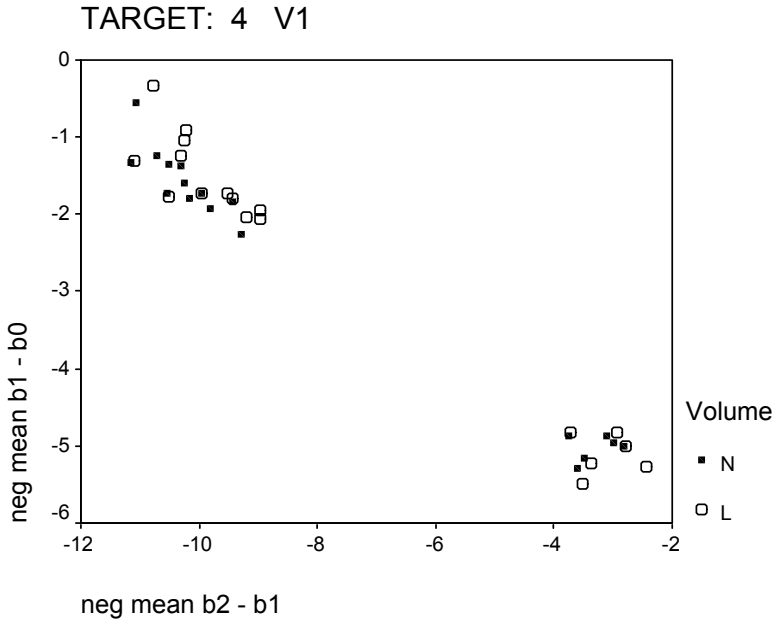


Figure 7-32b. Scatterplot of V1 means for all speakers and all vowel qualities. Bark-scaled b1-b0 vs b2-b1. Volume level is marked with different symbols.

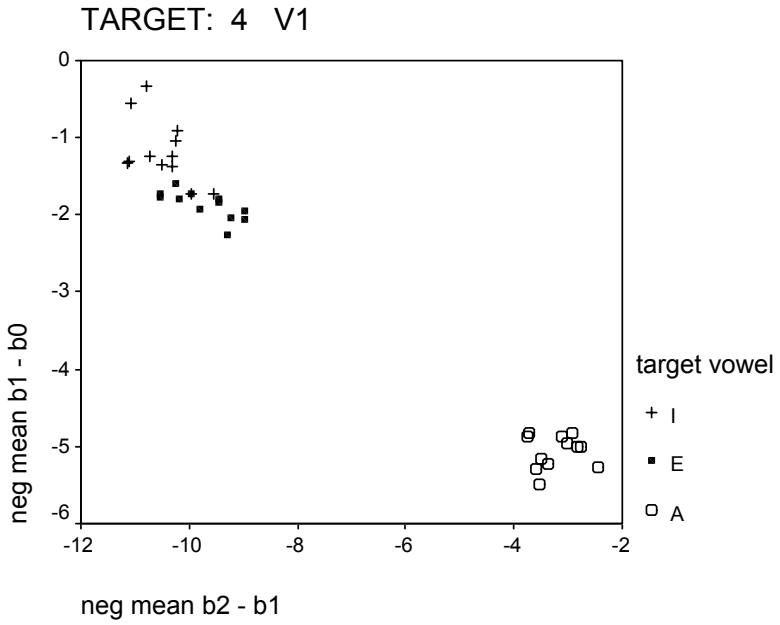


Figure 7-32c. Scatterplot of V1 means for all speakers and both volume levels. Bark-scaled b1-b0 vs b2-b1. Vowel quality is marked with different symbols.

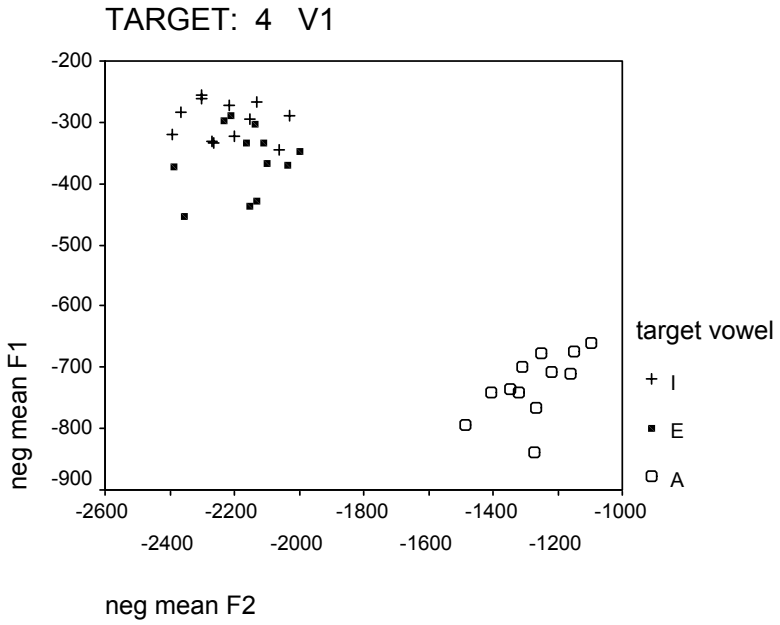


Figure 7-32d. Scatterplot of V1 means for all speakers and both volume levels. Hertz scaled F2 vs F1. Vowel quality is marked with different symbols.

7.4.5 Correlation of intensity and formants

Correlation of intensity with B1-b0 showed only a very weak correlation (see Figure 7-33) for individual vowel qualities and across all speakers. Pearson's correlation coefficient is for /i:/ -0.064; /e:/ 0.109; /a:/ 0.101. Correlation of intensity and B2-B1 (Figure 7-34) are only weakly

negative though highly significant. Pearson's correlation coefficient is for vowel /i:/ -0.164, /e:/ -0.310 and for /a:/ -0.257.

For intensity vs first formant in Hertz Pearson's correlation is for /i:/ 0.260; /e:/ 0.363; /a:/ 0.460. Correlations of intensity and second formant in Hertz (Figure 7-34) are not significant for vowel /i:/, Pearson correlation for /e:/ 0.114 and for /a:/ 0.135 (both highly significant). Correlations between intensity and formants within volume levels are even weaker. Correlation of intensity with second formant (Hertz) in loud voice condition is weakly negative at a highly significant level (Pearson correlation loud voice F2 vs intensity /i:/ -0.154; /e:/ -0.414, /a:/ -0.220).

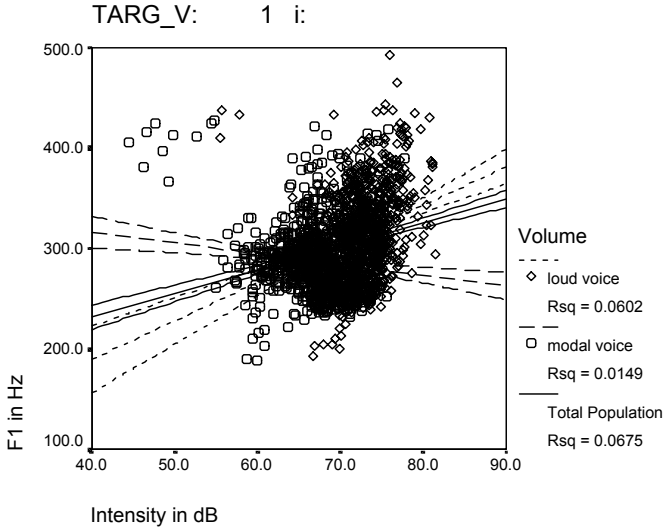


Figure 7-33a. Correlation of intensity and first formant (in Hz) for all speakers vowel /i:/. Volume level is marked with different symbols.

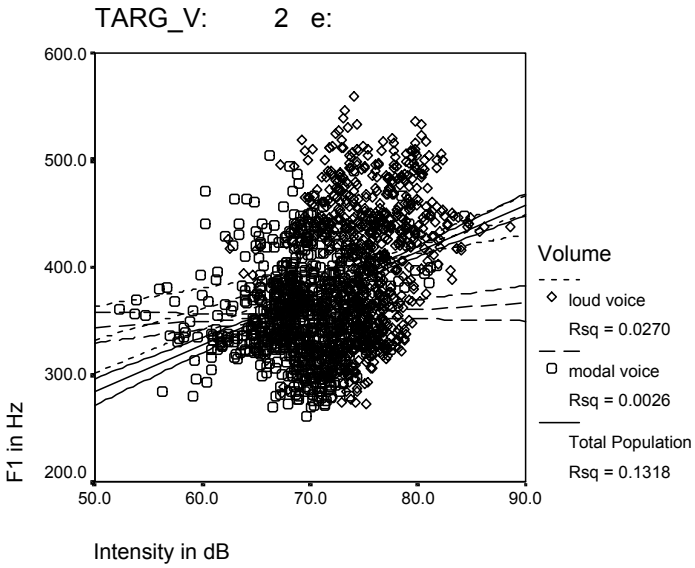


Figure 7-33b. Correlation of intensity and first formant (in Hz) for all speakers vowel /e:/. Volume level is marked with different symbols.

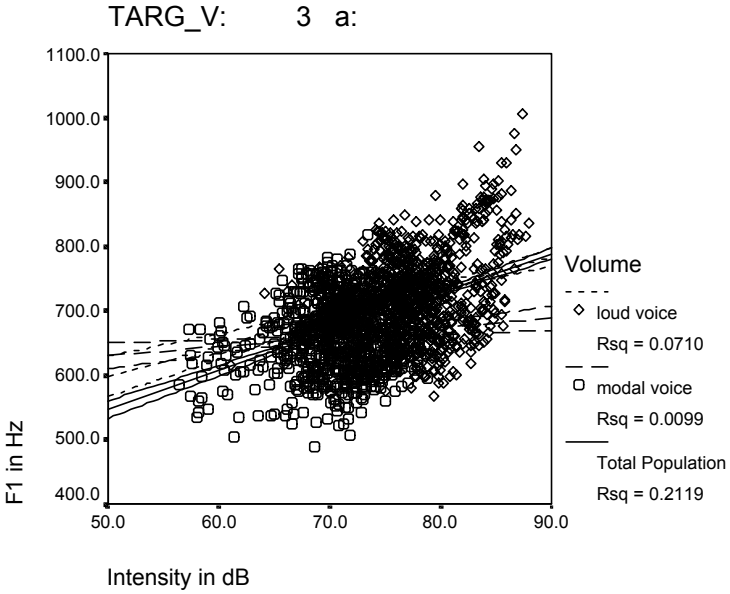


Figure 7-33c. Correlation of intensity and first formant (in Hz) for all speakers vowel /a:/. Volume level is marked with different symbols.

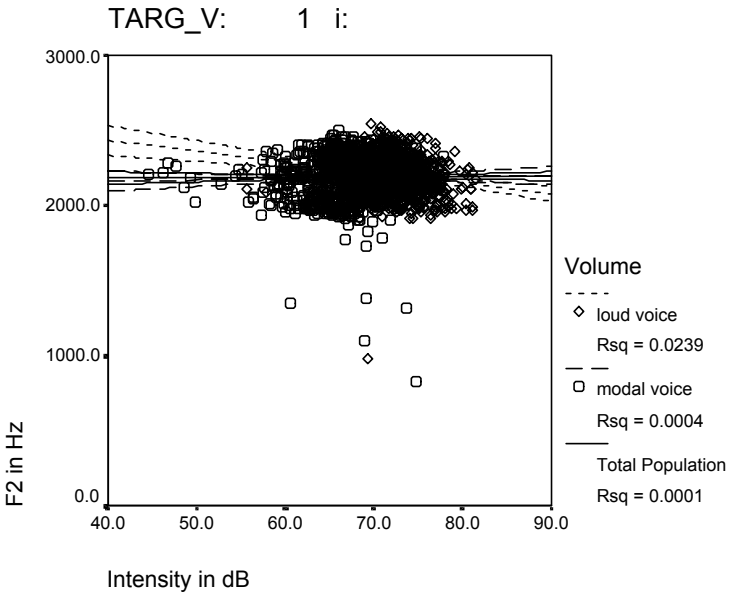


Figure 7-34a. Correlation of intensity and second formant (in Hz) for all speakers vowel /i:/. Volume level is marked with different symbols.

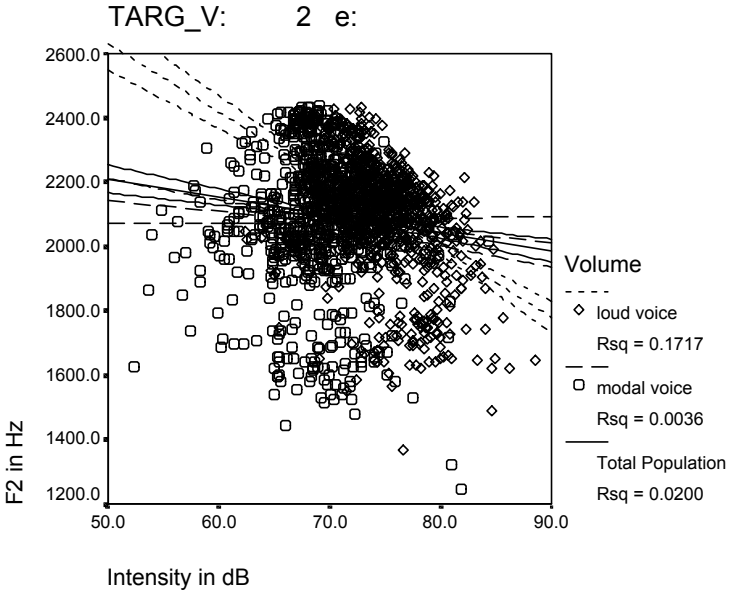


Figure 7-34b. Correlation of intensity and second formant (in Hz) for all speakers vowel /e:/. Volume level is marked with different symbols.

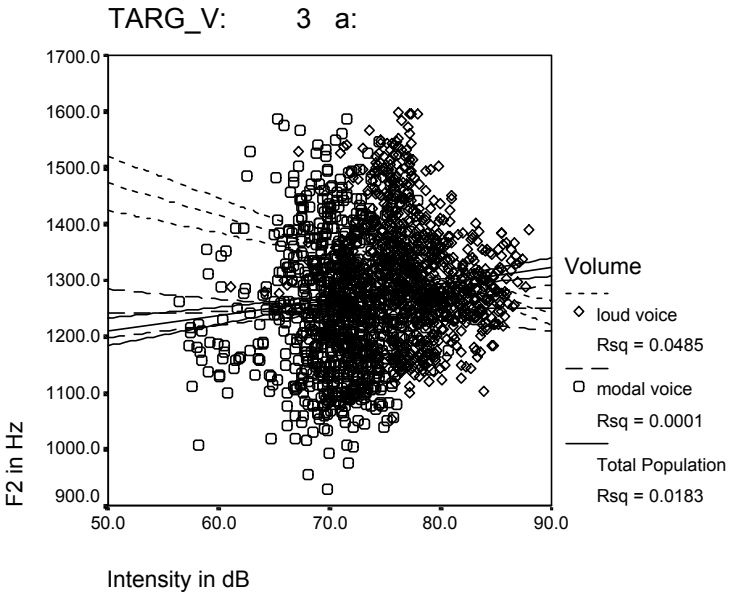


Figure 7-34c. Correlation of intensity and second formant (in Hz) for all speakers vowel /a:/. Volume level is marked with different symbols.

8. Discussion of Experimental Results

8.1 Discussion of articulatory data

8.1.1 Vowels

8.1.1.1 Jaw position in vowels

Jaw height in vowels varies as expected by categorical vowel height differences (Figure 6-19). A number of studies have shown that the vowel height is to a considerable extent produced by the jaw (e.g. Sussman et al. 1973, Johnson et al. 1993). Based on X-ray data Ladefoged discussed in (1972, 1990), that individual strategies for jaw and tongue interaction might be used, especially in the production of tense-lax oppositions. Unfortunately our data do not allow to comment on tense-lax production, since only three vowel categories have been analysed. However, we can state that jaw height differences even between /i:/ and /e:/ are very consistent for all our speakers (Figure 6-20). More tongue data have been analysed by Hoole and Kühnert (1996).

The jaw height of vowel V1 /a:/ and /i:/ is significantly influenced by the following consonantal context (see 6.5.1.1), the effect of perseverative C-to-V coarticulation is even stronger. For all vowel qualities jaw height in V2 is significantly influenced by the type of the preceding consonant.

Loud voice condition has a highly significant effect on jaw height in vowels (Table 6-2). For stressed and unstressed vowels the effect is similar. Only results for speaker hp in unstressed V2 are not highly significant ($p=0.022$). In all cases the jaw is lower in loud voice condition, as already observed by Schulman (1989).

8.1.1.2 Tongue - jaw interaction in vowels

In this study the position of the tongue tip reflects tongue height for vowels as well as for consonants. We have already discussed that the tongue tip in traditional descriptions is certainly not regarded as a central point of vowel articulation. However, studies of vowel articulation (e.g. Johnson et al. 1993, Ladefoged/Maddieson 1996, Hoole et al. 2000) show that vowel articulation can be interpreted as a general tongue body shaping. Even the tongue tip, then, is centrally involved in vowel articulation. The only example of apparently deviating tongue tip height from expected tongue height is found for speaker aw, where the tongue tip is higher in V2 /e:/ than in V2 /i:/. In all other speakers and vowels the tongue tip reflects expected tongue height patterns.

The volume level differences for tongue tip height are significant in V1 and V2 for most speakers (V1 speaker sr and V2 speaker aw not significant). Yet, intrinsic tongue tip height (over all vowel qualities) is not significantly influenced by volume level differences (exception: speaker kh in V1 shows a highly significant effect ($p<0.01$)).

Patterns of compensatory behavior for intrinsic tongue tip and jaw could be observed only for vowel /a:/, but not for vowels /e:/ and /i:/ (Figures 6-10 and 6-11). Since we have argued above that tongue tip might be involved in all vowel articulation this is a rather surprising finding. Consideration of different muscles involved in vowel articulation results in a better explanation. The tongue configuration for front high vowels needs activity of the m. genioglossus, which

connects the back part of the tongue body with the symphysis of the mandible (Borden/Harris 1980: 101, Gray 1901). If m. genioglossus contracts, the back part of the tongue is pulled down and frontward, while the front part of the tongue rises more passively. Insofar one can state some direct muscular coupling between lower jaw and tongue during /i:/ and /e:/ production. Production of a non-front vowel /ɑ/ (Borden/Harris 1980: 102) involves activity of m. hyoglossus (connected to the hyoid bone), and lowers the whole tongue body. These different muscles involved in different vowel production can explain why especially front vowels show up with strong jaw articulation and on the other hand why no patterns of compensatory articulation for front vowels can be found in our data.

8.1.2 Consonants

8.1.2.1 Jaw position in consonants

Jaw height in consonants is strongly influenced by the consonant type. The mean values of consonantal jaw height plotted in Figure 6-17ff show that consonantal jaw height range can be of about the same size as jaw height differences between the fairly high /e:/ and /a:/ (approx. 8mm). Higher jaw positions are found in /s/, /ʃ/, and /t/. Jaw position for /d/ is lower and lowest for /n/ and /l/.

Volume level has no significant effect on jaw height for pooled consonantal data or pooled subjects. A few significant effects of volume can be obtained, for individual speakers, and separate consonant types. They are reported here in Table 8-1, but it is very difficult to see any pattern besides a volume level effect for speaker ur.

p<0.05 = * p<0.001 = ** not significant = n.s.	/ʃ/	/d/	/l/	/n/	/s/	/t/
aw	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
hp	n.s.	n.s.	n.s.	n.s.	*	n.s.
kh	n.s.	n.s.	n.s.	**	n.s.	n.s.
rs	n.s.	n.s.	n.s.	n.s.	**	n.s.
sr	*	*	n.s.	*	n.s.	n.s.
ur	**	**	**	**	*	**

Table 8-1. Volume effect on jaw height in different consonants for individual speakers. Independent samples t test levels of significance are reported.

The coarticulatory effect of vowel context on consonantal jaw height depends clearly on the consonant type and was found for data pooled for speakers. Jaw height in /s/, /ʃ/, and /t/ is not affected by the jaw height of the vowel context. For /d/ the effect is significant, for /n/ and /l/ even highly significant. The coarticulatory effect on different consonants seems to be related to their intrinsic jaw height. Yet, it would be expected that jaw position for /s/, /ʃ/, and /t/, should be more strongly affected by the vowel context, since the height differences between vowels and consonants are larger. These data suggest that in general the jaw is a relatively precise articulator for the alveolar consonants, with special precision in the articulation of /s/, /ʃ/, and /t/

(small standard deviations). Similar results were obtained for other languages. Keating et al. (1989) reports consonantal jaw heights for Swedish and American English speakers. Lee (1994) presents data on consonantal jaw heights for Arabic, French, and Korean speakers. Very high jaw positions for /s/ (and /ʃ/ only reported for Arabic and French) were consistently found for all speakers and all 3 languages. High jaw positions for /t/, higher than for /p/ or /k/, were also consistently found. The case for the sibilants /s/ and /ʃ/ will be discussed in more detail further below. The stop /t/ might require a higher jaw position for a complete oral closure, behind which a relatively high air pressure arises. The /d/ (in German not necessarily (fully) voiced) in our data showed very short acoustic durations, and was for some speakers (especially hp) almost tap like. Differences in jaw height between /t/ and /d/ might result from some sort of target overshoot for the /t/ or some undershoot for the /d/. This is supported by the fact, that speakers with higher acoustic consonant duration (Figure 7-6a) e.g. aw, kh, and ur show higher jaw positions in /d/ (Figure 6-18), while speakers rs and sr have shorter consonant durations and a lower jaw position for /d/. Lower jaw position for velar stops (as reported by Keating et al. 1994 and Lee 1994) might have to do with a potentially higher influence of the jaw on more retracted tongue articulations.

Relatively low jaw positions for /n/ could be explained as the missing necessity to build up a very robust oral occlusion. The air pressure behind the closure will be negligible, since air escapes through the nasal cavity. A similar explanation may hold for /l/, the air escapes laterally, the medial oral closure has not to be very robust. Some introspective inquiry even revealed that a very high jaw position (up to dental occlusion) would produce lateral friction. So we can also interpret the lower jaw position for /l/ as intentional.

8.1.2.2 Tongue - jaw interaction in consonants

All consonants in our data require a complete oral closure in the midsagittal plane /t, d, n, l/ or very narrow constriction, produced by the tongue tip or slightly more retracted for /ʃ/. So, different jaw heights are expected to be compensated by higher intrinsic tongue movements. We have already seen that jaw height for consonants is rather precise, and not much largely affected by volume level. The overall pattern of covariation in Figure 6-9 is mainly due to the fact that jaw heights for consonants differ. A more detailed analysis of compensatory behavior for individual speakers and different consonant types shows (see chapt. 6.5) that there are further compensatory patterns. They differ somewhat for different speakers, but the most consistent interesting finding is the missing correlation for /t/, which can of course be caused by the low jaw variability for /t/. An alternative explanation is that as mentioned above /t/ needs a relatively robust oral closure. The wall of the palate is to some extent soft tissue that can be compressed by the tongue. Even a slightly lower tongue position would then still guarantee a full closure, and no compensatory higher tongue position is required.

8.1.3 Articulatory effort

A simple comparison of vertical positions of articulators allows only limited conclusions about intentional movement. In recent years in the literature there have been developed some ways to describe movements either as some kinematically defined points (peaks of velocity, acceleration or third derivative of the movement (jerk)). A further approach involves curve fitting based on damped spring-mass models (Saltzman 1989, Browman/Goldstein 1986ff). Since we did not find highly unpredictable patterns for equilibrium points in our data, we used a cosine model to

determine the articulatory effort for the consonantal articulation. Damped spring-mass models result in very similar sinusoidal movements, and are thus comparable. Figuratively speaking, we calculate the area below an assumed cosine-like movement, whose amplitude is defined by the difference. Downward movement can be accounted for if we take the absolute value.

It was already pointed out that mass differences between jaw and tongue tip are neglected. In the literature, articulation of coronal sounds is often described as relatively easy or fast since only the tongue tip would have to be involved. We have shown that in fact the jaw plays an important factor in articulation of most coronal sounds. If we then assume that jaw movement further involves displacement of a greater mass then any tongue movement will do, the results for relatively high work of the jaw are even more interesting. Instead of involving a relatively low effort in their articulation, coronals and especially /s/ and /ʃ/ seem to require a relatively high articulatory effort.

8.2 Discussion of acoustic data

8.2.1 Intensity

It is not surprising that higher volume level showed for all segment types a higher intensity. Yet, perceptual experiments have shown that the listener can reconstruct or estimate the speaker's loudness independent of the perceived sound pressure level (Wilkens/Bartel 1977, Eriksson/Traunmüller 1999). This effect can be noted e.g. in radio plays. Intensity differences between separate volume levels vary for individual speakers. Especially speaker ur showed very large intensity differences between the two volume levels. For the other speakers it appears that they produce generally similar intensities for loud voice, but rather differing volume levels for modal voice.

8.2.2 Duration

Clear durational effects in loud speech are found for vowels and consonants. Over all speakers all vowels showed a clear lengthening effect (Figure 7-5b). All consonant types except /t/ showed a significantly shorter duration in loud voice condition. Similar results were obtained by Bonnot/Chevrie-Muller (1991) for French data. Traunmüller and Eriksson (2000) get higher vowel durations but no shorter consonant durations for Swedish data. However, as they note, this might have to do with the fact that Swedish has a phonological lengthening effect in consonants (/V:C/ vs. /VC:/).

Furthermore, a longer sentence duration for loud volume level was significant for five speakers. This could not easily be explained as a side effect of longer vowel duration, since the sentence consists of 12 vowel phonemes (4 long) and 20 consonants. We can here only point to this result, and focus on segmental data.

Correlations between intensity and duration of vowels are positive for pooled data. Correlations within different volume levels are in general rather small. For unstressed vowels we can even find some negative correlations of intensity and duration within separate volume levels.

8.2.3 Vowel parameters

8.2.3.1 Fundamental frequency

In loud speech the vowel fundamental frequency is considerably higher for all speakers. Intrinsic fundamental frequency differences are kept. Speaker ur showed especially high f0

differences. Large f0 effects have been noted in the literature (Rostolland 1982a) for shouted as opposed to loud voice or were observed for speaking in noise (Summers et al. 1988). Interestingly Holmberg et al. (1988) found no f0 effects of such an extent. They asked their subjects to adjust their voice level to comfortable loud and normal conditions. This method is quite similar to ours. Based on the intensity data, fundamental frequency data and the auditory impression we can conclude for speaker ur that he came close to shouting in the loud condition. Traunmüller and Eriksson (2000) found very clear fundamental frequency changes for higher vocal effort resulting from larger speaker - listener distance. A very rough comparison of our f0 means with those would estimate an (assumed) speaker - listener distance in between 7.5 and 38.5 meters (around 25 to 30 meters) for our data.

8.2.3.2 Formants

Analyses of the first formant showed highly significant higher values for loud speech for all vowels and all speakers. This effect has been reported as well by Rostolland (1982b) and is summarized by Schulman (1989) for other earlier investigations.

An effect for the second formant was consistently found especially for /a:/, that is assumed to be produced with a fairly central quality. But even for the front vowel /i:/ across speakers a significantly higher second formant for loud speech was found. Somewhat in contradiction to our findings Lindblom/Sundberg (1971) found that jaw opening causes lowering of F2 for /ɑ:/, but F2 raising for /u/.

Third formant values showed some centralizing effect for loud voice condition toward a region above 2500 Hz.

The effects of loud voice on formants that we found can be summarized in articulatory terms as a clear lowering (raised F1), and some fronting (raised F2) especially of the non-front vowel /a:/.

As an explanation for the higher first formant in loud speech one could assume the need for keeping an appropriate distance to the simultaneously raised fundamental frequency. Since fundamental frequency is increased in loud speech, the first formant of high vowels might collapse with f0. However this is as well the case in normal speech and with a high fundamental frequency e.g. of female voices. A more plausible explanation is that vowel height quality is perceived similar to a difference F1-f0 on a Bark-scale (Syrdal/Gopal 1986). A higher fundamental frequency would then influence the perception of vowel height. Formant plots in Figures 7-26ff support this.

8.3 Articulatory - acoustic relations

8.3.1 Jaw height - first formant

We have discussed above that a higher first formant might be a perceptual need to compensate for the simultaneously increased fundamental frequency.

We find on the other hand a clear pattern of more open vowel quality in loud speech as well as jaw lowering. We find an articulatory aerodynamic explanation for lower first formant intriguing that was proposed by Schulman (1989). Increased intraoral pressure in loud speech but constant degree of high vowel constriction could result in friction. This could lead to a categorial change /i/ to /j/, /j̟/, or at least to not acceptable vowel quality. To compensate for the

heightened intraoral pressure the constriction is lowered for high vowels, other vowel qualities, then, are subsequently lowered.

8.3.2 Jaw targets for different consonants

Shadle (1985, 1990) demonstrated with mechanical models of the vocal tract which were aerodynamically excited that two different manners of noise generation underlie production of fricatives

- „wall source" - Noise produced at the place of the critical constriction, as for /ç/ and /x/.
- „obstacle source" - Here it is important that the obstacle is in close proximity to the critical constriction forming a jet, demonstrated for /s/ and /ʃ/. The obstacle then is formed by the incisors.

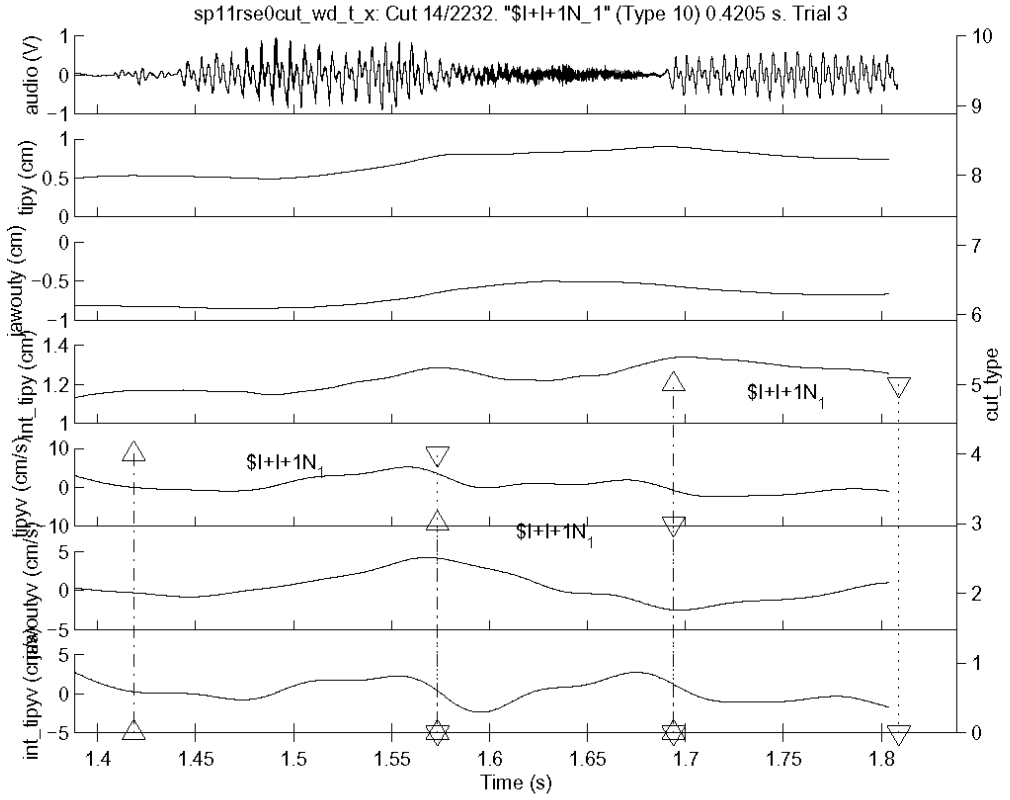


Figure 8-1. An example for tongue-jaw interaction in /i:ʃi:/ production of speaker sr. An intrinsic tip lowering for /ʃ/ can be observed in order to produce a higher jaw position. The three upper kinematic panels show tongue tip, jaw, intrinsic tip vertical movement, the lower three panels tongue tip, jaw, intrinsic tip velocity.

This was already implicated in the description of strident fricatives given by Jakobson, Fant, and Halle: „A supplementary barrier that offers greater resistance to the air stream is necessary in the case of the stridents. Thus beside the lips, which constitute the sole impediment in the production of the bi-labials, the labiodentals involve also the teeth. In addition to the obstacles

utilized in the corresponding mellow consonants, the sibilants employ the lower teeth and the uvulars, the uvula." (Jakobson/Fant/Halle 1952:24) In later descriptions the feature [strident] is described in more acoustical terms. „Strident sounds are marked acoustically by greater noisiness than their nonstrident counterparts" (Chomsky/Halle 1968:329). Since then, in many different works it was argued for or against [strident] being a feature only relevant for coronal sounds. For a survey of the relevant articles see Hall (1997:141f).

We can then differentiate the role of the jaw in /s/, /ʃ/ as opposed to /t/ (and to a lesser extent /d/). For /t/ as well as for the fricatives we found a very high jaw position. It seems to be a plausible explanation to assume that the jaw is needed as a supporting structure to ensure a robust oral closure. Tongue and jaw act as coordinated structures for stop production, that need to interact.

For /s/ the jaw as carrier of the lower incisors acts as an articulator in its own right. We find this supported by the fact that the jaw generally shows a larger portion of work for /s/ (and /ʃ/) than for /t/ (see Figure 6-15). In Figures 8-2 and 8-3 the production of fricative and stop are presented with one example each, both for speaker rs. They demonstrate our general interpretation for jaw activity of stops as supporting structure, and for jaw activity in fricatives as independent articulator, which in some cases even has to be compensated for, as shown in Figure 8-1.

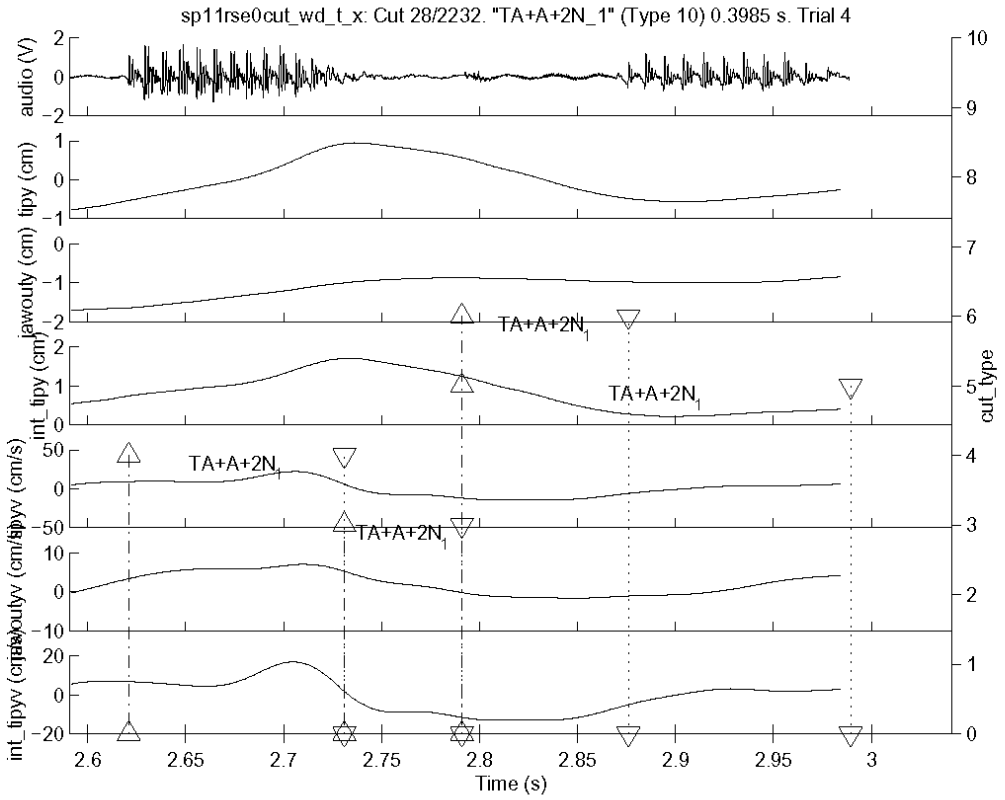


Figure 8-2. An example for tongue-jaw interaction in /a:ta:/ production of speaker rs. The three upper kinematic panels

show tongue tip, jaw, intrinsic tip vertical movement, the lower three panels tongue tip, jaw, intrinsic tip velocity.

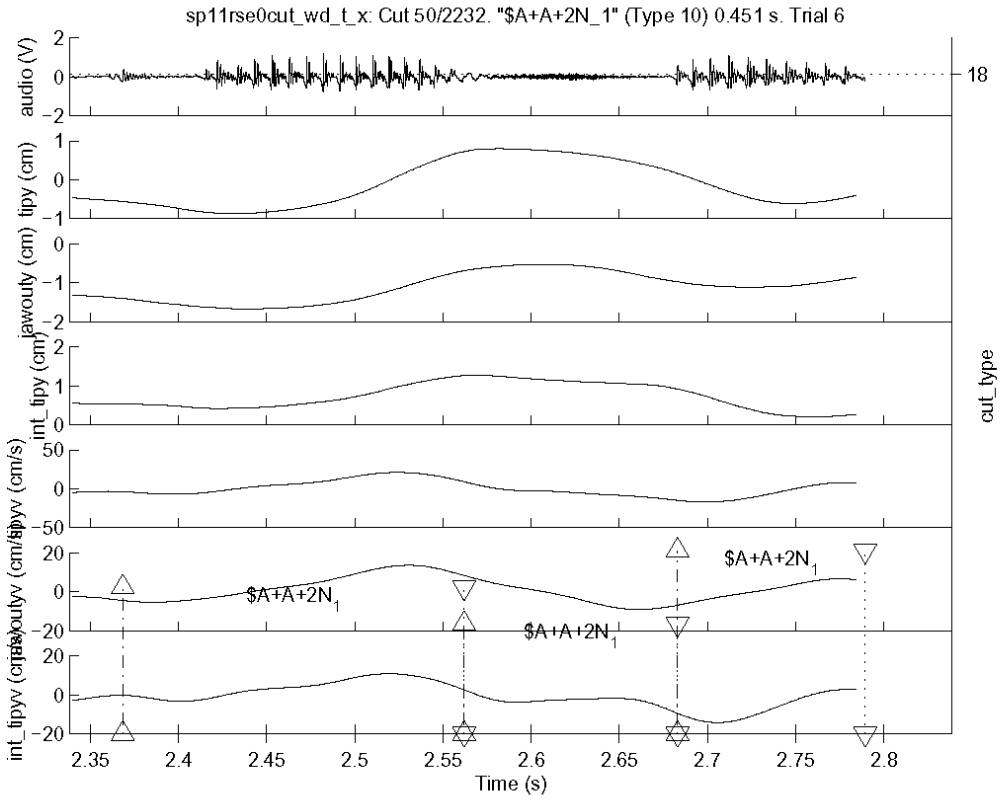


Figure 8-3. An example for tongue-jaw interaction in /a:ʃa:/ production of speaker rs. The three upper kinematic panels show tongue tip, jaw, intrinsic tip vertical movement, the lower three panels tongue tip, jaw, intrinsic tip velocity.

8.4 Speaker specific jaw activity

It was already mentioned that for consonant production general jaw activity was especially large for speaker sr and especially small for speaker ur. Based on Ladefoged et al. (1971), Johnson et al. (1993) discussed if differences for vowel production in jaw activity might have to do with back palate vault shape. Ladefoged et al. (1971) argued that a flat back part of the palate vault might be correlated with higher jaw activity in the production of sounds with a constriction in that region. A constriction within a highly domed palate would best be produced with higher tongue activity. According to this prediction the palate contours of the two speakers sr and ur would result in quite the opposite pattern of jaw activity. Speaker ur has a somewhat flat back palate contour but relatively small consonantal jaw activity, speaker sr has a domed contour and stronger jaw activity. The dome factor (see Table 6-5) of both is very similar. It could be argued that jaw activity for coronal consonants is more influenced by contours in the alveolar region, yet, they appear very similar for both speakers (Figure 6-17).

Only speaker rs showed a stronger deviating dome factor and a flat back part of the palate. However, his jaw activity patterns are comparable to the mean of other speakers.

Interestingly, speakers ur and sr have the strongest differences in translational jaw movement as indicated in Table 6-1. During /a:/ production speaker sr shows the largest proportion of translational jaw movement, while ur shows virtually no jaw translation in the speech tasks at all. It was already said that this translational effect is only noticable for lowest jaw position during /a:/ production, otherwise one could have assumed implications for our estimation of intrinsic tongue tip activity proportion especially for consonant production.

One can conclude that while there are sound specific patterns of jaw activity, individual speakers can show considerable differences (see also Hertrich/Ackermann 2000).

9. Conclusion

It could be shown that detailed articulatory and acoustic analysis is necessary and useful to evaluate phonetic theories. In this context, electromagnetic articulography was a useful and necessary technique to cope with movements of otherwise very difficult to analyse articulators. The simultaneous precision in temporal and spatial resolution would not have been available with any other technique, e.g. MRI. Additionally this technique allows simultaneous high quality acoustic recordings.

We will now summarize the overall findings and try to interpret them with respect to some theories of speech production.

A variety of articulatory and acoustic factors is involved in the production of loud speech. It has been shown that overall jaw behavior for speech production cannot be interpreted as a simple general lowering, comparable to a bite block effect. In fact jaw gestures increase in loud speech, thus have to be rescaled. Lowered jaw position together with the higher first formant effect can be explained by either an aerodynamic-acoustic constraint to avoid friction for high vowel production or acoustical spectral compensation for higher fundamental frequency. The perceptual result (Bark-scaled formant charts) remained as a consequence unaffected by louder voice. The articulation of most consonants was not significantly affected by louder voice condition. This articulatory behavior for consonant production supports theories of a coordinated structure for jaw and tongue articulation, insofar as both interact to produce an oral occlusion or constriction. Additionally, however, we must assume that for the production of at least coronal consonants the jaw height has to be rather controlled, and for strident fricatives even acts as an active articulator. This contradicts assumptions as that of Tuller and Kelso (1984) and others, that form a basis of the task-dynamics model (Saltzman 1986) and subsequently Articulatory Phonology (Browman/Goldstein 1986ff) that jaw opening is more strongly associated with vowel gestures and that jaw and tongue articulation can and do strongly compensate each other. An approach as Articulatory Phonology could certainly cope with that by further differentiating the specifications of individual articulators, but might become thus unattractively complexer.

We can state that the paradigm of compensatory articulation that had been first demonstrated in bite-block experiments does not play a major role in tongue tip - jaw interaction of vowels and coronal consonants.

It has been argued (Lindblom 1983, Keating 1983) that jaw articulation functions for a purpose of higher syllabic structuring. Keating (1983) assumes for „consonants as anchoring jaw height [in a syllable] with vowels and other consonants accomodating“ (Keating 1983 as in Keating et al. 1989: 77). While the analysis of consonantal and vowel jaw height variability in Keating et al. (1989) might be a little misleading, we think their basic question „Do consonants vary contextually more than vowels or the reverse“ (Keating et al. 1989: 77) could be commented by our data. We found contextual influences on jaw height in high and low vowels, but only for lower jaw height consonants /n/, /l/, /d/ and conclude that it is legitimate to assume larger jaw variability for vowels than for coronal consonants. With respect to this, we find Keatings (1983) description attractive. A jaw height parameter for syllabic organization would be able to explain syllable onsets like /st/ better than analyses with segmental sonority.

We can then state that in loud speech condition two patterns of extended contrast can be found. First, acoustic durational patterns increase the vowel/consonant ratio. More intensive vowels are lengthened, consonants that show generally less intensity are shortened. Second, the jaw amplitude and jaw work proportion are increased. We will now look for possible explanations: It seems to be possible and relevant for a listener to judge the volume of an utterance fairly precisely (Wilkens und Bartels 1977). Although volume is a paralinguistic phenomenon, which does not change the actual meaning of a word or a phrase it probably is of some pragmatic interest for the listener, since it may indicate distance to the speaker or give some hint to his emotional state. It might also carry linguistic information to mark an accent, in combination with segmental lengthening and/or fundamental frequency changes. There are some indications in our data that stressed and unstressed vowels show differing acoustic volume effects. However, a detailed analysis was not presented here since the long unstressed vowels in our nonsense material deviates from German real word stress patterns. Those changes can all be subsumed under the label „heightened effort“ (Ladefoged/Mc Kinney 1963).

As a more general implication for a model of speech production we want to conclude that neither a strong mapping to acoustic features nor a strong theory of motor units can hold on it’s own. We find it preferable to assume that speech perception and production use both articulation and acoustics as parallel channels to evoke a perceptual symbolic „event“ („Ereignis“ Tillmann 1980).

9.1 Outlook

The final recourse to perceptual targets seems to be a very interesting point for further research. It was beyond the limits of this works to perform perceptual experiments with audio-visual stimuli, although video data of the subjects exist. It would be interesting to examine to what extent listeners use the visual channel as well to estimate loudness, preferably by estimating jaw amplitude.

References

Badin, P., Perrier, P., Boë, L.-J., Abry, C. (1990) Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *JASA* 87(3), 1290-1300.

Baer, T., Gore, J.C., Gracco, L.C., Nye, P.W. (1991) Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *JASA* 90(2), 799-827.

Barry, W. G., Hawkins, S. (1992) Comments on Chap. 5 (On types of coarticulation, N. Hewlett/ L. Shockey). In: Docherty/Ladd (eds.), 138-145.

Barry, W.J. (1995) Cognitive aspects of phonetics. *Ms.* Summerschool of DGfS Saarbrücken 1995.

Bell, A.M. (1867) *Visible Speech*. London: Simpkin, Marshall and Co.

Benoît, C. (1986) Note on the use of correlations in speech timing. *JASA* 80(6), 1846-1849.

Benoît, C., Martin, J., Pelachaud, C., Schomaker, L., Suhm, B. (to appear) Audiovisual and multimodal speech systems. In: D. Gibbon, (ed.)*, Handbook of Standards and Resources for Spoken Language Systems - Supplement volume.* [On-line] http://www.citeseer.nj.nec.com/benoit98audiovisual.html

Bizzi, E., Mussa-Ivaldi, F.A. (1989) Geometrical and Mechanical Issues in Movement Planning and Control. In: Posner (ed.) 769-792.

Blackburn, C.S., Young, S.J. (1995) Towards improved speech recognition using a speech production model. Proc. Eurospeech 1995, Madrid, Spain, 1623-1626.

Boersma, P. (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 17, 97-110

Boersma, P. (1998) *Functional Phonology*. The Hague: Holland Academic Graphics.

Boersma, P., Weenink, D. Praat: doing phonetics by computer (Version 3.9.00 for linux) [Computer program], from http://www.praat.org/

Bonnot, J.-F.P., Chevrie-Muller, C. (1991) Some effects of shouted and whispered conditions on temporal organization. *J. Phonetics, 19, 473-483.*

Borden, G., Harris, C. (1980) Speech science primer. Baltimore, London: Williams & Wilkins.

Broe, M.B., Pierrehumbert, J.B., eds. (2000) Papers in Laboratory Phonology V: Acquisition and the lexicon. Cambridge: Cambridge University Press.

Browman, C.P., Goldstein, L. (1986) Towards an articulatory phonology. *Phonology Yearbook* 3, 219-252.

Browman, C.P., Goldstein, L. (1988) Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica* 45, 140-155.

Browman, C.P., Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology* 6, 201-251.

Browman, C.P., Goldstein, L. (1990a). Gestural specification using dynamically defined articulatory structures. *J. Phonetics* 18, 299-320.

Browman, C.P., Goldstein, L. (1990b) Tiers in articulatory phonology, with some implications for casual speech. In: Kingston/Beckman (eds.), 341-376.

Browman, C.P., Goldstein, L. (1991) Gestural Structures: Distinctiveness, Phonological Processes, and Historical Change. In: Mattingly/Studdert-Kennedy (eds.), 313-338.

Browman, C.P., Goldstein, L. (1992a) Articulatory Phonology: An Overview. (With Response to Commentaries). *Phonetica* 49, 155-180, 222-234.

Browman, C.P., Louis Goldstein (1992b) 'Targetless' schwa: an articulatory analysis. In: Docherty/Ladd (eds.), 26-56.

Bühl, A., Zöfel, P. (2000) *SPSS Version 9*. München et al.: Addison-Wesley.

Catford, J.C. (1977) *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

Cho, T. (2001) Effects of Prosody on Articulation in English. Ph. Diss. UCLA, New York, Routledge.

Chomsky, N., Halle, M. (1968) *The Sound Pattern of English*. New York, Evanston, London: Harper & Row.

Clauß, G., Finze, F.-R., Partzsch, L. (1995) *Statistik. Für Soziologen, Pädagogen, Psychologen und Mediziner. Band 1: Grundlagen.* 2nd revised edition, Thun, Frankfurt am Main: Harri Deutsch.

Clements, G.N., Keyser, S.J. (1983) *CV Phonology. A general theory of the syllable.* Cambridge et al.: MIT Press.

Docherty, G.J., Ladd, D.R., eds. (1992) *Papers in Laboratory Phonology. Vol. II: Gesture, Segment, Prosody.* Cambridge: CUP.

Dromey, C., Ramig, L.O., Johnson, A.B. (1995) Phonatory and articulatory changes associated with increased vocal intensity in parkinson disease: A case study. *JSHR* 38, 751-764.

Edwards, J. (1985) Contextual effects on lingual-mandibular coordination. *JASA 78*, 1944-1948.

Edwards, J., Harris, K.S. (1990) Rotation and translation of the jaw during speech. *JSHR* 33, 550-562.

Erickson, D. (1998) Effects of contrastive emphasis on jaw opening. *Phonetica* 55, 147-169.

Eriksson, A., Traunmüller, H. (1999) Perception of vocal effort and speaker distance on the basis of vowel utterances. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 3, 2469-2472.

Evers, V., Reetz, H., Lahiri, A. (1998) Crosslinguistic acoustic categorization of sibilants independent of phonological status. *J. Phonetics*, 26, 345-370.

Fowler, C.A. (1980) Coarticulation and theories of extrinsic timing. *J. Phonetics* 8, 113 - 133.

Fowler, C.A. (1981) A Relationship between Coarticulation and Compensatory Shortening. *Phonetica* 38, 35-50.

Fowler, C.A. (1986) An event approach to the study of speech perception from a direct-realist perspective. *J. Phonetics* 14, 3-28.

Fowler, C.A., Rubin, P., Remez, R.E., Turvey, M.T. (1980) Implications for speech production of a general theory of action. In B. Buttterworth (Ed.) *Language Production vol.1.: Speech and talk*. London et al.: Academic Press, 373-420.

Fowler, C.A., Smith, M.R. (1986) Speech Perception as 'Vector Analysis': An Approach to the Problems of Invariance and Segmentation. In: Perkell/Klatt (eds.), 123-139.

Fujimura, O. (1986) Relative Invariance of Articulatory Movements: An Iceberg Model. In: Perkell/Klatt (eds.), 226-233.

Fujimura, O. (1990) Toward a model of articulatory control: comments on Browman and Goldstein's paper. In: Kingston/Beckman (eds.), 377-381.

Fujimura, O. (1992) Phonology and Phonetics - A syllable based model of articulatory organisation. *J. Ac. Soc. Jpn, (E)* 13/1, 39-48.

Gay, T. (1977) Articulatory movements in VCV sequences. *JASA* 62, 183-193.

Geumann, A. (1994) Zur Beschreibung von Reduktionsprozessen in kontinuierlicher Sprache in einem gesturalen Ansatz. Unpublished masters thesis University Cologne.

Geumann, A., Kroos, C., Tillmann, H.G. (1999) Are there compensatory effects in natural speech? *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 1, 399-402.

Gray, H. (1901) *Anatomy, descriptive and surgical*. New York; Avenel, NJ: Gramercy Books. Reprint 1977, New York et al.: Random House.

Guenther, F.H. (1995a) Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102*, 594-621.

Guenther, F.H. (1995b) A modeling framework for speech motor development and kinematic articulator control. *Proceedings of the 13 th International Congress of Phonetic Sciences (ICPhS),* Stockholm, 1, 92-99.

Guenther, F.H., Hampson, M., Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review* 105, 611-633.

Hall, T.A. (1997) *The Phonology of Coronals*. Amsterdam, Philadelphia: John Benjamins.

Hardcastle, W., Vaxelaire, B., Gibbon, F., Hoole, P., Nguyen, N. (1996) EMA/EPG Study of lingual coarticulation in /kl/ clusters. *Proc. 1st ESCA Tutorial and Research Workshop on Speech Production Modeling - 4th Speech Production Seminar.* Autrans, France, 53-56.

Hawkins, S. (1992) An introduction to task dynamics. In: Docherty, G., Ladd, D.R. (eds.) *Papers in Laboratory Phonology. Vol. II: Gesture, Segment, Prosody*. Cambridge: CUP, 9-25.

Heid, S. (1998) Phonetische Variation: Untersuchungen anhand des PhonDat2-Korpus. (Phonetic variation - examinations based on the PhonDat2 corpus) Diss. University of Munich, in: *FIPKM* 36, 193-367.

Heid, S., Hawkins, S. (2000) An acoustical study of long domain /r/ and /l/ coarticulation. *Proceedings of the 5th seminar on speech production: models and data. & CREST Workshop on models of speech production: Motor planning and articulatory modelling*. Kloster Seeon Bavaria Germany May 1st to 4th, 2000, 77-80.

Heike, G. (1979) Articulatory measurement and synthesis. Methods and preliminary results. *Phonetica* 36, 294-301.

Heike, G. (1980) Articulatory Synthesis of German Monosyllables. In: G. Brettschneider/C. Lehmann (eds.). *Wege zur Universalienforschung. Beiträge zum 60. Geburtstag von Hans-Jakob Seiler.* Tübingen: Narr, 566-570.

Heike, G. (1987) The articulatory structure of german in the framework of an articulatory model. In: *Beiträge zu Phonetik und Linguistik. Festschrift H.-H. Wängler.* Hamburg: Buske, 95-110.

Heike, G. (1989) Ein phonologisches Artikulationsmodell des Deutschen. In: E. Feldbusch (ed.). *Ergebnisse und Aufgaben der Germanistik am Ende des 20. Jahrhunderts. Festschrift für L. E. Schmitt zum 80. Geburtstag.* Hildesheim/Zürich/ New York: Olms-Weidmann, 159-173.

Heike, G., Philipp, J. (1985) LISA - Ein Verfahren zur artikulatorischen Sprachsynthese. In: B. S. Müller (ed.). *Sprachsynthese*. Hildesheim: Georg Olms, 39-53.

Heldner, M., Strangert, E. (2001) Temporal effects of focus in Swedish. *J. Phonetics* 29, 329-361.

Henke, W.L. (1966) *Dynamic articulatory model of speech production using computer simulation*. Diss. MIT.

Hertrich, I., Ackermann, H. (1995) Coarticulation in slow speech: Durational and spectral analysis. *Language and Speech* 38(2), 159-187.

Hertrich, I., Ackermann, H. (1997) Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *JASA* 102(1), 523-536.

Hertrich, I., Ackermann, H. (2000) Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions. *JASA* 107(4), 2236-2247.

Holmberg, E.B., Hillman, R.E., Perkell, J.S. (1988) Glottal Airflow and transglottal air pressure measurements for male and female seakers in soft, normal, and loud voice. *JASA* 84(2), 511-529.

Hoole, P. (1987) Bite-block speech in the absence of oral sensibility. *Proceedings of the 11th International Congress of Phonetic Sciences (ICPhS)*, 16-19.

Hoole, P. (1993) Methodological considerations in the use of electromagnetic articulography in phonetic research. In *FIPKM* 31, 43-64.

Hoole, P. (1996a) Theoretische und methodische Grundlagen der Artikulationsananlyse in der experimentellen Phonetik. Diss. University of Munich, In: *FIPKM* 34, 3-174.

Hoole, P. (1996b) Issues in the acquisition, processing, reduction and parameterization of articulographic data, Published as part of Hoole (1996a) *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation, München (FIPKM)* 34, 158-173.

Hoole, P. Nguyen-Trong, N., Hardcastle, W. (1993) A comparative investigation of coarticulation in fricatives: electropalatographic, electromagnetic, and acoustic data. *Language and Speech 36(2, 3), 235-260.*

Hoole, P., Gfroerer, S., Tillmann, H.G. (1990) Electromagnetic articulography as a tool in the study of lingual coarticulation. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation, München (FIPKM)* 28: 107-122.

Hoole, P., Kühnert, B. (1996) Tongue-jaw coordination in German vowel production. *Proc. 1st ESCA Tutorial and Research Workshop on Speech Production Modeling - 4th Speech Production Seminar.* Autrans, France, 97-100.

Hoole, P., Wismüller, A., Leinsinger, G., Kroos, C., Geumann, A., Inoue, M. (2000) Analysis of the tongue configuration in multi-speaker, multi-volume MRI data. *Proceedings of the 5th seminar on speech production: models and data. & CREST Workshop on models of speech production: Motor planning and articulatory modelling*. Kloster Seeon Bavaria Germany May 1st to 4th, 157-160.

Isshiki, N. (1964) Regulatory mechanisms of voice intensity variation. *JSHR* 7, 17-29.

Jackendoff, R. (1997) *The Architecture of the Language Faculty.* Cambridge, MA, London: MIT Press.

Jackson, M.T.T. (1988) Analysis of tongue positions: Language-sepecific and cross-linguistic models. *JASA* 84(1), 124-143.

Jakobson, R., Fant, G., Halle, M. (1952) *Preliminaries to speech analysis. The distinctive features and their correlates*. 9th edition 1969, Cambridge, MA: MITPress.

Jespersen, O. (1904) *Lehrbuch der Phonetik*. Leipzig, Berlin: B.G. Teubner.

Johnson, K., Ladefoged, P., Lindau, M. (1993) Individual differences in vowel production. *JASA* 94(2), 701-714.

de Jong, K. (1995) The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *JASA* 91(1), 491-504.

de Jong, K., Beckman, M., Edwards, J.R. (1993) The interplay between prosody and coarticulation. *Language and Speech*. 36, 197-212.

Jordan, M.I., D.A. Rosenbaum (1989) Action. In: Posner, M.I. (ed.) *Foundations of cognitive science.* Cambridge, MA, London: MIT Press, 727-767.

Jordan, T.R., Sergeant, P. (2000) Effects of distance on visual and audiovisual speech recognition. *Language and Speech* 43(1), 107-124.

Keating, P.A. (1983) Comments on the jaw and syllable structure. *J. Phonetics* 11, 401-406.

Keating, P.A. (1985) CV Phonology, Experimental Phonetics, and Coarticulation. *UCLA WP in Phonetics* 62, 1-13.

Keating, P.A. (1988) Underspecification in phonetics. *Phonology* 5, 275-292.

Keating, P.A., Lindblom, B., Lubker, J., Kreiman, J. (1989) Jaw Position in English and Swedish VCVs. *UCLA WP in Phonetics* 74, 77-95 (also in: PERILUS XII (WP Stockholm) May 1991, 139-160.

Keating, P.A.., Lindblom, B., Lubker, J., Kreiman, J. (1994) Variability in jaw height for segments in English and Swedish VCVs. *J. Phonetics* 22, 407-422.

Kelso, J., Tuller, B., Vatikiotis-Bateson, E., Fowler, E. (1984) Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *J. Exp. Psychology: Human Perception and Performance* 10, 812-832.

Kelso, J.A.S., Saltzman, E.L., Tuller, B. (1986) The dynamical perspective on speech production data and theory. *J. Phonetics* 14, 29-59.

Kingston J.H., Beckman, M.E., eds. (1990) *Papers in Laboratory Phonology. Vol. I: Between the Grammar and Physics of Speech.* Cambridge: CUP.

Kirchner, R. M. (1998) An Effort-Based Approach to Consonant Lenition. Ph. Diss, UCLA.

Klapp, S.T., Anderson, W.G., Berrian, R.W. (1973) Implicit speech in reading, reconsidered. *J. Experimental Psychology* 100, 368-374.

Kohler, K.J. (1977) *Einführung in die Phonetik des Deutschen*. Berlin: E. Schmidt.

Kohler, K.J. (1984) Phonetic Explanation in Phonology: The Feature Fortis/Lenis. *Phonetica* 41, 150-174.

Kohler, K.J. (1986) Invariance and Variability in Speech Timing: From Utterance to Segment in German. In: Perkell/Klatt (eds.), 268-299.

Kohler, K.J. (1990) Segmental Reduction in Connected Speech in German: Phonological Facts and Phonetic Explanations. In: W. J. Hardcastle/A. Marchal (eds.). *Speech Production and Speech Modelling*. Dordrecht: Kluwer, 69-92.

Kohler, K.J. (1991) The Phonetics/Phonology Issue in the Study of Articulatory Reduction. *Phonetica* 48, 180-192.

Kohler, K.J., van Dommelen, W., Timmermann, G. (1981) Die Merkmalpaare stimmhaft/stimmlos und fortis/lenis in der Konsonantenproduktion des heutigen Standard-Französisch. *Institut für Phonetik, Universität Kiel. Arbeitsberichte 14*.

Kröger, B. (1993) A gestural production model and its application to reduction in German. *Phonetica* 50, 213-233.

Kröger, B.J., Schröder, G., Opgen-Rhein, C. (1995) A gesture-based dynamic model describing articulatory movement data. *JASA* 98(4), 1878-1889.

Kuehn, D.P., Moll, K. (1976) A cineradiographic study of VC and CV articulatory velocities. *J. Phonetics* 4, 303-320.

Kühnert, B. (1996) Die alveolare-velare Assimilation bei Sprechern des Deutschen und Englischen: Kinematische und perzeptive Grundlagen. Diss. University of Munich, in: *FIPKM* 34, 175-392.

Kühnert, B., Ledl, C., Hoole, P., Tillmann, H.G. (1991) Tongue-jaw interactions in lingual consonants. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm, (PERILUS) 14*, 21-25.

Ladefoged, P. (1990) On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein. In: J. Kingston & M. Beckman (eds.) *Papers in Laboratory Phonology 1*, Cambridge: Cambridge University Press, 398-405.

Ladefoged, P., DeClerk, J., Lindau, M., Papcun, G. (1972) An auditory-motor theory of speech production. *UCLA WP in Phonetics* 22, 48-75.

Ladefoged, P., Maddieson, I. (1996) *The Sounds of the World's Languages*. Oxford, Cambridge, MA: Blackwell.

Ladefoged, P., McKinney, N.P. (1963) Loudness, sound pressure, and subglottal pressure in speech. *JASA*, 35, 454-460.

Lane, H., Tranel, B. (1971) The Lombard sign and the role of hearing in speech. *JSHR* 14, 677-709.

Lee, S. (1994) A cross-linguistic study of the role of the jaw in consonant articulation. Ph.Dissertation, Ohio State, Department of Linguistics.

Lee, S., Beckman, M., Jackson, M. (1994) Jaw targets for strident fricatives. *Proc. of International Conference on Spoken Language Processing (ICSLP)* 1994 Yokohama, 37-40.

Levelt, W.J.M. (1989) *Speaking: From intention to articulation*. MIT Press: Cambridge, MA, London.

Levelt, W.J.M., Roelofs, A., Meyer, A.S. (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.

Liberman, A.M., Mattingly, I.G. (1985) The motor theory of speech perception revised. *Cognition* 21, 1-36.

Lindblad, P. Lundqvist, S. (1999) How and why do the tongue gestures of [t], [d], [l], [n], [s], and [r] differ? *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 417-420.

Lindblom, B. (1983) Economy of Speech. In: MacNeilage, P.F. (ed.) *The production of speech*. New York: Springer, 217-246.

Lindblom, B. (1986) On the origin and purpose of discreteness and invariance in sound patterns. In: Perkell/Klatt (eds.), 493-523.

Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H & H theory. In: Hardcastle/Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.

Lindblom, B. (1991) The Status of Phonetic Gestures. In: Mattingly/Studdert-Kennedy (eds.), 7-24.

Lindblom, B., Lubker, J., Gay, T. (1979) Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation, *J. Phonetics, 7*, 147-161.

Lindblom, B., Lubker, J., Gay, T., Lyberg, B., Branderud, P., Holmgren, K. (1987) The concept of target and speech timing. In: Channon, R., Shockey, L. (eds.) *In honor of Ilse Lehiste*. Dordrecht, Providence: Foris, 161-181.

Lindblom, B.E.F., Sundberg, J.E.F. (1971) Acoustical consequences of lip, tongue, jaw, and larynx movement. *JASA* 50(4), 1166-1179.

Löfqvist, A. (1997) Theories and models of speech production. In: W.J. Hardcastle and J. Laver (eds.), *Handbook of Phonetic Science*. Cambridge, MA: Blackwell, 405-426.

Lombard, E. (1911) Le signe de l'élévation de la voix. *Ann. Mal. Oreil. Larynx* 37, 101-119.

Lundqvist, S., Lindblad, P. (1999) The production of coronals in hyper and hypo speech. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 1, 421-424.

MacNeilage, P.F. (1998) The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, 499-546 (includes comments of various authors)

Mattingly, I.G. (1990) The global character of phonetic gestures. *J. Phonetics* 18, 445-452.

Mattingly, I.G., Studdert-Kennedy, M., eds. (1991) *Modularity and the motor theory of speech perception*: Proceedings of a conference to honor Alvin. M. Liberman. Hillsdale, New Jersey: Lawrence Erlbaum.

Menzerath, P., de Lacerda, A. (1933) *Koartikulation, Steuerung und Lautabgrenzung*. Berlin, Bonn: Ferd. Dümmlers.

Mooshammer, C. (1998) Experimentalphonetische Untersuchungen zur artikulatorischen Modellierung der Gespanntheitsopposition im Deutschen (The phonetics of the tense-lax opposition in German: experimental and modelling studies) Diss. Humboldt University Berlin, in: *FIPKM* 36, 3-192.

Munhall, K.,Tohkura, Y. (1998) Audiovisual gating and the time course of speech perception. *JASA* 104(1), 530-539.

Munhall, K.G., Kawato, M., Vatikiotis-Bateson, E. (2000) Coarticulation and physical models of speech production. In: Broe, M.B., Pierrehumbert, J.B. (eds.), 9-28.

Narayanan, S.S.(1995) *Fricative consonants: An articulatory, acoustic, and systems study*. Ph. Diss. UCLA.

Narayanan, S.S., Alwan, A. (1999) Noise source models for fricative consonants. *IEEE Transactions on Speech and Audio Processing.*

Nelson, W.L. (1983) Physical Principles for Economies of Skilled Movements. Biol. Cybernetics 46, 135-147.

Nelson W.L., Perkell, J.S., Westbury, J.R. (1984) Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. *JASA* 75(3), 945-951.

Nittrouer, S., Munhall, K., Kelso, J.A.S., Tuller, B., Harris, K.S. (1988) Patterns of interarticulator phasing and their relation to linguistic structure. *JASA* 84/5, 1653-1661.

Nolan, F. (1992) The descriptive role of segments: evidence from assimilation. In: Docherty/Ladd (eds.), 261-286.

Öhman, S.E.G. (1966) Coarticulation in VCV Utterances: Spectrographic Measurements. *JASA* 39, 151-168.

Öhman, S.E.G. (1967) Numerical model of coarticulation. *JASA* 41, 310-320.

Ostry, D.J., Shiller, D.M., Gribble, P.L. (2000) *Proceedings of 5th Seminar on Speech Production: Models and Data. & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling.* Kloster Seeon, Bavaria, 169-172.

Paradis, C., Prunet, J.-F. (1991) *Phonetics and Phonology 2. The special status of coronals: Internal and external evidence*. San Diego et al: Academic Press.

Paradis, C., Prunet, J.-F., eds. (1991) *Phonetics and Phonology Vol. II. The Special Status of Coronals*: Internal and External Evidence. New York: Academic Press.

Passy, P. (1890) *Etude sur les changements phonétiques et leur cararactères généraux*. Paris: Librairies Firmin-Didot.

Perkell, J.S. (1980) Phonetic features and the physiology of speech production. In: Butterworth, B. (ed.) *Language Production vol.1, Speech and Talk*. London: Academic Press.

Perkell, J.S. (1990) Testing theories of speech production: Implications of some detailed analyses of variable articulatory data. In: Hardcastle, W.J., Marchal, A. (eds.) *Speech production and speech modelling*. Dordrecht: Kluwer, 263-288.

Perkell, J.S. (1991) Models, theory and data in speech production. *Proc. of 12th ICPhS*. Aix en Provence, France, 182-191.

Perkell, J.S. (1997) Articulatory Processes. In: W.J. Hardcastle and J. Laver (eds.), *Handbook of Phonetic Science*. Cambridge, MA: Blackwell, 333-370.

Perkell, J.S., Cohen, M.H., Svirsky, M.A., Matthies, M.L., Garabieta, I., Jackson, M.T.T. (1992) Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *JASA* 92(6), 3078-3096.

Perkell, J.S., Klatt, D.H., eds. (1986) *Invariance and Variability in Speech Processes*. Hillsdale, New Jersey: Erlbaum Associates.

Perkell, J.S., Matthies, M.L., Svirsky, M.A., Jordan, M.I (1993) Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'motor equivalence' study, *JASA* 93, 2948-2961.

Perkell, J.S., Zandipour, M. Matthies, M. Lane, H. (1999) Articulatory kinematics: preliminary data on the effects of speaking condition, articulator and movement type. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco*, 1773-1776.*

Perkins, W.H., Kent, R.D. (1986) *Textbook of functional anatomy of speech, language, and hearing*. London, Philadelphia: Taylor & Francis

Pick, H.L., Siegel, G.M., Fox, P.W., Garber, S.R., Kearney, J.K. (1989) Inhibiting the Lombard effect. JASA 85(2), 894-900.

Pickett, J.M., Bunnell, H.T., Revoile, S.G. (1995) Phonetics of intervocalic consonant perception: Retrospect and prospect. *Phonetica* 52, 1-40.

Pols, L.C.W. (1999) Flexible, robust, and efficient human speech processing versus present-day speech technology. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 1, 9-16.

Posner, M.I., ed. (1989) *Foundations of cognitive science*. MIT Press: Cambridge, MA, London.

Prince, A., Smolensky, P. (1997) Optimality: From neural networks to universal grammar. *Science* 275, 1604-1610.

Rostolland, D. (1982a) Acoustic features of shouted voice. *Acustica*, 50, 118-125.

Rostolland, D. (1982b) Phonetic structure of shouted voice. *Acustica*, 51, 80-89.

Rostolland, D., Parant, C. (1983) The influence of voice sound level on the duration of french long vowels. *Proceedings of the 10th International Congress of Phonetic Sciences (ICPhS),* Utrecht. Dordrecht, Cinamminson: Foris Publications, 1984, vol. IIb, 205-209.

Rousselot, P.J. (1897-1901) *Principes de phonétique expérimentale*. 2 vols. Paris: H. Welter.

Saltzman, E. (1986) Task dynamic coordination of the speech articulators: a preliminary model. In: Heuer, H., Fromm, C. (eds.) *Generation and modulation of Action Patterns*. New York: Springer, 129-144.

Saltzman, E., Munhall, K. (1989) A dynamic approach to gestural patterning in speech production. *Ecological Psychology, 1(4)*, 333-382.

Savariaux, C., Perrier, P., Orliaguet, P.P. (1995) Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *JASA* 98, 2428-2442.

Schulman, R. (1989) Articulatory dynamics of loud and normal speech. *JASA, 85*, 295-312.

Shadle, C.H. (1985) The acoustics of fricative consonants. Ph.D. thesis, MIT.

Shadle, C.H. 1990. Articulatory-Acoustic Relationships in Fricative Consonants. In: W. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 187-209.

Shiller, D.M., Laboissière, R., Sanguineti, V., Ostry, D.J (2000) Jaw mechanical behavior during speech production. *Proceedings of 5th Seminar on Speech Production: Models and Data. & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling.* Kloster Seeon, Bavaria, 13-16.

Sievers, E. (1881) *Grundzüge der Phonetik*. Leipzig: Breitkopf und Härtel.

Simon, H.A. Kaplan, C.A. (1989) Foundations of cognitive science. In: M.I.Posner (ed.), *Foundations of cognitive science*. Cambridge, MA, London: MITPress, 1-47.

Sluijter, A.M.C., Shattuck-Hufnagel, S., Stevens, K.N., van Heuven, V.J (1995) Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English. *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS*), Stockholm, vol 2, 630-633.

Sluijter, A.M.C., van Heuven, V.J (1996) Spectral balance as an acoustic correlate of linguistic stress. *JASA* 100, 2471-2485.

Sluijter, A.M.C., van Heuven, V.J, Pacilly, J.J.A. (1997) Spectral balance as a cue in the perception of linguistic stress. *JASA* 101, 503-513.

Sock, R. (1998) *Organisation temporelle en production de la parole Émergence de catégories sensori-motrices phonétiques*. Presses universitaires du septentrion: Villeneuve d'Ascq Cédex.

Stathopoulos, E.T., Sapienza, C. (1993) Respiratory and laryngeal function of women and men during vocal intensity variation. *JSHR* 36, 64-75.

Steininger, S. (1998) *Handeln und Wahrnehmen: Eine experimentelle Analyse einer Wahrnehmungsbeeinträchtigung bei simultan ausgeführten Handlungen*. Diss University of Munich. Aachen: Shaker.

Sternberg, S., Monsell, S., Knoll, R.L., Wright, C.E. (1978) The latency and duration of rapid movement sequences.: Comparison of speech and typewriting. In Stelmach, G.E. (ed.*) Information processing in motor control and learning*. New York: Academic Press.

Sternberg, S., Wright, C.E., Knoll, R.L., Monsell, S. (1980) Motor programs in rapid speech: Additional evidence. In: Cole, R.A. (ed.), *Perception and production of fluent speech*. Hillsdale, N.J.: Lawrence Erlbaum.

Stevens, K.N. (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In E.E. David and P.B Denes (eds.) *Human communication: a unified view.* New York, NY: Mc Graw Hill, 51-66.

Stevens, K.N. (1989) On the quantal nature of speech. *J. Phonetics,* 17, 3-45, 145-157.

Stevens, K.N. (1998) *Acoustic Phonetics*. Cambridge, MA, London: MIT Press.

Stevens, K.N., Keyser, S.J., Kawasaki, H. (1986) Toward a phonetic and phonological theory of redundant features. In Perkell/Klatt (eds.), 426-449 (comments 449-463).

Stone, M., Lundberg, A. (1996) Three-dimensional tongue surface shapes of English consonants and vowels. *JASA* 99(6), 3728-3737.

Sussman, H.M., MacNeilage, P.F., Hanson, R.J. (1973) Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *JSHR* 16, 397-420.

Sweet, H.(1877) *A handbook of phonetics*. Oxford: Clarendon Press.

Syrdal, A.K., Gopal, H.S. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels. *JASA* 79(4), 1086-1100.

Tatham, M. (1984) Towards a cognitive phonetics. *J. Phonetics,* 12, 37-47.

Tillmann, H.G. (1994) Phonetics: early modern, especially instrumental work. In: R. Asher (ed.) *The encyclopedia of language and linguistics*. Oxford: Pergamon, vol. 6, 3082-3094.

Tillmann, H.G. with Mansell, P. (1980). *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. Stuttgart: Klett Cotta.

Traunmüller, H. (1994) Conventional, biological and environmental factors in speech communication: a modulation theory. *Phonetica* 51, 170-183.

Traunmüller, H., Eriksson, A. (2000) Acoustic effects of variation in vocal effort by men, women, and children. *JASA,* 107(6), 3438-3451.

Tuller, B., Kelso, J.A.S. (1984) The timing of articulatory gestures. Evidence for relational invariants. *JASA* 76, 1030-1036.

Ungeheuer, G. (1962) Elemente einer akustischen Theorie der Vokalartikulation. Berlin, Göttingen, Heidelberg: Springer.

Van Son, R.J.J.H., Pols, L.C.W. (1995) Acoustic consonant reduction: a comparison. *Institute of Phonetic Sciences, University of Amsterdam, Proceedings* 19, 83-91.

Van Son, R.J.J.H., Pols, L.C.W. (1999) An acoustic description of consonant reduction. *Speech Communication* 28, 125-140.

Van Summers, W., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., Stokes, M.I. (1988) Effects of noise on speech production: Acoustic and perceptual analyses. JASA 84(3), 917-928.

Warren, R.M. (1977) Subjective loudness and its physical correlate. *Acustica*, 37, 334-346.

West, P. (1999) The extent of coarticulation of English liquids: an acoustic and articulatory study. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 1901-1904.

West, P. (2000) Long-distance coarticulatory effects of British English /l/ and /r/: An EMA, EPG and acoustic study. *Proceedings of the 5th seminar on speech production: models and data. & CREST Workshop on models of speech production: Motor planning and articulatory modelling*. Kloster Seeon Bavaria Germany May 1st to 4th, 2000, 105-108.

Westbury, J.; Lindstrom, M. & McClean, M. (2002). Tongue and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language and Hearing Research* 45, 651-662.

Wilkens, H. Bartel, H.-H. (1977) Wiedererkennbarkeit der Originallautstärke eines Sprechers bei elektroakustischer Wiedergabe. *Acustica*, 37, 45-49.

Wood, S. (1991) X-ray data on the temporal coordination of speech gestures. *J. Phonetics*, 19, 281-292.

Yehia, H., Kuratate, T., Vatikiotis-Bateson, E. (2000) Facial animation and head motion driven by speech acoustics. *Proceedings of 5th Seminar on Speech Production: Models and Data. & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling.* Kloster Seeon, Bavaria, 265-268.

Zlokarnik, I. (1995a) Adding articulatory features to acoustic features for automatic speech recognition. *ASA 129th Meeting* Washington ,DC 1995-May 30 - June06. [Abstract]

Zlokarnik, I. (1995b)Articulatory kinematics from the standpoint of automatic speech recognition. *ASA 129th Meeting* Washington ,DC 1995-May 30 - June06. [Abstract]