An investigation of coarticulation in rare and frequent syllables with regard to (the existence of) a mental syllabary

Diplomarbeit No. 49 im Fach Computerlinguistik

von

INES FLECHSIG

Beginn der Arbeit: 15.02.2006
Abgabe der Arbeit: 20.09.2006

Institut für Maschinelle Sprachverarbeitung
Azenbergstraße 12
D-70174 Stuttgart
Germany

Betreuer:	PD Dr. Bernd Möbius, IMS Phonetik
Prüfer:	PD Dr. Bernd Möbius, IMS Phonetik

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Stuttgart, den 20. September 2006

(Ines Flechsig)

Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank Bernd Möbius, who has always been a great source of ideas and information, for being my tutor and providing me with this work's exciting topic.

Then I am very thankful to my parents who have had a lot of concerns but who have always supported me with their special mixture of mental support and hard-earned patience.

Next I want to thank Andreas Madsack for sharing his great knowledge of programming with me and for introducing me together with Ralf Jankowitsch to the spirits of LaTeX. I am especially grateful to Andreas, Ralf and Manuel Kountz for their general help and ideas and for the joy they brought me during this year.

I would also like to thank Vera Demberg for answering my questions on English. My thanks go to Antje Schweitzer, too, for discussing problems with the German probability list with me and for being my contact in terms of Festival.

Last, I want to thank lenchen for the bottomless carafes of black and green tea and hot chocolate, for kindling ara on cold winter days and for always being there.

1 Introduction

A mental syllabary is a supposed mental storage for syllables that are frequently used in speech production. It is thought of to be present for efficiency reasons during speech production, retrieving frequent syllables from the storage instead of assembling them new each time. If a syllable stems from the syllabary or has been assembled from segments is supposed to result in differences in the syllable's properties and behaviour. Frequent and therefore presumably retrieved syllables are said to be more prone to coarticulatory effects than assembled syllables. Coarticulation is an ubiquitous phenomenon, existent in every human language that is spoken around the world. There are different levels of coarticulation varying from one phone influencing another phone in terms of place and manner of articulation to elision of whole syllables. The present work presents and reviews three different approaches to find evidence or counter-evidence for the existence of a mental syllabary in German. For the investigations minimal pairs were used that had been retrieved from a large speech corpus, their distinction being syllable probability which had been attained by means of a statistical calculation. The first approach was an investigation of syllable duration and the factors that interact there. It will be shown that the main influencing factor on syllable duration is syllable stress, followed by the position of the syllable in the word. Syllable probability will be reported as being only marginally significant. The second approach was investigating fundamental frequency of syllables. It will be shown that only minimal differences can be found in minimal pairs when stress is not taken into account. In fact, not respecting stress may mislead calculations to exhibit effects that contradict the mental syllabary theory. Therefore it will be found that frequent syllables are more prone to coarticulatory effects than rare syllables but also that unstressed rare syllables are more influenceable by coarticulation than stressed frequent syllables. As a third approach formant transitions have been investigated, showing significant differences between rare and frequent syllables. Here also it will be shown that these differences become even more clear when syllable stress is taken into account.

2 Theory

This section will introduce the main topics and their theories that are the basis for the present work. First, coarticulation will be explained along with its various specifications, and examples will be given for the different types of coarticulation. Then the mental syllabary theory will be introduced, its place in the speech production process and its relevance for the present work.

2.1 Coarticulation

2.1.1 Definition

Coarticulation denotes an act of articulatory smoothing: during speech production, different processes such as word form retrieval and phonological word formation are involved. The phonetic part comprises the access of stored information on how to produce a certain speech sound ideally, i.e., what articulators are involved and how they have to be triggered. This information is given in form of commands to the articulators that are involved in producing the respective speech sound. The articulators now are supposed to perform according to the phonological rules, but as they are subject to inertia and additionally have to create transitions from one speech sound to the next one, it is natural that the ideal pronunciation is hardly ever achieved. The tongue is the most flexible articulator, being fast and able to perform complex muscular coordination. Especially the tongue tip can be positioned very exactly even at a high speech rate. Moreover, the tongue is equipped with the most sensors in comparison with the other articulators, and among these are proprioceptive sensors which have the fastest way of signalling information to the brain (Kohler, 1977: 208). Thus, the tongue differs in its way of movement from other articulators. During coaction of organs with differing articulatory precision, time shifts occur during movement, i.e., the result is either persistence or anticipation of an articulatory parameter (Kohler, 1977: 208). The result of what the articulators are doing is nearly always a compromise between the phonological prototype of a speech sound and the efficiency needs of speech. According to Clark & Yallop (1995) when they refer to Lindblom (1983), distinctiveness and communicative effectiveness are primary motives in speech production. Coarticulation is therefore an ubiquitous phenomenon. It is exhibited in everyday speech and evidence has been found in languages all over the world. However, there is no uniformity across languages (Clark & Yallop, 1995); indeed, in each language, a language-dependent “phonological conditioning” (Clark & Yallop, 1995: 88) takes place. Coarticulation is also context-sensitive, i.e., the gradation and type of coarticulation depends on the surrounding speech sounds.

2.1.2 Types of coarticulation

In fact, the term *coarticulation* has at least two readings: in its narrow sense it means simultaneous movement of two different articulators as it occurs e.g., when lip rounding takes place in a consonant that precedes a rounded vowel; in its broader sense coarticulation means ’adaption’ or ’accommodation’ which describes sort of a compromise in articulatory positioning (Clark & Yallop, 1995). The present work will adapt to Clark & Yallop (1995) in not differentiating between the narrow and the broader sense. Here, coarticulation will refer to any kind of articulatory phenomena where influences between segments of speech can be found and that can be claimed to belong to the following classes of coarticulation: (the items of the list are often not unrelated to each other and may interact)

- *Left-to-right coarticulation*, *perseverative coarticulation* or *carry-over coarticulation* is due to a lag in articulatory movement which is caused by inertia, i.e., the (or some) articulators are not fast enough in moving from one articulatory position to the next one, therefore a delay in pronunciation is caused. An example would be devoicing which occurs quite frequently in German when a phonologically voiced consonant is preceded by a voiceless consonant like in /plats/ (Platz) where the /l/ may be devoiced because of the preceding /p/. Another occurrence would be the upkeep of a velar articulatory production as in /e:b@n/ (eben) which would then be produced as /e:bm/(example from Kohler (1977: 209)).

- In *right-to-left coarticulation*, which is also often referred to as *anticipatory coarticulation*, an articulator takes up a position already at a stage of production that precedes that stage where this position is necessary, i.e., the articulator anticipates a later stage in speech production. For example, /z/ as in /zy:s/ (süß) is normally rounded whereas it is not in /zi:/ (sie). But an anticipatory effect would also be voicing premature to the voiced segment. In German this may happen at boundaries that start with /h/. /h/ then can only be followed by a vowel and since /h/ is phonologically underspecified for voicing (Kohler, 1977: 160), /h/ may be voiced before the vowel itself starts.

- The term *assimilation* is an older term for coarticulation and is used "more or less synonymously with coarticulation" Clark & Yallop (1995). It often refers only to that kind of coarticulatory phenomena where two phonemes overlap at at least one articulatory parameter, e.g., in the example from above where /e:b@n/ becomes /e:bm/. The upkeep of the velar production for /@n/ is a matter of assimilation. This example was *progressive assimilation*, but *regressive assimilation* is also possible like in "an beide" which can be produced as /ambaId@/, thereby /b/ having an anticipatory effect on /n/ (example from Kohler (1977: 213)). (Kohler, 1977: 213) instances even an both-way assimilation: in "ich habe" /b/ may be assimilated to /B/.

- *Elision* refers to an act of deleting whole segments. Elisions are divided in elisions of consonants and vowels. Moreover, they are categorisable dependent on their position in linguistic units, e.g., whether they occur on word-, morpheme-, or syllable boundaries or mid-syllable, but also dependent on the context and other factors (Kohler, 1977: 213). A frequent form of elision in German is the deletion of syllable-final vowels like in "kommen" where /@/ is elided to leave /kOmm/ (/n/ being assimilated to /m/).

- *Target undershoot* can be viewed especially in rapid speech, when the tongue (or other articulators) do not have the time to reach the target position for the pronunciation of a phoneme because the next phoneme has already to be produced. The inability of (e.g.) the tongue to move to peripheral positions in the appropriate time results in a more centralised pronunciation of (some) segments (Clark & Yallop, 1995: 86).

- *Vowel reduction* is caused by the aforementioned target undershoot: The centralisation of peripheral vowels results in reduced forms of them (Clark & Yallop, 1995: 86). Normally unstressed vowels are reduced. An example would be /i:/ in "die" which could be reduced to /I/ (Hakkarainen, 1995: 65). Hakkarainen (1995: 65) reports different gradations in vowel reduction, elision (see above) being then the last consequence in reduction.

Now, as could be seen from the listing, coarticulation occurs in various forms and is context-dependent as well as context-sensitive. It is always language-dependent although some forms of coarticulation occur in more than one language. The present work will try to discover coarticulatory effects in order to find out whether there are differences in the articulation of frequent and rare syllables with regard to the mental syllabary theory.
Modarresi et al. (2004) investigated bi-directionality of coarticulation in VCV utterances in American English, differentiating between open sequences (V.CV) and closed sequences (VC.V), the point indicates the syllable boundary. Given a fixed vowel set (/i/,/O/,/e/,/u/) and a changing vowel set (/i/,/O/), they varied the vowel combination that surrounded the consonant, the changing vowel set constituting either the first or last vowel whereas the fixed vowel set constituted the respective other vowel in the given syllable. The authors reported anticipatory coarticulation exceeding carry-over coarticulation in open syllables for labial and velar stops. Carry-over coarticulation, however, predominated in alveolar contexts. Moreover, alveolars resisted anticipatory effects the most but were very influenceable by carry-over effects. For closed syllables, the authors reported that carry-over coarticulation exceeded anticipatory effects across all stops.

2.2 The mental syllabary theory

The mental syllabary, assuming that it exists, is said to be part of the phonetic encoding stage during speech production (Levelt et al., 1999: 5). At this stage, the gestural score of a phonological word is computed. This takes place after the "selection and ordering of lexical items" (Crompton, 1982: 109) and belongs to the "specification of how they [the lexical items] are pronounced" (Crompton, 1982: 109). The mental syllabary is thought of as a storage for syllables that are used frequently during speech. Levelt et al. (1999) report that for languages like English and Dutch that have many different syllables, a speaker can handle 50% of their speech with about 80 different syllables and about 80% with 500 syllables. German is a language which is similar in this respect (Croot & Rastle, 2004: 1), and the present work will try to account for the mental syllabary effect in German. The basic assumption behind the mental syllabary theory is that "syllables play a functional role in the process of word-form encoding" (Cholin et al., 2004: 1). Cholin et al. (2004: 49) argue with reference to Levelt et al. (1999) that syllables, rather than segments, are the basic programming units of speech articulation. They refer to Fujimura & Lovins (1978) and Lindblom (1983) by stating that "only the syllable can form the appropriate source for late phonological processes, such as allophonic variation, coarticulation and assimilation". In detail, the role of a mental syllabary in speech production is thought of as follows: during phonetic encoding, speakers access a mental syllabary that stores the gestural programs for at least the frequent syllables. These gestural programs are activated by the segments of the phonological syllables (Levelt et al., 1999). Low-frequency words, on the contrary, are assembled "using the segmental and metrical information provided in the phonological representation" (Levelt et al., 1999: 32). High-frequency syllables may also be assembled by means of the phonological representation but access on readily stored syllable is likely to be faster.

Whiteside & Varley (1998: 1) speak of a "dual-route phonetic encoding", distinguishing between a "direct route" and a parallel "indirect route". The direct route is said to access the mental syllabary and to retrieve syllables from there. The indirect route, on the contrary, relies on the assembly of syllables from subsyllabic units. These two routes concur with each other with regard to speed: the one that produces the questioned syllable faster will be the one where the pronunciation of a syllable comes from. The two routes differ in the sort of syllables they are evoked for: The indirect route is accessed rather for novel and low frequency syllables, whereas the direct route is used for frequent syllables. But also in "careful speech scenarios" and "when more conscious attention is being directed to speech" (Whiteside & Varley, 1998: 1) the indirect route is likely to be evoked.

Crompton (1982) was the first who suggested that there may be a duality in the phonetic encoding in speech production to account for particular speech errors[1] that involve phonemes and syllables. Levelt[2] elaborated further on the idea and introduced the notion *mental syllabary*. Levelt & Wheeldon (1994: 245) argue that "most syllables a speaker uses are highly overlearned articulatory gestures" and that therefore it is highly probable that these syllables are readily stored especially from the view of efficiency: it is much more convenient to store syllables that are frequently used than to assemble them newly every time. The idea has been debated on since and different phenomena in speech that support this theory have been reported in the literature.

Levelt & Wheeldon (1994) conducted four experiments to investigate the relation between word and syllable frequencies with regard to naming latencies and the position of the syllable in the word. They conducted a naming task where subjects had to learn to associate arbitrary symbols with target words. Levelt & Wheeldon (1994) found that syllable frequency affects naming latency in bisyllabic words and that this effect is independent of word frequency. Moreover, they found that this effect is dependent on the last syllable's frequency of a word and that the complexity of the second syllable does not affect naming latency. Especially the independence of complexity from naming latency is an argument for the assumption that "gestural scores for syllables are retrieved as whole entities" (Levelt & Wheeldon, 1994: 263).

Whiteside & Varley (1998) examined in a repetition task experiment whether a mental syllabary may be operating in the phonetic encoding of high and low frequency monosyllabic words. They reported longer response latencies for low frequency words and a tendency of low frequency words to have a longer duration. Also, high frequency words showed a lower degree of formant movement.

---

[1] e.g., segmental errors such as exchanges of segments: they seem to be restricted to occur only for identical internal syllable positions. This is referred to as *syllable position constraint* (cf. Cholin et al. (2004))

[2] Levelt (1989), Levelt (1992), Levelt (1993)

Schweitzer & Möbius (2004) who made investigations by means of the same corpus as the present work argue that the 326 most frequent syllable types in the corpus, which occur more than 20 times each, account for 22,638 syllable tokens, thereby covering approximately 67% of the corpus. They concluded that at least for the frequent syllables there have to be enough exemplars that can be used as reference during speech production without having to make use of the assembly. Schweitzer & Möbius (2004) provided evidence for a frequency effect on syllables for German by using the same speech corpus like the present work. Assuming that if speakers had a storage for auditory representations of perceived speech tokens to serve as perceptual target regions in speech production then high frequent syllables will be stored whereas low frequent syllables will not. The target region is thereby defined by the many high frequent exemplars of a syllable. Target regions for low frequent syllables, instead, are thought of to be assembled from sub-components of a syllable, e.g. segments. The authors reported according to their expectation a better fit for the regression of z-scores of the durations for whole syllables and their corresponding segments in infrequent syllables than for the duration of syllables and their corresponding segments of frequent syllables.

Cholin et al. (2004) tried to find out whether the emergence of the syllable is traceable during the phonetic/phonological processing for Dutch. Therefore they conducted an odd-man-out experiment which was expected to reduce a preparation effect in the variable set as compared to the constant set, meaning that variable sets contained a syllable that had a segmental design different from the other syllables of this set. Thus, subjects should be prevented from concluding from one syllable of a variable set to all other syllables of the respective set. Cholin et al. (2004) reported that shorter item sets are produced faster than larger item sets which was exhibited by shorter voice onset latencies. Moreover the data confirmed their expectation that a syllable structure effect would occur: variable sets had significantly larger latencies than constant sets, thus indicating a preparation effect and providing support for the syllable as an important factor in late speech production.

Croot & Rastle (2004) conducted three experiments for English to examine whether high frequency non-word syllables and non-existing syllables show differences in response latency, accuracy, duration and coarticulatory effects. The first two experiments showed no significant effects that could be ascribed to a mental syllabary, and the third experiment exhibited only slight indication of a mental syllabary effect by showing differences in spectral correlates of the examined syllables.

Concluding from these and other investigations reported in the literature there is evidence of an existing mental syllabary, naming latencies that coincide with syllable frequency and slips of the tongue indicate that an assembly of syllables alone can not account for certain findings. But often enough, a mental syllabary effect can not be detected where it is assumed to exhibit influence. Outcomes of investigations are often only – if at all – slightly indicating the possibility of a mental syllabary and unambiguous results remained rare.

3 Database

In this chapter, I will introduce the speech corpus that provided the speech data for the investigation presented in this work. Then the German probability list will be introduced that provided the statistical background for the syllables examined in this work. At last, I would like to draw the attention to some problems that existed from the beginning and had to be taken into account during this work.

3.1 Speech corpora in general

In order to investigate coarticulation there is need of a speech corpus. A speech corpus is a collection of spoken language. Nowadays speech corpora are typically stored digitally, because storage and access are much simpler than for corpora that are stored analogically, e.g., on tapes, and digitally stored data allows searches on a corpus that go beyond the possibilities of analogically stored corpora. These properties are directly linked to the usability of a corpus.

A speech corpus for the purpose of scientific research should fulfil some basic requirements.

- **Size**
  The speech corpus should attain a minimum size of tokens to be representative of a language or a dialect. To obtain a speech corpus that should allow for certain phenomenons in speech such as coarticulation, it is not sufficient to simply agglomerate recordings of any kind. Rather the speech data for the recordings has to be chosen thoroughly and adequately.

- **Coverage**
  So, along with size the coverage of a corpus is equally important for representativeness. The measurement of the coverage quality however is dependent on what shall be investigated by means of this corpus or what will be its intention. A closed-domain corpus for example has as primary objective to cover the full vocabulary and special speech properties of the respective domain. A domain-independent corpus in contrast must be designed to cover a vocabulary from which ideally any possible word of the respective language can be generated since every language has theoretically an infinite vocabulary. Furthermore statistical effects like LNRE (Large Number of Rare Events[3]) must be considered as well during corpus construction. One typical application of a speech corpus is being the basis for a speech synthesis system: either as pool of segments for a unit selection or as basis for the extraction of diphones for diphone synthesis. Another reason to build up a speech corpus can be to attain a certain section of speech as evidence for language habits over a specific period of time or as database of a specific dialect.

---

[3] The term LNRE denotes a language phenomenon: while there are only a few high frequent words (or syllables) in a language (or a corpus), there are, on the other hand, lots of very infrequent words that often remain unknown or unseen and therefore add up to a real problem with regard to open-domain systems (see Möbius (2003) for a discussion on coping with LNRE in TTS systems). The LNRE effect is directly connected to the presumed infinity of a language's vocabulary.

- **Annotation**
  A speech corpus that shall be of use in any way beside being a mere collection of speech has to be annotated in some way in order to provide accessibility. Annotation may be restricted to labelling the content of a whole paragraph or sentence. But standard annotation includes a more fine-grained annotation like annotation of words and syllables as far as annotation of segments and phonemes.

The investigation on coarticulation of syllables with distinct probabilities in this work has been carried out on a large speech corpus that shall be called *SmartKom corpus* from here on.

3.2 SmartKom corpus

The SmartKom corpus has been designed during the SmartKom project (Schweitzer et al., 2006). The aim of the SmartKom project was to develop a multimodal dialogue system that allows mixed-initiative interaction between a user and the SmartKom system. As a multi-modal dialogue system the SmartKom system has a built-in speech synthesis system which "is realized by a corpus-based unit selection strategy" (Schweitzer et al., 2006: 411). The corpus behind this unit selection solution is in particular suitable for the investigation described in this work for it has been designed with respect to a special requirement demanded by the objective of the SmartKom project: to have a database for unit selection that is appropriate to closed-domain speech processing as well as to open-domain vocabulary. "The vocabulary is therefore unlimited, although it is biased toward domain-specific material." (Schweitzer et al., 2006: 422). The requirement of domain-independence led to a database content that took LNRE effects into account (cf. footnote 3 p. 8). This was achieved by enlarging the closed-domain corpus with diphones in different contexts. The result was a corpus with full diphone coverage.
The recordings were made with a professional male speaker. The corpus has a size of 160 minutes and it contains 2601 sentences and 17489 words (Möbius, 2004). The speech data has been assigned labels of the following content:

- word level: information about the word class
- syllable level: information about word accent (stress), phrase boundaries and syllable position (initial, mid, final or single)
- segment level: information about position within a syllable (onset, coda)
- intonation: intonation was annotated by use of a further developed version of ToBi (Baumann et al., 2000)

Furthermore the SmartKom corpus was enlarged by English words which was required by some domains, e.g., cinema, to be able to process for example foreign movie titles and names.

3.3 German syllable probability list

In her dissertation, Karin Müller (Müller, 2002) applied a probabilistic classification technique, viz. EM-based clustering, for an automatic classification of German and English syllables. She exemplified her approach by the induction of 3- and 5-dimensional probabilistic syllable classes (Müller, 2002: 45). A 3-dimensional syllable class considers a syllable's onset, nucleus and coda. A 5-dimensional syllable class considers moreover two more syllable properties: a syllable's stress and a syllable's position in the respective word. During her work, Müller derives a probability distribution, among other things, for the 5-dimensional syllable classes and their constituents (for more information see Müller (2002)). For German - from the probabilities of the single syllable constituents - probabilities for the whole syllables were derived independently of Müller's actual work. This yielded a list of German syllables together with their respective probability which serves as basis for the investigations in the present work. The German syllable probability list will be referred to as *SP list* from here on. The list comprises a total of 330,401 different syllables. The syllables are annotated with their respective probability, syllable stress and syllable position. The probability rate reflects the probability of a syllable in its specified environment varying theoretically from 0 to 1. Although the probability rates allow for fourteen decimal spaces, a few syllables do have nonetheless a probability smaller than that which led to a representation as zero.

3.4 Data sanity and preprocessing

	Onset	**Nucleus**	**Coda**	**Position**	**Stress**
class 46 0.007	NOP[E] 0.630 ts 0.256 d 0.074 k 0.024 m 0.007 t 0.002 kv 0.001 n 0.001	E 0.990 e: 0.004 i: 0.003	nt 0.602 t 0.128 n 0.092 l 0.053 m 0.037 Rn 0.02 N 0.013 s 0.012 pt 0.010 Rnst 0.005 ks 0.004 Rns 0.003 nts 0.003 Rt 0.002 Rts 0.001 Rp 0.001	INI 0.627 FIN 0.331 MED 0.040	STR 0.596 USTR 0.403

Table 1: *Class no. 46 out of fifty classes from Müller (2002: 57)*

The size of the databases necessitated to take a close look on the data before the investigation. As an important matter for the whole outcome of the investigation it was revealed that the German syllable probability list seems to contain false data to some extent. The assumed explanation is as follows: During the computation of the probabilities for whole syllables together with their stress and position[4] it seems that some onsets and codas had been assigned to nuclei which led in each case to an impossible onset-nucleus-coda combination. I will explain this in more detail: Müller's algorithm did a classification on syllables,

---

[4] This computation generated the SP list. It was taken out independently from the original dissertation by Jörg Mayer.

i.e., nuclei that have some properties in common formed a common class, other nuclei formed other classes. For example when the algorithm was set to build up 50 syllable classes, class no. 46 (cp. table 1) comprised the nuclei /E/, /e:/ and /i:/ in certain contexts (Müller, 2002: 57). By definition of Müller (2002: 47), syllables had to consist at least of a nucleus (i.e., one or more vowels), an onset and a coda. Thus empty nuclei were not allowed whereas empty onsets and codas were. To be able to keep track of empty onsets and codas and to assign them probabilities they were coded as "NOP[*]", where * is the respective nucleus of the syllable (Müller, 2002: 50). To return to the above mentioned class no. 46 (cf. table 1) it may be asserted that it contains eight different onsets among which one is "NOP[E]"; neither "NOP[e:]" nor "NOP[i:]" are present in the table (syllable parts below a certain threshold did not get an entry in the table). There is no evidence in Müller's work that a certain NOP[x]-onset or NOP[x]-coda is allowed to merge with nuclei other than [x]. This is the problem that arises when surveying the German syllable probability list that emerged from Müller's work. Table 2 shows a cutout of the original SP list. According to the definition of the NOP[*]-coding, the * always gives the nucleus that belongs to the NOP in front of the brackets and not for example a class of nuclei (Müller, 2002: 50). Therefore all rows except for the first one in table 2 are unacceptable by definition since the bracketed nucleus /E/ does not correspond to the proper syllable nucleus (in row 2 it is /Y/, row 3: /9/, row 4: /a/, row 5: /e:/, row 6: /i:/). In the last row onset *and* coda of the syllable are empty, therefore the row holds two NOPs. Although the first NOP matches the /E:/-nucleus, the second one does not. So the whole row is as corrupt as the other five. The consequence from this discovery was to completely remove syllables from the database where (one of) the NOP nucleus did not correspond to the syllable nucleus. It was a total of 6,461 syllables that had to be removed which is approximately 1.96% of the entire SP list.

probability	**onset nucleus coda**	**position**	**stress**
0.00001276154649	NOP[E]Ek	MED	USTR
0.00000008651644	bRYNOP[E]	MED	USTR
0.00000000005698	NOP[E]9lf	FIN	STR
0.00000001278217	NOP[E]alf	FIN	STR
0.00000001356055	NOP[E]e:ks	FIN	USTR
0.00000026403622	NOP[E]i:t	FIN	USTR
0.00000010747239	NOP[E:]E:NOP[OY]	ONE	USTR

Table 2: *Cutout of the SP list*

After having deleted the false entries from the SP list, some main differences between the SmartKom database and the SP list required preprocessing steps. Table 3 shows on the one hand a generally accepted SAMPA[5] transcription for German as it has been used in SmartKom. On the other hand it reveals the basic differences of the SP list towards SmartKom. (Where there is no entry in the SP list column, SmartKom and SP list transcriptions are identical.)

- As could be seen above, the SP list contains text strings, viz. "NOP",

---

[5] Abbrev. for 'Speech Assessment Methods Phonetic Alphabet', an ASCII-based alphabet, http://www.phon.ucl.ac.uk/home/sampa/index.html

SmartKom	example	SP list	SmartKom	example	SP list
p	Pein		a:	Stab	
b	Bein		a	hat	
t	Tisch		E:	Käse	
d	Dorf		e:	Beet	
k	Kohle		E	Bett	
g	gut		i:	Liebe	
m	Mann		I	Kind	
n	Name		o:	Boot	
N	Rang		O	Kopf	
l	Lied		u:	Hupe	
R	Ruf		U	Butter	
f	Fahrt		2:	tödlich	
v	wild		9	Knöpfe	
s	Gras		y:	süß	
z	Suppe		Y	dünn	
C	ich	x	OY	Freund	
x	Dach	x	aU	Haus	
S	Schule		aI	Reifen	
h	Hund		@	Gewinn	
Z	Genie		6	Wetter	@R
j	Junge		?	?ich	—
ts	Platz				
tS	Matsch				
pf	Pferd				
dZ	Gin				

Table 3: *German phonetic transcription in SmartKom and the SP list*

directly attached to the rest of a syllable when there is a missing onset or coda.

- The SP list was not constructed with respect to glottal stops, according to Celex conventions (Cranfield, 2002), so no glottal stops are present in the list. The SmartKom files, however, differentiate between syllables with and without glottal stops; they are transcribed as /?/. Glottal stops occur only syllable initial in German.
- The SmartKom files contain the allophonic transcription of the phoneme /C/, viz. [C] and [x], whereas the SP list confines to the phone [x] which is also a Celex notation form (Cranfield, 2002). (It is possible to do so, for the distribution of the two allophones in German is complementary[6]).
- The SmartKom files occasionally contain syllables with foreign phonemes (which can be put down to the fact that the SmartKom corpus was extended with for instance English words, see section 3.2 for details) such as /D/ and /A/. The SP list, since it incorporates only German syllables, has no occurrences of foreign phonemes.
- Müller adopted in her work the handling for vocalised /r/ in syllable final position from the Celex conventions, too (Cranfield, 2002): it is coded /@R/ (Müller, 2002: 48) which is not at all compatible with SmartKom where it is coded as a single phoneme /6/.

---

[6] The uvular allophone [x] of the phoneme /C/ follows back vowels such as /u/, /a/, /o/ and /au/, other speech sounds are followed by the palatal allophone [C] (see Pompino-Marschall (2003: 265)).

I overcame the mentioned problems by the following solutions. From SP list syllables which contained one or more NOP entries, the NOP entries themselves were cut directly off the syllable, thus only the nucleus was processed further. Glottal stops were removed from the SmartKom syllables since otherwise the syllables would not match any SP list syllable. For the allophone problem I decided to transcribe all /C/-allophones of the SmartKom syllables by [x] according to the SP list convention, since a reversion to [C] of the SP list phonemes in the adequate vowel surrounding would have been much more effort. Syllables with foreign phonemes are ruled out automatically during the merging process described in chapter 5.1. The /@R/ problem from the SP list was overcome by changing the transcription to /6/ according to the SmartKom corpus convention.

4 Methods and tools for analysis

The following paragraphs provide an overview of the programs and statistical methods used to display and interpret the results.

4.1 Festival

Festival is a speech synthesis system developed in 1996 by Alan Black in the first place and has been developed since to the current version 1.95. Festival "offers a general framework for building speech synthesis systems as well as including examples of various modules" (Black et al., 1999). At the IMS, Festival has been further developed to synthesise German speech. For the purpose of the present investigation, the so called *Festival Speech Tools* had been used (Taylor et al., 1999). These speech tools, however, have been developed independently of Festival, but they are a major part of the Festival system since they provide the accessibility of the Festival data and the corpora.

4.2 ESPS get_f0 & formant

These two tools belong to the Entropic Signal Processing System which is a signal processing suite including a speech analysis toolkit for UNIX systems.
get_f0 is a command-line-based program that takes one or more speech recordings, identifies voiced areas in the recordings and writes out, among other things, the respective fundamental frequency of the phonation.
formant is also command-line-based but can calculate beside the fundamental frequency speech formants up to the seventh one. Both *get_f0* and *formant* deliver their results in binary code. Therefore another ESPS tool is required to decode the result files into normal text files: this command line tool is called *pplain*.

4.3 Box plots

A box plot is a diagram that is used to graphically depict numerical data basing on a five-number summary. The five-number summary comprises median, lower and upper quartile, minimum and maximum. A box plot therefore consists of a box, a median-bar, two whiskers and, dependent on the data, sometimes outliers. The box indicates by its size and location in the diagram where 50% of the data are allocated; it therefore comprises the second and third quartile of the data. The first and the fourth quartile are indicated by the whiskers which are lines that are drawn from the box to the maximum endpoint and to the minimum endpoint respectively. The length of a whisker may not exceed 1.5 times the box' size. Data values that go beyond this maximum or minimum are displayed as points beyond the extremes, so-called outliers.
Box plots can help to indicate variance, skew and outliers of a data set. The median is the value that divides the higher half of the data values and the lower half of the data values. It is relatively robust towards outliers in contrast to the mean score. The box plots used in this work have been generated with R (R, 2003), a free software environment for statistical computing and graphics.

4.4 Analysis of Variance (ANOVA)

ANOVA is the abbreviation for Analysis of variance. It is a collection of statistical models to analyse statistical variance. An ANOVA is suitable for example to investigate whether presumed independent factors like subjects in an experiment behave really independent from the outcome of the experiment, that is if chance can be assumed. For the actual research, the ANOVA that is implemented in R (R, 2003) is used.

5 Investigation

In this chapter, the course and outcome of investigations on coarticulation in rare and frequent syllables will be presented with regard to the theory of a mental syllabary. The first section will introduce the syllables on which the research will be carried out. Then, in section 5.2, the first investigation will be presented which is about syllable duration. In section 5.3, influences of fundamental frequency will be taken into account, and in section 5.4 formant transitions will be investigated.

5.1 Approach

In section 3.4 the preprocessing was explained that was needed to ensure that the two databases – SmartKom syllables and SP list – are comparable. For all investigations, minimal pairs are used for they are the most comparable units, while at the same time failures basing on syllable differences can be excluded to a large degree. The data coverage of the SmartKom corpus allowed to choose stress as single distinctive factor. (Otherwise one would have had to search for only similar syllables or had to neglect word position information.) Minimal pairs in this work are therefore defined as follows: they have to be identical in their positional information and in their syllables themselves. The stress information, however, *must* be different from one another, and the probability has to be disjunctive in the meaning that it is frequent for one part of a minimal pair and rare for the other part. Table 4 should make this definition clearer: apart from the probability, stress is the only differentiating factor between each of the syllable pairs. The probability, that divides the syllables into rare and frequent and therefore is the main basis of the whole work, is defined as follows: the mean of all syllable probabilities from the SP list was calculated and used as point of separation between rare and frequent. With this approach I followed Levelt & Wheeldon (1994) who orientated the syllables for their tests by the mean of the frequency distribution of all syllables. But in contrast to the present work they set up the point of separation a little higher than their calculated mean value. It has to be emphasised furthermore that Levelt's and Wheeldon's calculation based on real corpus frequency (they used the CELEX[7] database for Dutch), whereas the present work's probabilities derive from a statistical approach by Müller (2002), cf. section 3.3. The mean value calculated

probability	**syllable**	**position**	**stress**
0.00000737324231	aUf	MED	USTR
0.00025011182370	aUf	MED	STR
0.00006728772923	baI	FIN	USTR
0.00022633920458	baI	FIN	STR
0.00000376030469	baI	INI	USTR
0.00077711016596	baI	INI	STR
0.00004299399235	ba:	INI	USTR
0.00016603535533	ba:	INI	STR

Table 4: *Extraction from the minimal pair list*

---

[7] The Centre for Lexical Information (CELEX), Max Planck Institute, The Netherlands

	German syllable probability list	**SmartKom syllables**
stress	STR vs. USTR	stress 1 vs. stress 0
position	INI MED FIN ONE	(initial mid final single)[8]
probability	yes	n.a.
time index	n.a.	yes
size	330,400 syllable types	33,821 syllable entries 4,812 distinct syllables

Table 5: *Table of information stored with the syllables*

for the present work is 0.0001339165. Compared to the total amount of 33,821 SmartKom syllable entries in table 5 it may seem as if 4,812 distinct syllables are too few to make up a whole database, but it must be taken into account that the SmartKom database contains multiple occurrences of some syllables e.g., to be able to provide them in different syllable surroundings and in different prosodic contexts for unit selection (cf. section 3.2).

There were three major demands that had to be fulfilled to attain minimal pairs like those in table 4 and before investigation could take place. Firstly, a list had to be generated that contained all distinct SmartKom syllables. This list had then to be searched for syllables that in principal could make up a minimal pair, i.e., any two syllables that have a disjunctive stress distribution. Secondly, the thus attained syllables had to be searched by means of Festival in the SmartKom database again in order to reveal their positional information. Thirdly, the syllables had to be merged with the syllables of the SP list to get additionally probability information. At last, minimal pairs had to be extracted. To be able to follow the necessity and order of the processing steps, it is convenient to know what information is provided by each of the databases and how it is stored together with the syllables. An overview is given in table 5. While the SP list comes in list format by default, the SmartKom database consists of many different single files, files that contain e.g., the audio data in wave format[9], the syllable information or the syllable duration. Festival (cf. chapter 4.1) makes the information from the different SmartKom data files accessible. As can be seen in table 5, SmartKom information about syllable position can not be drawn directly from the SmartKom syllable files. So, the first step was to generate a list of the SmartKom syllables, removing syllables that would never find a minimal partner and thus reducing early the mass of data that has to be processed further. The next step was requesting Festival for a syllable's position for each single SmartKom syllable. The list thus contained for each syllable now a syllable's positional information in addition to its stress. Having the SmartKom syllables now in list format, like the SP list, together with the necessary information, the next step was to merge the two lists with respect to each syllable's stress and its position. Thus a new list was attained that consists of all SmartKom syllables that could also be found[10] in the SP list in the same environment. Details are shown in table 6: every syllable entry is unique (independent from the last column). The list consists of 1,705 syllable entries on the

---

[8] the information on syllable position is not directly stored within the syllable files

[9] a common audio format

[10] English syllables did not match of course since the SP list consists only of German syllables, cf. section 3.3

probability	syllable	position	stress	origin
0.00000000036754	pfalts	ONE	STR	ms_sk626_1
0.00000003604831	pfan	MED	STR	ms1630_1
0.00000000088587	pfats	FIN	USTR	ms392_1
0.00000885361141	pfe:6	INI	STR	ms1297_1
0.00000003078203	pfe:6	INI	USTR	ms360_1

Table 6: *Cutout of the candidate syllables after the fusion of SmartKom syllables and SP list*

whole. The last column exists only for further processing purposes and gives the name of *one* of the SmartKom files that contained the respective syllable in the respective environment. From this list, now, the minimal pairs can be discovered and extracted by imposing the calculated mean score on it. It divides the list into two probability-separated parts, and the syllables now defined as rare are then compared to the syllables defined as frequent in order to find matching minimal pairs. Thus 234 syllables were discovered that sum up to 117 minimal pairs. The full minimal pair list is given in the appendix (A.1). A small cutout can be viewed in table 4. The minimal pair list has the following properties:

- It consists of 117 minimal pairs which comprise 101 distinct syllables (of course, with regard to their environment all minimal pair entries[11] are distinct to each other).
- A total amount of 2,767 occurrences was retrieved for the whole minimal pair list from the SmartKom database.
- 1,764 of these occurrences were for word initial position, 545 for word medial position, 112 for final position, and 346 occurrences could be retrieved for the monosyllabic words.
- 938 occurrences of minimal pair entries were discovered for unstressed environments, 1,829 for stressed environments.
- Regarding the probability distribution, 2,159 of the occurrences could be found for frequent minimal pair entries and 608 for rare/infrequent minimal pair entries.

These minimal pairs are used for all of the following investigations. The next chapter (5.2) will be concerned with the examination of the syllable duration which is then followed by a chapter (5.3) on examination of the minimal pairs' fundamental frequency, and the last investigation chapter (5.4) will take into account differentiating values of the first three formants.

5.2 Syllable duration

Syllable duration is very suitable to investigate differences in syllables with regard to the mental syllabary theory. Fast spoken syllables have a shorter duration than the same syllables spoken in a slower manner. Thus, if a mental

---

[11] *entry* refers to a syllable together with its respective environment, in contrast to *syllable* which means the syllable alone.

syllabary is engaged in producing speech, frequent syllables should exhibit a tendency to be faster produced for the retrieval from the syllabary is supposed to be faster than the assembly of a syllable. Therefore frequent syllables can be pronounced faster and thus should be shorter than rare syllables. The results from this investigation are divided into the factors that come along with the syllables, viz. stress, position and above all syllable probability.

5.2.1 Method

Syllable duration for the minimal pair syllables was again attained by the use of Festival. For each syllable in the minimal pair list, the whole SmartKom database was searched for any occurrence of the respective syllable in its corresponding environment. The command line call was as follows (in one line):

```
/usr/local/bin/festival --script <script> -feats
"(syllable_duration name position_type)" -relation Syllable
```

Every detected occurrence was then recorded to files for later statistical analysis. Because of the differing database frequencies of the distinct syllables, the frequency recordings varied from one occurrence (the syllable /taI/ in word initial position in either of the two stress variations where one is frequent and the other one is rare) to over two hundred occurrences (264 for the frequent syllable /aI/ in initial stressed position).

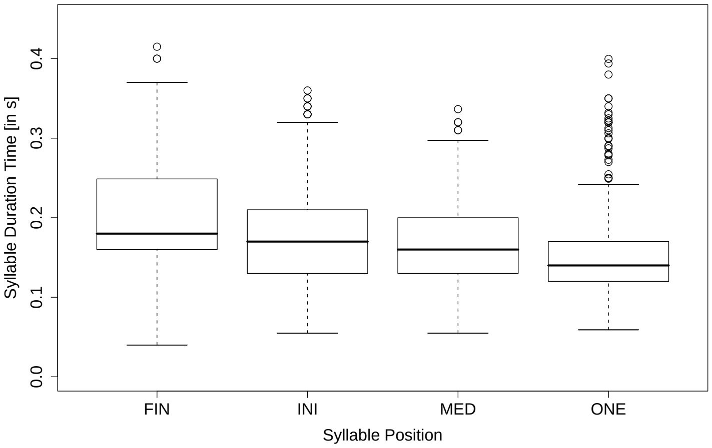


Figure 1: *Relation between syllable duration and position*

5.2.2 Results

Figure 1 shows the distribution of the average duration time classified according to the positional information. The graphic shows quite clearly the fundamental differences between the position of syllables in a word. The longest syllable duration in average have final syllables with a mean of 204ms. Initial syllables have a value of 172ms, mid syllables have a mean of 166ms and monosyllabic words have a mean of 158ms. Overall syllable duration varied from 40ms to 415ms, the mean was 174ms. The shortest duration with 40 milliseconds had the frequent syllable /a:/ in final and unstressed position. It stemmed from the word *Media* /mi:di:a:/. Since this is a foreign word, I looked for the next shortest syllable which was /?aI/, stemming from the word *eine* /?aIn@/ (a). It was in initial stressed position, had 54ms in duration and was frequent – which is no surprise since it is an determiner. The longest duration with 415 milliseconds was recorded for the frequent syllable /baI/ in final and stressed position, stemming from the word *vorbei* /fo:6baI/ (over, gone). I observed an interesting tendency in this outcome: Final syllables – when they are stressed – tend to have a longer duration than stressed syllables in other positions. This is strongly supported by my data (cf. figure 2): stressed final syllables have a duration mean of 232ms, whereas stressed syllables in other positions have a mean of only 179ms. Breaking the other positions of stressed syllables down produces the following results: Mid ones have the second highest mean with 196ms, followed by initial ones that have a mean of 182ms. The shortest duration have monosyllabic stressed words with a mean of 158ms. This observation may not surprise as it seems comprehensible that – in a language like German where final syllables are normally unstressed – stressed final syllables are lengthened deliberately to lay special emphasis on them.

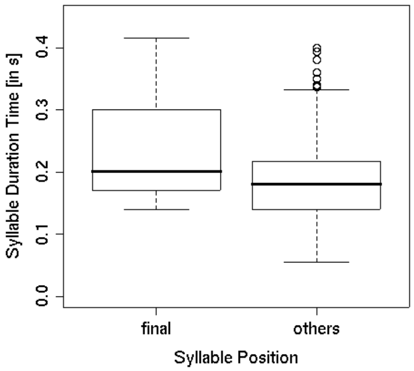


Figure 2: *Syllable duration in relation to stressed syllables' position: final stressed syllables have a longer duration in average than stressed syllables in other positions.*

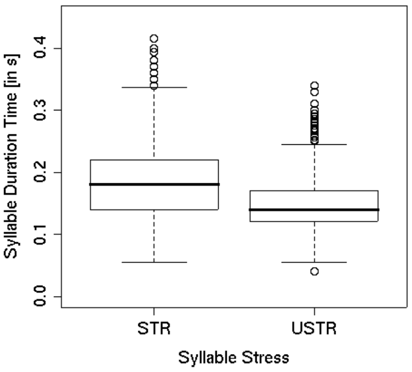


Figure 3: *Relation between syllable duration and syllable stress*

Figure 3 shows the relation between syllable duration and syllable stress. It is unambiguous that there is a vast difference between stressed (STR) and unstressed (USTR) syllables. Stressed syllables have a mean in duration of approximately 180ms whereas unstressed syllables are much shorter in average with a mean of 151ms. This observation directly supports what had been widely reported throughout the literature: namely that unstressed syllables have a shorter duration in general.

In Figure 4, the relation between duration and syllable probability is presented: frequent syllables, having a median of 168ms seem to have a longer duration

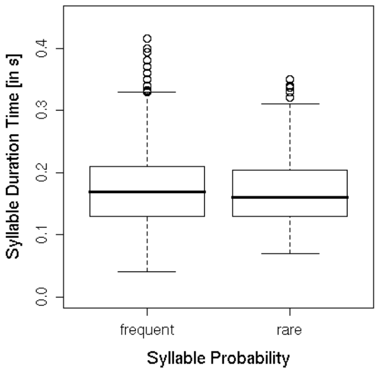


Figure 4: *Relation between syllable duration and syllable probability*

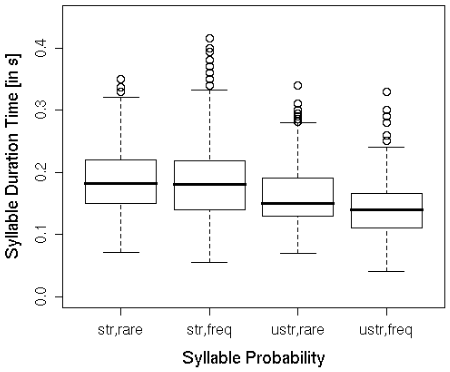


Figure 5: *Relation between syllable duration and syllable probability with regard to stress*

than rare syllables which have a median of only 160ms. If that observation was true, the expected effects of an assumed mental syllabary theory would have been turned into their contrary. A look on the means of the data however imparts a different view: frequent syllables have a mean of 170.3ms and rare syllables have a mean of at least 170.7ms. Since it is only fractions of milliseconds, the outcome is nonetheless disappointing. But what about differentiating between stressed rare and frequent syllables and the ones that are unstressed? Figure 5 gives a demonstration of this issue. The obvious marks are again the medians: in numbers, frequent stressed syllables had a median of 180ms which is a bit less than the median of rare stressed syllables which was 182ms. The median of frequent unstressed syllables was 140ms whereas rare unstressed syllables had a median of 150ms. The means give yet a much clearer indication of the distribution: Frequent stressed syllables had a mean of 180ms, rare stressed syllables 188ms. For unstressed syllables the distance is really broad with a 24ms difference: frequent unstressed syllables had a mean of only 140ms whereas unstressed rare ones had a mean of 164ms. Nonetheless, an ANOVA reported the probability factor as only marginally significant (see below).

The median for the syllable duration of all syllables was 164ms, the mean was 170ms. For the following syllable duration factors, an ANOVA analysis (cf. section 4.4) reported high significance: position (p-value $\ll .0001$, F-value $= 34.4$) and stress (p-value $\ll .0001$, F-value $= 293.2$). The syllables themselves are also highly significant (p-value $\ll .0001$, F-value $= 23.9$) but this significance is not orthogonal to that of position and stress since it is very likely that the syllables' significance interacts with the other factors. As mentioned before, syllable probability is only marginally significant (p-value $= .086$, F-value $= 2.9$).

Additionally to the data so far I would like to take a look at the extremities of the syllable duration results: Table 7 gives the ten syllables that had the shortest duration and the ten syllables that had the longest duration. Besides, the table shows the words from where the syllables originate. Similar syllables stem

duration	syllable	position	stress	probability	word
40	a:	FIN	USTR	frequent	Media (1)
54.7	aI	INI	STR	frequent	ein (2)
58.4	E6	MED	USTR	frequent	Reiseerleichterung (3)
58.4	E6	MED	USTR	frequent	Wählerbetrugs (4)
59	e:6	ONE	STR	frequent	er (5)
59.9	aI	INI	STR	frequent	ein (2)
59.9	aI	INI	STR	frequent	ein (2)
59.9	e:6	ONE	STR	frequent	er (5)
59.9	e:6	ONE	STR	frequent	er (5)
59.9	e:6	ONE	STR	frequent	er (5)
350	Spi:	INI	STR	frequent	Spielen (6)
350	RUm	FIN	STR	rare	herum (7)
360	fa:	INI	STR	frequent	Fahrerin (8)
370	tsu:	FIN	STR	frequent	dazu (9)
380	fo:6	ONE	STR	frequent	vor (7)
394	man	ONE	STR	frequent	Mann (10)
399	fo:6	ONE	STR	frequent	vor (7)
400	fy:6	FIN	STR	frequent	dafür (11)
400	tsu:	FIN	STR	frequent	hinzu (7)
415	baI	FIN	STR	frequent	vorbei (11)

Table 7: *A list of the ten longest and the ten shortest syllables and the word they stem from. The bracketed numbers give additional information. 1: "media"; 2: indefinite determiner; 3: "facilitation of travelling"; 4: "voter fraud"; 5: male pronoun; 6: "playing"; 7: verb particle; 8: "female driver"; 9: "in addition"; 10: "man"; 11: preposition*

despite of their seeming identity from different sentences or sentence positions. Three observations can be made from this data excerpt: Firstly, there is no rare syllable among the short syllables. Secondly, there is no unstressed syllable among the long syllables. Thirdly, the distribution of the sort of syllables is interesting: The short syllables comprise mainly very frequent German words, like pronouns and determiners but none of the word classes that are mainly comprised by the long syllables part, namely prepositions and verb particles. We will not go deeper into details here as these observations presumably are not directly linked to the mental syllabary effect.

5.2.3 Discussion of syllable duration

From the data found from syllable duration measuring, it can be stated that the two factors stress and position exhibit high significance when it comes to syllable duration. The most significant factor found was stress, followed by position, whereas syllable probability is only marginally significant. Regarding the mental syllabary theory, the data found exhibits no clear tendency that there is a difference between rare and frequent syllables. It is expected that during speech production more accuracy is employed when building syllables on-line (as it is with rare syllables) than with syllables that are readily stored. This should result in duration differences. But since no bigger difference between rare and frequent syllables could be found, the data does not assure us whether the observed difference can be attributed to a mental syllabary effect or not.

5.3 Voicing probability

Fundamental frequency denotes the lowest frequency that is evoked by vibration of the vocal folds. Whenever there is voicing in an utterance, a fundamental frequency value can be detected. Consequently, if no fundamental frequency can be measured, it is probable that no voicing is intended or devoicing takes place. The fundamental frequency of a male normal speaker's voice may range from 100 to 220 Hz (Hakkarainen, 1995: 33), although the range of a single speaker is a lot smaller than the full, generally possible range. In general, one distinguishes between voiced consonants (e.g., /b/, /z/, /v/) and voiceless consonants (e.g., /p/, /s/, /f/). In particular situations, devoicing of a phonologically voiced consonant can take place – this event often follows some phonotactical patterns in German. Devoicing in German takes place preferably at word or syllable boundaries or before and after voiceless consonants. For example, /z/ that is phonologically voiced may be devoiced before /t/ like in *reisen* (travel) vs. *reiste* (travelled). Especially in fast speech, devoicing occurs more often than in slower speech. This is said to be due to efficiency needs of the vocal tract where the articulators have to change more quickly from one point of articulation to another the higher the speech rate is. An exception are fricatives and stops which are likely to be devoiced the longer they are. Devoicing effects are counted to coarticulation and thus it should be possible to detect differences between rare and frequent syllables. Therefore, fundamental frequency of the minimal pairs will be investigated, looking for coarticulatory effects that can be ascribed to a mental syllabary effect if a difference between frequent and rare syllables can be detected.

5.3.1 Method

Fundamental frequency was measured for all of the minimal pairs from chapter 5.1. A Perl script was used to grep the names of the files that contain in each case the respective syllable in the demanded stress environment. For each of the tracked syllables, get_f0 (cf. chapter 4.2) was then called up with the following command:

```
get_f0 -i 0.01 <file>
```

As one can see from the command, get_f0 was invoked with a frame rate of 10ms, i.e., the fundamental frequency was measured every ten milliseconds of a recording. The result files are binary and had to be converted by the use of `pplain` which is also an ESPS command line tool (cf. chapter 4.2). The resulting file is a list of columns of digits, every line indicating a ten milliseconds step in the recording. Therefore, it had to be retrieved which lines correspond to the syllable in the recording of whom the values are wanted. This was resolved by the same Perl script which subsequently processed the fundamental frequency values at ten equidistant time intervals and computed the mean values of these values for a normalised time frame. It is useful to generalise all syllables' fundamental frequency values over a normalised time frame, for it makes the fundamental frequency comparable when all syllables are virtually having the same duration. The presentation of voicing probability over a normalised time is called a *voicing profile* (cf. Möbius (2004)).

5.3.2 Results

Results were obtained for all minimal pair syllables. However, in order to get statistically reliable data plots, a threshold was applied concerning the number of occurrences in the SmartKom corpus. To be a representative minimal pair, one partner of the pair had to have at least ten occurrences in the corpus, the other one had to have at least five occurrences, e.g., given the syllable /aIn/ the frequent /aIn/-syllables had to have to occur at least ten times in the corpus and the rare /aIn/-syllables had to occur at least five times in the corpus or vice versa. This 10-5-threshold is applied to prevent minimal pairs with less than five or less than ten occurrences to give the impression that they stand for a really existing distribution, for it can not be made sure that they are. It might as well be that these few occurrences have a completely different distribution by chance than what the real distribution is for their kind. After applying the threshold, 30 minimal pairs were attained. The plots of their fundamental frequency course can be found entirely in the appendix (A.2).

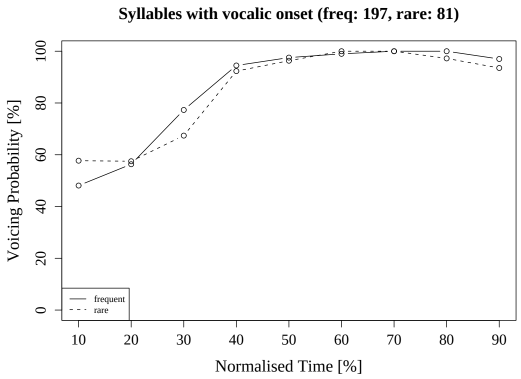


Figure 6: *Voicing profile for minimal pairs that have a vocalic onset and a voiced consonantal coda. These syllables expanded into the plot: /al/, /En/, /In/, /Um/, /Un/, /aIn/.*

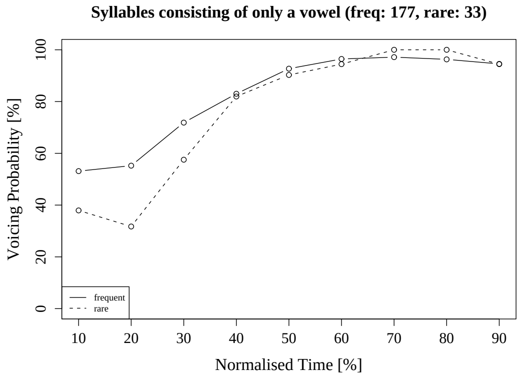


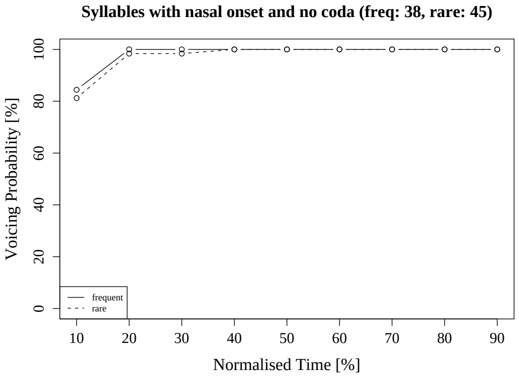
Figure 7: *Voicing profile for minimal pairs consisting of only a vowel or a diphthong. These syllables expanded into the plot: /a:/, /aU/, /E6/.*

At first the fundamental frequency course of syllables will be regarded that have a vocalic onset followed by a voiced consonant. This is displayed in figure 6. The numbers in brackets in the graphic give the number of occurrences in the SmartKom corpus for the frequent syllables, respectively for the rare syllables. As can be seen from the graphic, the lines of fundamental frequency run quite parallel except for the end of the syllables and especially except for the beginning. It seems as if rare syllables generally start more often voiced than frequent syllables. The generally low voicing probability in the beginning is likely to be due to the occurrence of a glottal stop in some of the syllables. Here, frequent syllables tend to begin more often with a glottal stop than rare syllables. It is also a bit remarkable that rare syllables' voicing probability at 30% of the time is considerably lower than that of frequent syllables.

Next I would like to present the results of syllables that consisted only of a long vowel or diphthong. The results can be seen in figure 7. In contrast to the result above, rare syllables here start out considerably more often voicelessly than frequent syllables, namely with a probability of only 40% and then the probability yet decreases, whereas the probability of frequent syllables, which starts out at about 55%, increases. In contrast to figure 6, here rare syllables exhibit a stronger tendency to start with a glottal stop. It can also be seen from the graphic, that voicing probability of frequent syllables never reaches 100% whereas that of rare syllables does at 70% and 80% of the time. This fulfils a prediction of the mental syllabary theory, viz. that rare syllables are generally produced with more effort because they are assembled newly during speech production instead of being retrieved from the syllabary. Thus they are very unlikely to be devoiced since during assembly phonological rules are applied and vowels have the phonological voiced. Frequent syllables, however, are thought of to be retrieved from the syllabary and therefore are much more prone to context-dependent effects like devoicing.

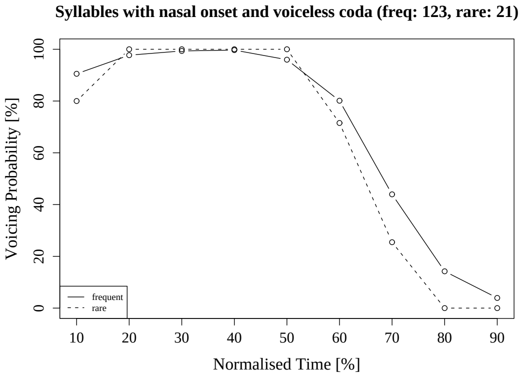
Now, nasals will be reviewed since they are phonologically voiced (Kohler, 1977: 165) and therefore have similarities to vowels. Figure 8 shows the fundamental frequency results for syllables with a nasal in the onset and no coda. There is just a tiny difference between rare and frequent syllables. Rare syllables start out a bit less frequently voiced than frequent syllables. Interesting is the difference to figure 7: both syllable combinations end with a vowel, but nonetheless only the nasal combinations hold the voicing up to 100% until the end of the syllable. The vowel-only syllables, however, do not. But this is likely to be dependent on the following syllable and not on vocalic properties itself. The most striking difference between figure 7 and figure 8 is the first 40% in time: syllables with nasal onset reach nearly 100% at already 20% of the time whereas vowels come close like this only at 40% of time. This is surely dependent on what precedes these syllables. The /mo:/ and /ma:/ values stem only from initial and medial syllables, as well as the /E6/ values. The /aU/ values originate from initial syllables only and the /a:/ values from final syllables only. This may be an important factor for differences in fundamental frequency therefore a distinction with regard to syllable position will be made at a later point in this chapter.

Figure 9 shows the voicing profile for syllables with a nasal onset and a phonologically voiceless coda. It can be seen that rare syllables start out with less probability of voicing than frequent syllables and this effect is larger than in figure 8. Despite of this they reach full voicing probability already at 20% of time and keep it up until the half of the syllable. Frequent syllables on the contrary



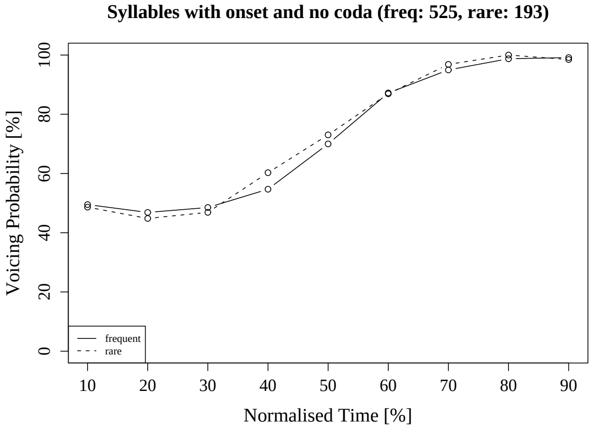
Syllables with nasal onset and no coda (freq: 38, rare: 45)

Figure 8: *Voicing profile for minimal pairs with nasal onset and no coda. These syllables expanded into the plot: /ma:/, /mo:/.*



Syllables with nasal onset and voiceless coda (freq: 123, rare: 21)

Figure 9: *Voicing profile for minimal pairs with nasal onset and voiceless coda. These syllables expanded into the plot: /mIs/, /mIt/, /nIs/.*



Syllables with onset and no coda (freq: 525, rare: 193)

Figure 10: *Voicing profile of minimal pairs with onset and no coda. These syllables expanded into the plot: /baI/, /bi:/, /ga/, /ga:/, /la:/, /ma:/, /mo:/, /Ra:/, /Ro:/, /hE/, /fE6/, /fo:6/, /fO6/, /kaI/, /ta/, /tsu:/, /tsvaI/*

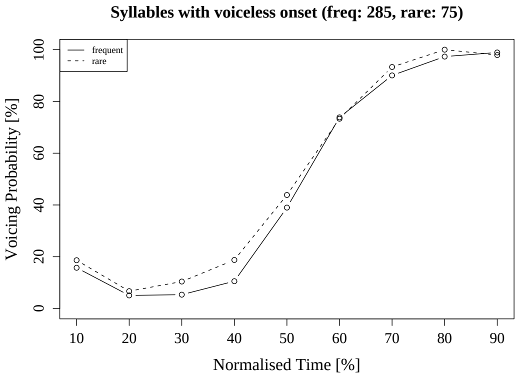


Figure 11: *Voicing profile for minimal pairs with voiceless onset and no coda. These syllables expanded into the plot: /fE6/, /fo:6/, /fO6/, /hE/, /kaI/, /ta/, /tsu:/, /tsvaI/.*

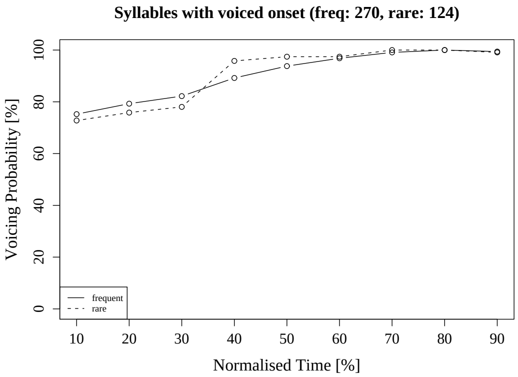


Figure 12: *Voicing profile for minimal pairs with voiced onset and no coda. These syllables expanded into the plot: /baI/, /bi:/, /ga/, /ga:/, /hE/, /la:/, /ma:/, /mo:/, /Ra:/, /Ro:/.*

maintain full voicing probability only between 30% and 40% of the syllable. Then they discard voicing smoothly until the end of the syllable, however, complete voicelessness of the consonant is not always achieved. The contrary applies for the rare syllables: after having passed 50% of the syllable, voicing probability decreases fast and reaches total voicelessness at 80% of time to provide complete voicelessness of the consonant for the rest of the syllable. Here, another effect may have been found that may be ascribed to the mental syllabary. The voicing profile differences allow to assume that during the production of the rare syllables more effort was applied, whereas the frequent syllables show a typical attribute of coarticulatory effects, viz. a carry-over effect of voicing from the vowel to the consonant.

Figure 10 displays the voicing profile for syllables with consonantal onset and no coda. As found with the vowels (figure 6) and the nasals (figure 8), no fundamental difference seems to take place between rare and frequent syllables. Therefore now a distinction is made between different kinds of consonants,

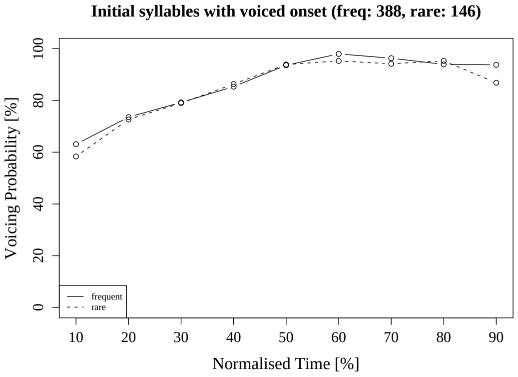


Figure 13: *Voicing profile for initial minimal pairs with voiced onset and no coda.*

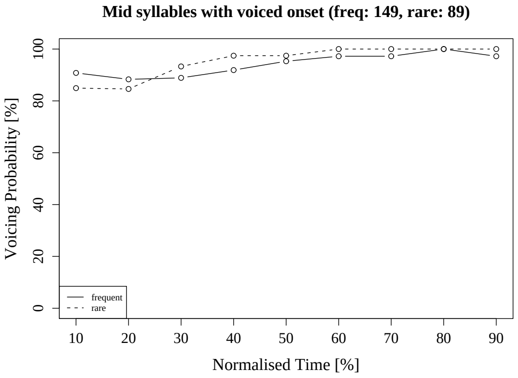


Figure 14: *Voicing profile for mid minimal pairs with voiced onset and no coda.*

namely between phonologically voiced and phonologically voiceless consonants. The syllable /hE/ occurs in both of the voicing profiles because /h/ is said to be underspecified (Kohler, 1977: 160) regarding the feature voice. Both the voiceless consonant profile (figure 11) and the voiced consonant profile (figure 12) show at first sight no big differences. But a closer look uncovers nonetheless some slight differences. In figure 11, apart from starting out with somewhat higher voicing probability, the rare syllables prepare full voicing of the vowel earlier than the frequent syllables. It is likely that here again a carry-over effect of the voicelessness of the consonant causes the later initiation of voicing in frequent syllables. In figure 12 another sort of coarticulatory effect can be assumed to be the reason for the slightly different course of the curve: the consonants are likely to have a duration not longer than a third of the syllable. This can also be seen from the rare syllable where the curve at this point suddenly makes a leap of over 20% in probability straight towards full voicing. The frequent syllable curve, however, runs smoothly across this boundary and this may be an indication for phonetic smoothing between different modes and places of articulation. Besides, it is not quite clear, why rare syllables start out with a lesser voicing probability than frequent syllables, but perhaps an answer can be found when looking at the position the syllables originate from.

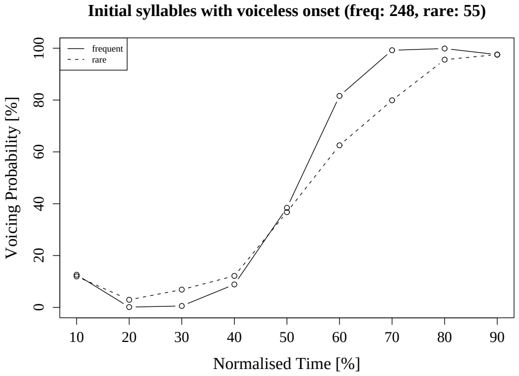


Figure 15: *Voicing profile for initial minimal pairs with voiceless onset and no coda.*

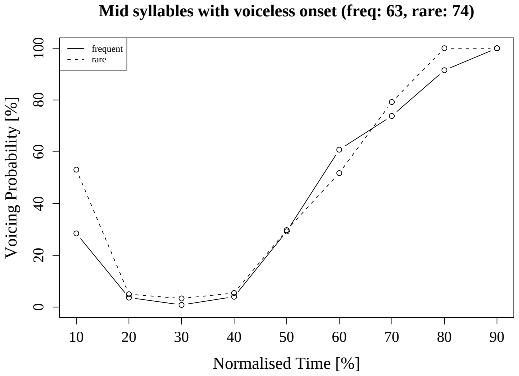


Figure 16: *Voicing profile for mid minimal pairs with voiceless onset and no coda.*

The voicing profiles that are now presented to investigate differences that occur from positional factors of the syllables consist of all syllables that could provide the demanded environment, not only of those that achieved to pass the 10-5-threshold that was mentioned in the beginning. The figures 13 and 14 show the voicing profiles of initial syllables with voiced onset, respectively mid syllables with voiced onset. Final syllables are not regarded since there were not enough occurrences to provide statistical reliability. As can be seen from both figures, rare syllables tend to start with a lower voicing probability regardless of the position. Nonetheless a small difference can be observed from 14: Vowels in rare syllables tend to have a voicing probability of nearly 100% whereas vowels of frequent syllables are generally more voiceless or devoiced and voicing probability increases towards 80% of the syllable. Again, the course of frequent syllables is smoother than that of rare syllables.

Now, let us have a look at initial and mid syllables that have a voiceless onset instead. Figure 15 and figure 16 exhibit some differences in their course of curves. In initially voiceless syllables it can be seen that voicing of frequent syllables starts only after 30% of the syllable. Rare syllables tend to start voicing earlier. A reason may be that frequent syllables transfer devoicing into the vowel, i.e. the first part of the vowel tends to be devoiced and this would be a clue to coarticulation. On the other hand it may be as well that frequent syllables generally have a longer consonantal onset. The reason for the difference in onset voicing can not be assured here to a particular factor . The course of the curve in the second half of the syllable is contradictory to the expectations raised by the mental syllabary theory because in this figure it is the rare syllables that have a much smoother course than the frequent syllables and moreover, they reach a high voicing probability only at 80% of the syllable whereas frequent syllables reach *full* voicing probability already at 70% of the syllable.

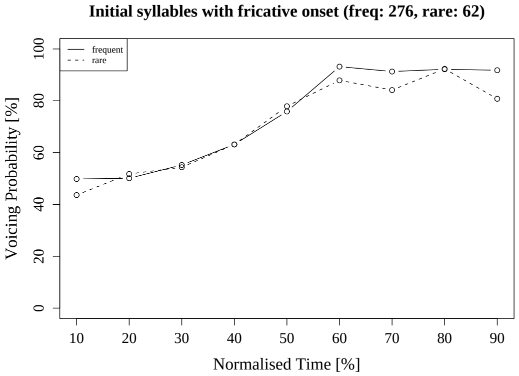


Figure 17: *Voicing profile for initial minimal pairs with fricative onset and no coda.*

In mid syllables (figure 16), however the curve in the second half of the syllable runs as predicted from the mental syllabary theory: rare syllables achieve full voicing probability in contrast to frequent syllables and moreover, they attain it at an earlier point in time. Regarding the beginning of the syllable, rare syllables start out with a much higher voicing probability than frequent syllables. Since voiceless consonants are supposed to start voiced in German (Shih & Möbius, 1998: 6) before becoming voiceless, it is probable that frequent syllables tend to anticipate the voicelessness of the consonant at its very beginning. Thus the lower voicing probability in frequent syllables may be due to anticipatory coarticulation.

Finally, syllables with a fricative onset will be regarded but only those that were in initial and in mid position, for syllables in final position did not deliver enough corpus occurrences to be statistically reliable. In figure 17 initial syllables that begin with a fricative are displayed. Dependent on corpus occurrence, syllables beginning with /f,s,S,v,z/ expanded into the calculation. It can be seen from the figure that rare syllables have a trend to start with a slightly lower voicing probability than frequent syllables. The course of the curves from 60% onwards is irritating with regard to expectations that arise from the theory of a mental syllabary since this time the rare syllables achieve less voicing probability than frequent syllables although forming a vowel. The expectation would be that rare syllables at least achieve full voicing probability regardless of frequent syllables' voicing probability. Here it seems as if the vowel of rare syllables was influenced by anticipatory effects of the following syllable. To provide an explanation for this the calculation was yet adapted to take additionally accentuation as a differentiating factor into account. The result showed that accentuation indeed

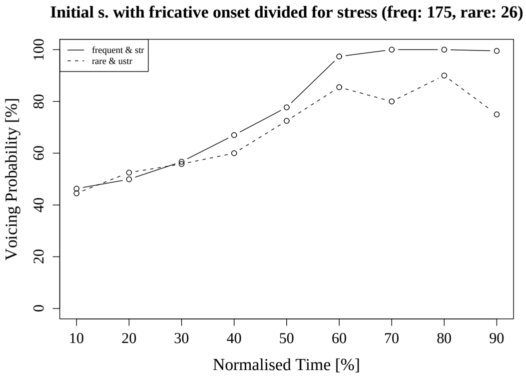


Figure 18: *Voicing profile for initial minimal pairs with fricative onset and no coda. Frequent syllables are stressed, rare syllables are unstressed.*

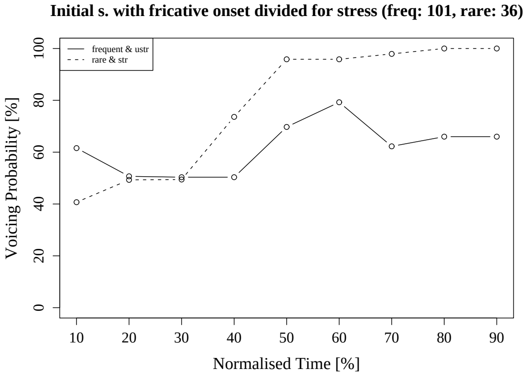


Figure 19: *Voicing profile for initial minimal pairs with fricative onset and no coda. Frequent syllables are unstressed, rare syllables are stressed.*

was a major factor: when only stressed/frequent syllable combination was compared to unstressed/rare syllable combination (cf. figure 18) the difference that could be seen in figure 16 was enlarged in the way that frequent syllables now have a voicing probability of 100% from 60% of the time onwards whereas rare syllables on the other hand lost voicing probability, though not significantly. Now comparing figure 18 to figure 19 there is a striking difference between the outcome of the two calculations. At first sight it seems as if the curves had swapped places. Thus, the second graphic (figure 19) fulfils the expectations mentioned above: the rare syllables take the phonologically more ideal course of curve and the frequent syllables seem to be influenced by anticipatory effects of the next syllable. However, the first graphic (figure 18) provides the contrary to the expectations and that should be accounted for: the only thing that has changed from figure 16 to the latter two graphics is the respect for accentuation differences. Thus it has been made quite clear that stress plays a major role coarticulation and that its role seems to bear more weight than the role of a syllable probability effect alone.

More evidence should be presented for this hypothesis with regard to syllable probability therefore CV syllables beginning with a plosive shall be examined. The hypothesis holds for initial CV syllables as can be seen from the graphics 20 and 21. Rare syllables start in both calculations with a lesser voicing probability than frequent syllables. However, differences can be observed, for rare/stressed syllables (figure 20) start with a lower voicing probability than rare/unstressed syllables (figure 21). Moreover rare/stressed plosives seem to be less prone to anticipatory effects of the following vowel since the voicing probability status is maintained until 20% of the syllable whereas in rare/unstressed syllables voicing probability increases directly after 10% of the syllable. The voicing probability maintenance of the plosives, however, can not be reported for frequent syllables, neither for stressed nor for unstressed ones. Therefore syllable probability seems to be an important factor here regardless of stress. The two graphics also show vast differences for the vocalic part of the syllable. In comparison to rare/stressed syllables frequent/unstressed syllables exhibit a slow and smooth transition from consonant to vowel. Full voicing probability is first achieved at about 60% of the syllable. Rare/stressed syllables on the contrary exhibit a fast and steep transition towards full voicing probability which is reached already at

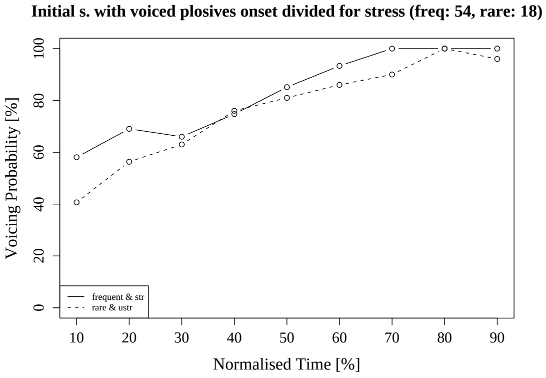


Figure 20: *Voicing profile for initial minimal pairs with plosive onset and no coda. Frequent syllables are stressed, rare syllables are unstressed.*

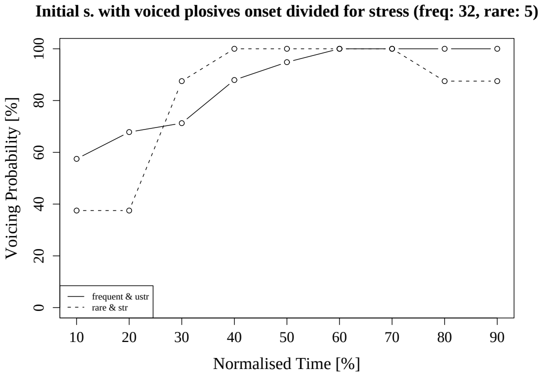


Figure 21: *Voicing profile for initial minimal pairs with plosive onset and no coda. Frequent syllables are unstressed, rare syllables are stressed.*

about 40% of the syllable. In the other distribution (figure 21) a somewhat contrary situation is present. Here the frequent/stressed syllables exhibit a course of curve through the vowel that is a little more ideal than that of rare/unstressed syllables, but both syllables show significant coarticulatory effects. But, similar to the phenomenon that was observed in fricatives, the unstressed syllables is more prone to coarticulation, that is here the rare syllables. Frequent/stressed syllables reach full voicing probability late in the vowel, only at about 70% of the syllable, but rare/unstressed syllable reach it yet later, viz. only at about 80% of the syllable. These are very likely carry-over effects from the voicelessness of the consonant. At the end of the syllable also anticipating effects can be observed. Rare syllables in both distributions tend to become voiceless to some extent whereas frequent syllables in both distributions keep up full voicing. It cannot be assured whether the rare syllables exhibit anticipatory effects to prepare following voiceless segments or whether frequent syllables exhibit a perseverative effect in taking voicing into the next segment.

5.3.3 Discussion of voicing probability

Investigation on voicing probability began with VC syllables with voiceless consonants and syllables that consisted of a vowel or a diphthong. Differences could be found but the reasons for differences could not always be found. Then differentiation was applied regarding the consonant class that were pooled to a calculation. CV syllables, CV syllables with nasal onset and with and without coda and CV syllables with voiced and voiceless onset were presented. For vowel-only syllables a possible mental syllabary effect could be reported which was expressed by the fact that frequent syllables exhibited a trend to a lower voicing probability than rare syllables, i.e., frequent syllables are more often devoiced. A larger difference between rare and frequent syllables could be found for syllables with a nasal onset and a phonologically voiceless coda: a perseverative effect in frequent syllable could be found. They reduced voicing significantly later than rare syllables towards the voiceless consonant. For syllables beginning with a voiceless consonant a carry-over effect was observable regarding the frequent syllable which seemed to keep voicelessness up although the vowel was already being formed. in syllables with voiced consonantal onset the curve of voicing probability ran very smooth in comparison to that of rare syllables, the smoothness indicating that frequent syllables are more free in their variation of voicing onset than rare syllables. After that positional information is taken into account and it can be confirmed that position definitively is linked to coarticulation. For CV syllables in initial and mid position starting with a voiced consonant a trend could be detected for rare syllable to start with a higher voicing probability. Mid syllables additionally show a trend of rare syllables to have a generally higher voicing probability than frequent syllables. Also, the curve of frequent syllables in mid position is smoother during CV transition than that of rare syllables. For syllables starting with a voiceless consonant rare syllables show a tendency to have an earlier voice onset. The reason for this may again be perseverative coarticulation, viz. the consonant transferring its voicelessness over to the vowel. Initial syllables exhibited here a seeming contradiction to a prediction of the mental syllabary theory: rare syllables were affected more by coarticulation than frequent syllables since frequent vowels reached full voicing probability and rare syllables did not. Mid syllables, however, behaved as predicted by the mental syllabary and anticipatory effects could be reported in frequent syllables achieving high voicing probability of the vowel only after rare syllables and never achieving full voicing probability. This investigation was followed by the review of syllables that start with a fricative and this calculation too exhibited courses of curves that are contradictory to the predictions of the mental syllabary theory: frequent syllables showed again a higher voicing probability for vowels than rare syllables. In order to find a reason for the contradictory effects, additionally to positional respect accentuation was regarded, too. It could be made obvious from four syllable samples that accentuation has the highest effect on coarticulation from all three factors syllable probability, position and accentuation. Accentuation is such a high influencing factor that it is able to convert the outcomes of syllable probability effects to their contrary, i.e., rare syllables are less influenceable by coarticulatory effects than frequent syllables but unstressed rare syllables are more prone to coarticulation than stressed frequent syllables.

5.4 Formant transitions

Formants are an acoustic property of speech. They are the result of resonances that emerge in the vocal tract during speech production. The resonances cause energy peaks in certain frequencies of the speech, and these frequencies are referred to as formants. Formants have the characteristic to denote certain proportions in the vocal tract by their frequency itself and the course of their frequency which is referred to as *formant transition.* It is the formants, in particular the first three formants, that make vowels distinguishable because they are directly linked to vowel quality. Formant transitions also give clues to the identity of consonants by reflecting the rapid movement of articulatory position from consonant to vowel and vice versa.
In this chapter, formant transitions of 30 minimal pairs is surveyed. It is the same minimal pairs as in chapter 5.3, viz. those minimal pairs that accomplish the following threshold: one partner in each minimal pair has to occur at least ten times in the corpus, the other one at least five times (see also chapter 5.3.2). Comparisons of formant courses will be conducted between frequent and rare syllables with regard to steady states in the formant trajectories and transition courses during VC/CV-changes.

5.4.1 Method

For each occurrence of a minimal pair partner in the SmartKom corpus, `formant` (cf. chapter 4.2) was called up by a Perl script for the first three formants and with a frame rate of 10ms:

```
formant -i 0.01 -n 3 <file>
```

The generated data from `formant` was then transferred by use of `pplain` (cf. chapter 4.2) from binary code into text format. This resulting format had the form of six columns where the first three of them denoted the first three formants and the latter three denoted the bandwidth of each of the formants. The digits in each line corresponded to the formant measurement results, each line indicating a step of 10ms in time. A Perl script associated a minimal pair syllable's time with the respective lines and extracted the respective formant values for

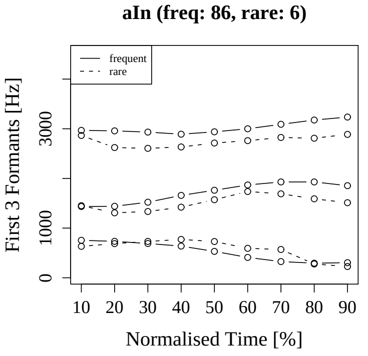


Figure 22: *Formant transitions in initial syllable /aIn/. The solid line represents the frequent and stressed syllable tokens.*

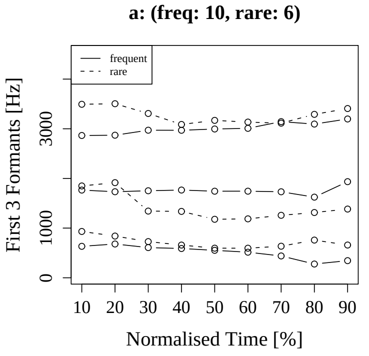


Figure 23: *Formant transitions in final syllable /a:/. The solid line represents the frequent and unstressed syllable tokens.*

the syllable. For the resulting files, time was normalised and the mean of the formant frequencies was calculated at nine equidistant steps in time over all occurrences of a single minimal pair partner and for all three formants. Then the graphics of formant transitions of frequent and rare syllables were created by use of R (2003). The graphics for all of the 30 minimal pairs can be found in the appendix (A.3).

5.4.2 Results

For nearly all considered minimal pairs noticeable differences in the course of formants could be observed between frequent and rare syllables. As expected the differences in syllable probability appeared in form of differences in frequency height inside vowels and in form of differences in transition courses at CV/VC transitions. Three syllables, viz. /En/, /In/, /ga:/, had only slight differences and can be viewed in the appendix (A.3). Nine syllables will be reviewed in detail in representation for all other syllables that can also be found in A.3.

The syllable /aIn/ in figure 22 exhibits fundamental differences in the formant course of frequent and rare syllables, in particular at the second and third formant. The frequent syllables were the ones that were stressed and all /aIn/-corpus occurrences that were calculated are initial syllables. As can be seen from in the figure, the second and third formant of rare syllables run nearly parallel to that of frequent syllables, but with a distance of about 200Hz. Additionally, the F1 of frequent syllables is lower than that of rare syllables, whereas the F2 is higher. Since F1 indicates the opening degree of the jaw it can be stated that the frequent /aIn/-syllables tend to be produced with a lesser jaw opening. The F2 indicates the retraction degree of the tongue, alternatively the rounding of the lips. Thus, the general higher F2 value of frequent syllables suggest a more central pronunciation of the syllable /aIn/ in contrast to rare syllables. In frequent syllable formant courses as well as in rare syllable courses, F2 starts out at about 1400Hz. For German, Neppert (1999: 147) reports a mean F2-value for /a/ of 1301Hz. During the phase of 20-30% of time where

the /a/ seems relatively stable, the F2-value of rare syllables is shortly above this benchmark whereas the F2-value of frequent syllables is about 100-200Hz higher than the benchmark which suggests an anticipatory effect towards the production of /I/. On the other side, during the production of /I/, frequent syllables are quite nearer in their F2-value to the reported mean F2-value for /I/ of 2088Hz (Neppert, 1999) than rare syllables, by having an F2-value of approximately 1900Hz during the stable phase of /I/ between 60% and 80% of time. But this may be due to the fact that /I/ in a diphthong is likely to never reach the F2-value which is reported for its context-free pronunciation. During the production of /aIn/ in normal speech, a good deal of jaw movement is involved, beginning with a quite big aperture that is reduced towards the pronunciation change to /I/ and /n/. The F3 is known to reflect the jaw movement by falling frequency during jaw aperture with less obstruction in the vocal tract at the same time. As can be seen from the figure, F3 of frequent syllables remains relatively stable in comparison to that of rare syllables. Like the F2-movement, this is also an indication to a more central pronunciation of frequent syllables which can be denoted as coarticulatory effect.
In the next figure the formant transitions for syllable /a:/ (figure 23) are displayed, where /a:/ was always in final position. The frequent syllables are the ones in unstressed position. Beside some differences in the onset that may be due to transitions from preceding sounds, some differences in the course of F1 and fundamental differences in the course of F2 are observable, in particular at the stable part of the syllable between 30% and 70% of time. When one regards the distances of the frequent syllables' three formants to each other, one can observe that they are quite equidistant to each other. Actually, this is a property of central vowels like /@/ and exhibits the fact that the frequent syllables

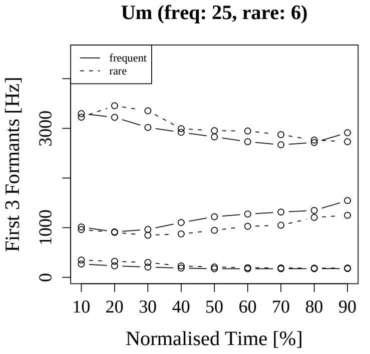


Figure 24: *Formant transitions in initial syllable /Um/. The solid line represents the frequent and stressed syllable tokens.*

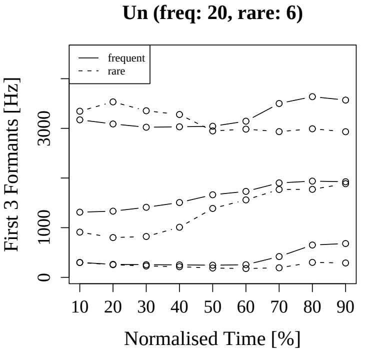


Figure 25: *Formant transitions in mid syllable /Un/. The solid line represents the frequent and unstressed syllable tokens.*

have a strong tendency to be pronounced in a central shape. German low and back vowels normally are pronounced like the rare syllables in the figure: with an F2-value near 1213Hz (Neppert, 1999: 147), with F1 and F2 being close to each other and a remarkably bigger distance between F2 and F3 than that of F1/F2. Regarding the figure, this can be only observed for rare syllables. Con-

clusively, it is likely that the observed properties of the formant courses here are a coarticulatory effect, namely a target undershoot that results in vowel reduction.
The formant transitions of the syllables /Um/ (figure 24) and /Un/ (figure 25) show some common properties, especially in the beginning of the course of F3. The /Um/-syllables are initial syllables where the frequent ones are the stressed; the /Un/-syllables are mid syllables where the frequent ones are the unstressed. The higher F3-values in the onset in both rare syllables' courses indicate that the syllables were pronounced with a more retracted tongue at a moderate jaw aperture and with presumably more lip rounding which enlarges the vocal tract and thus leads to a rise in F3 frequency. Regarding the F1/F2 courses, it is observable that in rare syllables F1 and F2 are closer to each other than in frequent syllables, viz. about 200-300Hz more than frequent syllables. The distance of F1 and F2 for /U/ is reported with approximately 380Hz (Neppert, 1999: 147) for male speakers. Here, in the stable time of /U/ between 10% and 40% of normalised time, the F1/F2 distance in /Um/ varies in rare syllables from approximately 580Hz to 640Hz and for /Un/ the distance varies from approximately 540Hz to 790Hz. However, for frequent syllables the distance is even larger: in /Um/ it varies from approximately 680Hz to 900Hz and in /Un/ from approximately 1010Hz to 1250Hz. This is a significant difference in F1/F2 distance and only coarticulatory effects can be the cause which seems to be again target undershoot in the pronunciation of /U/.

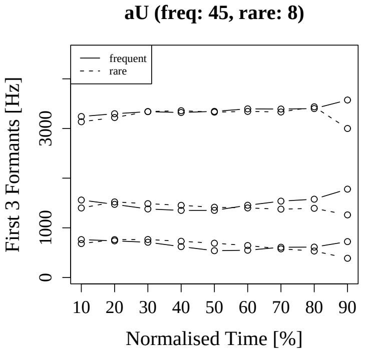


Figure 26: *Formant transitions in initial syllable /aU/. The solid line represents the frequent and stressed syllable tokens.*

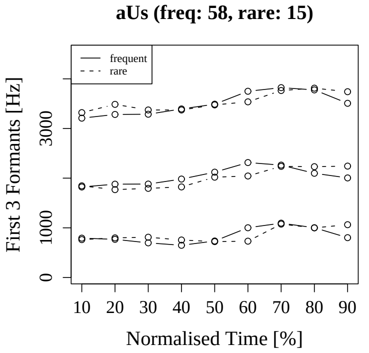


Figure 27: *Formant transitions in initial syllable /aUs/. The solid line represents the frequent and stressed syllable tokens.*

The syllables /aU/ (figure 26) and /aUs/ (figure 27) are reviewed for their interestingness regarding their similar behaviour at VC-boundaries. Both syllables are initial and in both syllable probability distributions the frequent syllables are the stressed. Regarding the VC-boundaries which lie around 90% in figure 26 and around 60% in time in figure 27, there is great similarity of the courses of the rare syllables' formants and those of frequent syllables: whereas the rare syllables' formants display a tendency to maintain their frequency level or to fall, frequent syllables' formants rather tend to rise. In numbers, the difference of the three frequent syllables' formants to their rare counterpart is as follows:

in /aU/ there is, at 90% of the time, a difference of 336 Hz in F1, of 518Hz in F2 and of 573Hz in F3. In /aUs/, at 60% of the time, there is a difference of 270Hz in F1, of 271Hz in F2 and of 213Hz in F3. (It should be noted that the direct indication of time indices is somewhat unreliable because of a certain inaccuracy in segmentation.) Considering these differences between rare and frequent syllables and the other slight differences in the course of the formants, it can be stated that coarticulatory effects are the cause, presumably induced by a mental syllabary effect.

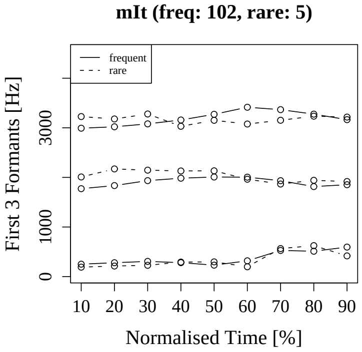


Figure 28: *Formant transitions in initial syllable /mIt/. The solid line represents the frequent and stressed syllable tokens.*

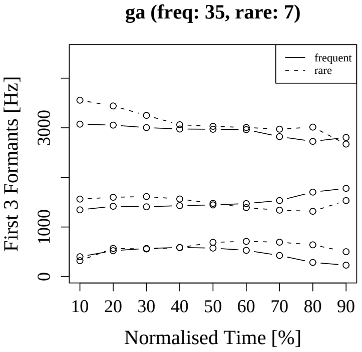


Figure 29: *Formant transitions in mid syllable /ga/. The solid line represents the frequent and unstressed syllable tokens.*

Now, the outcomes of syllables that start with a consonant will be reviewed. At first, the syllable /mIt/ (figure 28) and the syllable /ga/ (figure 29) will be regarded. The calculation of /mIt/ consists only of initial syllables where the frequent syllables are the stressed. The calculation of the /ga/ syllables consists only of mid syllables where the frequent syllables are the unstressed. Having a first look at the figures 28 and 29, some similarities between the two syllables can be noticed immediately: the frequencies of F2 and F3 at the onset are significantly higher for rare syllables than for frequent syllables. It can not be determined here why the onsets resemble each other at the beginning of the consonant. Although /mIt/ is syllable-initial it might be that the preceding word has an influence on /m/. For /ga/ it is quite likely that the beginning of /g/ is influenced by the preceding syllable for /ga/ is a mid syllable. Nonetheless, F2 maintains its higher frequency in rare syllables until 40% of /ga/ and even 50% of /mIt/. This is unlikely to be due to a anticipatory coarticulation, rather the F2 target frequency is reached faster in rare syllables which is confirmed by the sudden shift of F2 from 10% to 20% in /mIt/. In rare syllables the F2 remains from here on constant until the mentioned 50% of the syllable. In contrast, frequent syllables lack this sudden shift; their F2 curve exhibit a quite smoother transition from consonant to vowel to consonant, however, the level of the rare syllables' F2 frequency during the vowel duration is never achieved. First when the F2 curve of the rare syllables falls in preparation of the pronunciation of /t/, the F2 curve of the frequent syllables reaches its highest point. Regarding the overall frequency distances of /mIt/ it can be stated that the /I/ in rare syllables

is more fronted than that of frequent syllables which is more centred. Thus it seems that frequent syllables undergo a target undershoot effect in /mIt/.
Returning to /ga/ (figure 29) the steady phase of the vowel is reached at approximately 30% of the syllable. Both frequent and rare syllables exhibit strong transitions at the end of the syllable towards the sound patterns of the next syllable. But differences can be observed here: The frequent variants of /ga/ reach their most stable specification of /a/ between 40% and 60% of the syllable. Afterwards transitions of all three formants are drawn rapidly to the loci of the next sound. The rare variants of /ga/, in contrast, reach their most stable specification of /a/ not before 50% of the syllable, but then they maintain it until 80% and only then they exhibit rapid transitions to the loci of the next sound. This difference may be due to the fact, that fast produced syllables have an tendency to be more easily influenced by carry-over effects like anticipatory and perseverative coarticulation. As can be seen from figure 29 moreover, F1 and F2 come closer together in rare syllables than in frequent syllables and thereby a better prototype of /a/ is produced. That leaves the impression that here again coarticulation in form of vowel reduction takes place in frequent syllables but not in rare syllables.

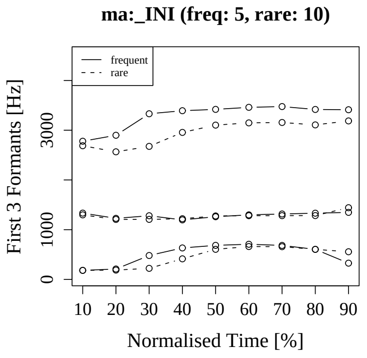


Figure 30: *Formant transitions in syllable /ma:/ in word initial position. The solid line represents the frequent and stressed syllable tokens.*

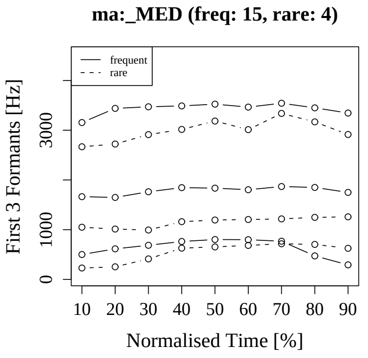


Figure 31: *Formant transitions in syllable /ma:/ in word mid position. The solid line represents the frequent and unstressed syllable tokens.*

In the appendix (A.3) the formant transition figure for /ma:/ can be viewed. However, that figure is composed of an calculation of initial as well as mid syllable occurrences, i.e. it comprises /ma:/ syllables regardless of their position. Therefore two figures will be reviewed here instead to allow for differences that may be caused by positional differences, viz. a calculation of /ma:/ (figure 30) in initial position and a calculation of /ma:/ (figure 31) in mid position. In initial /ma:/ syllables' probability distribution the frequent syllables are the stressed ones. For mid /ma:/ syllables, the frequent syllables are the unstressed ones. The comparison of the two figures displays fundamental differences in the course of F2: In initial position F2 of rare and frequent syllables are congruent, whereas in mid position F2 runs about 300-400Hz higher for frequent syllables than for rare syllables. It is also noticeable that the curve of rare syllables in mid position has about the same shape and height (at around 1200 Hz) as the curves in initial syllables. Therefore it is convenient to assume that it is the frequent syllables in syllable mid position that take the form of an outlier curve. The probable cause for this behaviour is likely to be coarticulation: target undershoot that results in vowel reduction. The reason why only the frequent syllables are prone to coarticulation may be their duration: while the rare syllables in mid position have a mean of 200ms in duration, frequent syllables are about one third shorter having a duration mean of only 137ms, and short syllables are much more influenceable by coarticulatory effects than longer syllables. Moreover, according to Modarresi et al. (2004), stressed vowels are less sensitive to context-dependent coarticulatory effects than unstressed syllables. Since here, in initial /ma:/ the frequent syllables are stressed but, conversely, in mid /ma:/ the frequent syllables are unstressed, the increased context-sensitivity of unstressed vowels may be present here, too. Thus, the great difference in F2 regarding the position of /ma:/ and the probability may be a cumulative effect of the differences in duration, the difference in position and the difference in accentuation. These differences resulting in vowel centralisation can be seen from F2 but not solely. Also deviations in the courses of F1 and F3 are present – in mid syllables as well as in initial syllables. Both F3 courses of the frequent variants suggest that /a:/ was pronounced with a more retracted tongue which led to more constriction in the vocal tract than was present during the pronunciation of rare syllables, therefore F3 in frequent syllables is significantly higher than in rare ones. The coarticulatory effect for the consonant is bigger in mid syllables than in initial syllables. This can be seen particularly from F1. Concluding from the course of the formants, in initial syllables the /m/ seems to have a longer duration: for frequent syllables the pronunciation of /m/ seems to be maintained until 20% of the syllable, maximally until 30%. At 40% of the syllable, F1 has already reached the target position of F1 for the pronunciation of /a:/ for frequent syllables. Whereas for rare syllables, the /m/ is pronounced until 30% of the syllable, perhaps until 40%, and it is not until 50% that F1 reaches its target position for /a:/. Thus, F1 CV-transition in figure 30 in rare syllables starts later but is processed faster which can be seen from the steepness of the curve. The transition course of F1 of frequent syllables is much more smoother and starts earlier. For mid syllables (figure 31) coarticulatory effects in CV-transition are even stronger. For rare syllables, F1 transition starts already at 20% of the syllable (30% in initial position) reaching its target position at 40%. For frequent syllables, no stable course of F1 is present in /m/, instead the syllable starts directly with F1 transition towards the vowel target position, reaching it at already 40% of the syllable. Additionally, like in initial syllables, the transition course of frequent syllables is much smoother than that of rare syllables. Another coarticulatory effect should be mentioned here, too: perseverative coarticulation. At the end of initial syllables as well as mid syllables, transitions can be observed again. These are transitions that anticipate future frequency positions of the formants. While for rare syllables the F1 transition is only slightly present, since F1 runs nearly horizontally, for frequent syllables the F1 transition is much more obvious, especially for frequent syllables in mid position. F1 of frequent syllables in figure 31 starts to fall already after 70% whereas F1 for rare syllable remains stable until 90%. Although smaller, this effect can be viewed also in figure 30. Consequently at least four different kinds of coarticulation, partly interacting with each other, have been shown from this syllable: target undershoot, vowel reduction, perseverative and anticipatory coarticulation.

5.4.3 Discussion of formant transitions

Reviewing the formant transition investigation, it can be formulated that the examined syllables exhibited strong differences in formant courses, formant heights and formant transitions. These differences can mainly be ascribed to coarticulatory effects. These coarticulatory effects in turn are likely to have their origin in a mental syllabary effect. No other effect is known by the author that could account for such differences in principally identical syllables whose only difference is their probability distribution with regard to stress. The mental syllabary is thought of to retrieve frequent syllables from some sort of a mental repository instead of assembling them newly every time. These frequent syllables are supposedly highly overlearned and therefore are much more prone to coarticulation than rare syllables. Rare syllables on the other side are thought of to be assembled during speech production, i.e. they are not retrieved from the mental syllabary. Therefore they are supposed to be produced with more effort and accuracy, thus being less influenceable by coarticulatory effects. This structural condition may account for the effects and differences that were found during this investigation.

6 Summarising discussion

The present work showed by means of syllable duration measurement that two factors have high significance with regard to syllable duration: stress and position of the syllable. Syllable probability, however, seems to play only a minor role thus being no strong indicator for the existence of a mental syllabary. The investigation by means of a comparison of fundamental frequency between rare and frequent syllables exhibited larger support for the existence of a mental syllabary. It could be found that frequent syllables are more prone to coarticulation than rare syllables but, additionally, that unstressed rare syllables are more prone to coarticulatory effects than stressed frequent syllables. This is a main finding and should be investigated further. At last, formant transitions have been investigated and here, too, large differences could be found between rare and frequent syllables, in formant courses and formant heights as well as in transitions. Exhibited coarticulatory effects throughout the investigations were target undershoot, vowel reduction, right-to-left and left-to-right coarticulation. More evidence should be provided for the findings in this work by evidence from other syllables as well as evidence from corpora other than the SmartKom corpus.

Moreover, it would be convenient to regard fundamental frequency and formant transitions with respect to the segments of the syllable, i.e., it would be easier and partly much more reliable to interpret the outcomes if the segment boundaries were drawn. Unfortunately, at the moment, the simultaneous access on syllable and segment information that would be necessary to attain segment boundaries within a syllable remains a problem, for Festival does not provide features to allow such simultaneous access.

For further investigation it should be also taken into account that the threshold for the definition of minimal pairs was attained by a statistical approach in the present work, therefore other approaches would be interesting, too.

It could be seen from this work but has remained only a side remark that the high frequent syllables of the minimal pairs were nearly exclusively function words. Jurafsky[12] investigated English functions words with regard to their probability and their predictability in context. He also found some interesting coarticulatory phenomena on English functions words, e.g., a higher tendency to have reduced vowels. Concluding from the present work, the findings of Jurafsky et al. may be also due to a mental syllabary and it would be interesting to investigate German function words in a similar way like Jurafsky and with regard to the possibility that function words due to their high occurrence in speech are likely to be present in a mental syllabary above all other syllables.

---

[12] Bell et al. (1999), Jurafsky et al. (2001.) and Jurafsky et al. (2001)

References

Baumann, S., Grice, M., Benzmüller, R., 2000. GToBI – a phonological system for the transcription of German intonation. In: S. Puppel, G. Demenko (eds.), *Prosody 2000 – Speech Recognition and Synthesis*. Poznań, 21–28.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gildea, D., 1999. Forms of english function words - effects of disfluencies, turn position, age and sex, and predictability. In: Proceedings of ICPhS-99.

Black, A. W., Taylor, P., Caley, R., 1999. *The Festival speech synthesis system—system documentation*. CSTR Edinburgh.
URL `http://www.cstr.ed.ac.uk/projects/festival`

Cholin, J., Schiller, N. O., Levelt, W. J. M., January 2004. The preparation of syllables in speech production. *Journal of Memory and Language* **50** (1), 47–61.

Clark, J., Yallop, C., 1995. *An introduction to phonetics and phonology*, 2nd Edition. Blackwell.

Cranfield, T., 7 2002. *Celex Reference Manual*.
URL `http://europa.eu.int/celex/htm/doc/de/referencemanual_de.pdf`

Crompton, A., 1982. *Slips of the tongue*. Mouton, Ch. Syllables and segments in speech production, 109–162.

Croot, K., Rastle, K., 8 - 10 December 2004. Is there a syllabary containing stored articulatory plans for speech production in english? In: Proceedings of the 10th Australian International Conference on Speech Science and Technology. Australian Speech Science & Technology Association Inc., Macquarie University, Sydney, 376–381.

Faraway, J. J., 2004. *Linear Models with R*. CRC Press.

Farnetani, E., Recasens, D., 1999. *Coarticulation: Theoy, data and techniques*. Cambridge, Ch. 2: Coarticulation models in recent speech production theories, 31–65.

Fujimura, O., Lovins, J. B., 1978. *Syllables and segments*. Amsterdam: North-Holland, Ch. Syllables as concatenative phonetic units, 107–130.

Hakkarainen, H. J., 1995. *Phonetik des Deutschen*. Wilhelm Fink Verlag, München.

Hedelin, P., Huber, D., 1990. Pitch period determination of aperiodic speech signals. In: Proceedings of the IEEE International Conference on Acoustics and Speech Signal Processing. 361–364.

Hillenbrand, J. M., Clark, M. J., 2001. Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America* **109** (2), 748–763.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W. D., 2001. The effect of language model probability on pronunciation reduction. In: Proceedings of ICASSP-01. Salt Lake City, Utah, 801–804.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W. D., 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, Ch. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production, 229–254.

Klankert, T., April 2003. *Hybride Unit Selection für ein Sprachsynthesesystem*. Master's thesis, Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.

Kohler, K. J., 1977. *Einführung in die Phonetik des Deutschen*. Erich Schmidt Verlag.

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

Levelt, W. J. M., 1992. Accessing words in speech production: Stage processes and representations. *Cognition* **42**, 1–22.

Levelt, W. J. M., 1993. *Theorie und Praxis des Lexikons*. de Gruyter, Berlin, Ch. Lexical Selection, or how to Bridge the Major Rift in Language Processing, 164–172.

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

Levelt, W. J. M., Wheeldon, L., 1994. Do speakers have access to a mental syllabary? *Cognition* **50**, 239–269.

Lindblom, B., 1983. *The production of speech*. New York: Springer, Ch. Economy of speech gestures, 217–245.

Möbius, B., 2003. Rare events and closed domains: Two delicate concepts in speech synthesis. *International Journal of Speech Technology* **6**, 57–71.

Möbius, B., 2004. Corpus-Based Investigations on the Phonetics of Consonant Voicing. *Folia Linguistica* **38**, 5–26.

Modarresi, G., Sussman, H., Lindblom, B., Burlingame, E., 2004. An acoustic analysis of the bidirectionality of coarticulation in VCV utterances. *Journal of Phonetics* **32**, 291–312.

Müller, K., 2002. *Probabilistic Syllable Modeling Using Unsupervised and Supervised Learning Methods*. PhD thesis, University of Stuttgart, Institute of Natural Language Processing (IMS), Stuttgart.

Neppert, J. M. H., 1999. *Elemente einer akustischen Phonetik*. Buske Verlag.

Pierrehumbert, J., Beckman, M. E., Ladd, D. R., 2001. Conceptual foundations of phonology as a laboratory science. In: N. Burton-Roberts, P. Carr, G. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford University Press, Oxford, UK.

Pompino-Marschall, B., 2003. *Einführung in die Phonetik*. deGruyter.

R, 2003. *The R project for statistical computing*. http://cran.r-project.org.

Schweitzer, A., Braunschweiler, N., Dogil, G., Klankert, T., Möbius, B., Möhler, G., Morais, E., Säuberlich, B., Thomae, M., 2006. *SmartKom: Foundations of Multimodal Dialogue Systems*, 1st Edition. Springer, Ch. Multimodal Speech Synthesis, 411–435.

Schweitzer, A., Braunschweiler, N., Klankert, T., Möbius, B., Säuberlich, B., 2003. Restricted unlimited domain synthesis. In: Proceedings of Eurospeech-2003 (Geneva). 1321–1324.

Schweitzer, A., Möbius, B., 2004. Exemplar-based production of prosody: Evidence from segment and syllable durations. In: Speech Prosody 2004. Nara, Japan, 459–462.

Shih, C., Möbius, B., 1998. Contextual effects on voicing profiles of German and Mandarin consonants. In: Proceedings of the International Conference on Spoken Language Processing (Sydney, Australia). Vol. 7. 3135–3138.

Shih, C., Möbius, B., Narasimhan, B., 1999. Contextual effects on consonant voicing profiles: A cross-linguistic study. In: Proceedings of the 14th International Congress of Phonetic Sciences (San Francisco, CA). Vol. 2. 989–992.

Taylor, P., Caley, R., Black, A., King, S., 1999. *The Edinburgh speech tools library: System Documentation*. The Centre for Speech Technology Research, University of Edinburgh.
URL `http://www.cstr.ed.ac.uk/projects/speech_tools/`

Whiteside, S. P., Varley, R. A., 1998. Dual-Route Phonetic Encoding: Some Acoustic Evidence. In: Proceedings of the 5th International Conference on Spoken Language Processing. Vol. 7. 3155–3158.

Wiese, R., 1996. *The Phonology of German*. Clarendon Press, Oxford.

A Appendix

A.1 The 117 minimal pairs

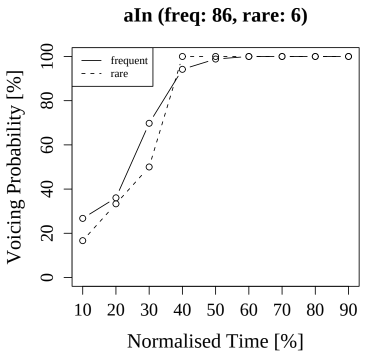
117 minimal were investigated during the present work. Their acquisition is described in chapter 5.1.

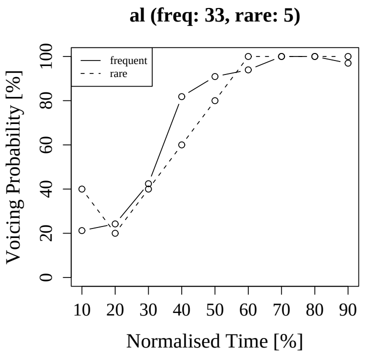
probability of rare syllables	syllable	position	stress of the rare syllables	probability of frequent syllables
0.00003319994217	a:	FIN	STR	0.00015283477786
0.00000072036095	aI	INI	USTR	0.01051601574842
0.00000010564658	aIn	INI	USTR	0.00463675880839
0.00004516732143	al	INI	USTR	0.00042129243578
0.00000536628619	an	ONE	USTR	0.00413954592885
0.00003256102357	ap	INI	USTR	0.00076147102104
0.00001353217894	at	INI	USTR	0.00031467457152
0.00000737324231	aUf	MED	USTR	0.00025011182370
0.00003327443338	aU	INI	USTR	0.00112872042858
0.00005411238639	aUs	INI	USTR	0.00183557613942
0.00000616600026	aUs	MED	USTR	0.00020916029968
0.00006728772923	baI	FIN	USTR	0.00022633920458
0.00000376030469	baI	INI	USTR	0.00077711016596
0.00004299399235	ba:	INI	USTR	0.00016603535533
0.00000302125319	baI	ONE	USTR	0.00134610564672
0.00001256343106	bE6	INI	STR	0.00033370048563
0.00008080141092	bi:	INI	USTR	0.00018045846996
0.00011399523590	bi:	MED	STR	0.00026707814716
0.00001888737913	bIs	INI	USTR	0.00022328488995
0.00004413194204	bli:	MED	STR	0.00020045702477
0.00007029952726	bo:	MED	USTR	0.00013975108023
0.00005373495784	by:	INI	USTR	0.00014268771528
0.00000020644797	dEn	ONE	USTR	0.00048745048738
0.00008933839654	dI	INI	STR	0.00022790091337
0.00000033622480	dRaI	INI	USTR	0.00014774253027
0.00000568602322	e:6	INI	USTR	0.00034968560230
0.00006398827419	E6	INI	STR	0.00555844873471
0.00004792874277	E6	MED	STR	0.00059944664868
0.00000007208097	e:6	ONE	USTR	0.00188843977825
0.00008843255708	En	MED	USTR	0.00013419894254
0.00008428035955	fa:	INI	USTR	0.00032547615799
0.00010358498382	fE6	INI	STR	0.00903485908439
0.00010163294487	fE	MED	USTR	0.00015456303414
0.00007173429539	fo:6	INI	USTR	0.00055953430159
0.00008212949652	fO6	INI	USTR	0.00058649146680
0.00002836110165	fO6	MED	USTR	0.00021053582968
0.00000004245145	fo:6	ONE	USTR	0.00091055667708
0.00008122513200	fRy:	INI	USTR	0.00021568507678
0.00000039396677	fy:6	FIN	USTR	0.00015315897466
0.00011991339623	fy:	INI	USTR	0.00031841782755
0.00011403658038	ga	MED	STR	0.00015081614924
0.00004769622234	ga:	MED	STR	0.00016033919806
0.00011333565308	gi:	INI	USTR	0.00028975092923
0.00007435517006	gI	MED	STR	0.00018755877976
0.00000060829602	haI	INI	USTR	0.00017821419019
0.00006408279128	hal	MED	STR	0.00014245299731
0.00012315411342	hE	INI	USTR	0.00018729235349
0.00008255889364	ho:	MED	USTR	0.00016412193677
0.00008789334094	i:	FIN	USTR	0.00019629728794
0.00012319877007	In	INI	USTR	0.00133554529154
0.00010286627630	ja	INI	STR	0.00013604315059
0.00000047654820	kaI	INI	USTR	0.00073228431942
0.00008664737960	ka:	INI	USTR	0.00033461717963
0.00008857228394	kOn	MED	USTR	0.00015657162818
0.00009654852112	kRi:	INI	USTR	0.00021562740305
0.00010352093052	laI	MED	STR	0.00014766730079
0.00008633438521	la:	INI	USTR	0.00033340844949
0.00011064983236	la:	MED	USTR	0.00013639500804
0.00006524487208	lan	INI	USTR	0.00026324238547
0.00000080993515	lIS	FIN	STR	0.00027866529977
0.00010458220717	lIs	MED	STR	0.00025497978186
0.00000104710052	maI	INI	USTR	0.00028280904979
0.00003704980761	ma:	INI	USTR	0.00014307994294

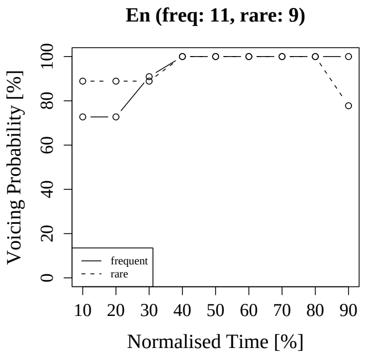
probability rare syllables	**syllable**	**position**	**stress of the rare syllables**	**probability of frequent syllables**
0.00013354339358	ma:	MED	STR	0.00039279244367
0.00004979025859	man	INI	USTR	0.00068990865841
0.00000276638021	man	ONE	USTR	0.00080072342286
0.00005565020630	mEn	INI	USTR	0.00016879542635
0.00005492091506	mIs	INI	USTR	0.00026970829388
0.00008834883131	mIs	MED	STR	0.00013860280127
0.00005150772497	mIt	INI	USTR	0.00022097646818
0.00007785436464	mIt	ONE	USTR	0.00029948733968
0.00010328652120	mo:	INI	USTR	0.00020532716893
0.00013189586770	mo:	MED	USTR	0.00026220076728
0.00002869839429	mY	INI	USTR	0.00019589207132
0.00000001004477	na:x	INI	USTR	0.00033820793489
0.00005831129721	nIs	MED	STR	0.00014873279413
0.00004867334507	nOY	INI	USTR	0.00025154589840
0.00008175942681	O6	INI	USTR	0.00015061872656
0.00001487907252	pE6	INI	STR	0.00037033833345
0.00005775091089	pi:	MED	STR	0.00027177132381
0.00011046005359	Ra	MED	STR	0.00014608610563
0.00009935240337	Ra:	MED	USTR	0.00020128990429
0.00011696100327	RE	INI	USTR	0.00017787389279
0.00012873263852	RI	INI	STR	0.00032632481738
0.00010661932341	Ro:	INI	USTR	0.00021195257208
0.00000246708374	RUm	FIN	STR	0.00014638792764
0.00004244898301	si:	MED	STR	0.00032893969401
0.00009475981786	Si:	MED	STR	0.00023684546007
0.00003361425696	Sli:	MED	STR	0.00013430223887
0.00006355644509	Spi:	INI	USTR	0.00014194428917
0.00001819202225	Stan	INI	USTR	0.00035835104268
0.00013054821133	Ste:	INI	USTR	0.00018983593915
0.00012439481923	StEn	MED	USTR	0.00018295793426
0.00008105541704	StI	MED	STR	0.00020577685245
0.00009078800722	ta:	FIN	STR	0.00041793943026
0.00000215796232	taI	INI	USTR	0.00017079755907
0.00012945860632	ta	INI	STR	0.00017121215338
0.00011225122327	ta:	INI	USTR	0.00043349479134
0.00010154850640	ta	MED	STR	0.00013430036788
0.00011537104425	tE	MED	USTR	0.00017545588857
0.00009698062658	tRi:	INI	USTR	0.00021659244918
0.00006452554877	tsi:	INI	USTR	0.00014410864454
0.00009700945343	tsI	MED	STR	0.00024723331580
0.00006289751096	tsu:	FIN	USTR	0.00021107379697
0.00000042592346	tsvaI	INI	USTR	0.00038888612808
0.00002880119580	Um	INI	USTR	0.00056307225755
0.00010196215512	Un	MED	STR	0.00019432146192
0.00000194392307	vaI	INI	USTR	0.00073907808048
0.00012442350617	vE	INI	USTR	0.00018922284162
0.00007305847505	vI	INI	STR	0.00014077541390
0.00000300580672	vIl	ONE	USTR	0.00022282914291
0.00012076102642	vo:	INI	USTR	0.00024006539656
0.00000309267761	vU6	INI	USTR	0.00088988193779
0.00011465176053	zi:	INI	USTR	0.00049758968653
0.00012048803194	zI	INI	STR	0.00024285927078
0.00010898259101	zO6	INI	USTR	0.00020592929107
0.00012316616813	zOn	INI	USTR	0.00022786681910

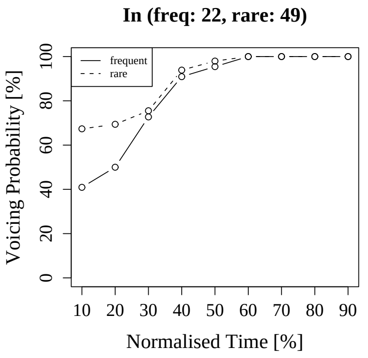
A.2 Voicing profiles of 30 minimal pairs

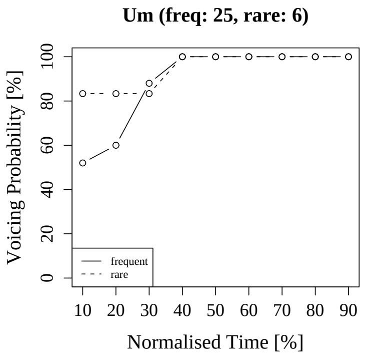
Here all single voicing profiles of the 30 statistically most reliable minimal pairs (see 5.3.2) are given. Voicing profiles are presented and discussed in chapter 5.3.

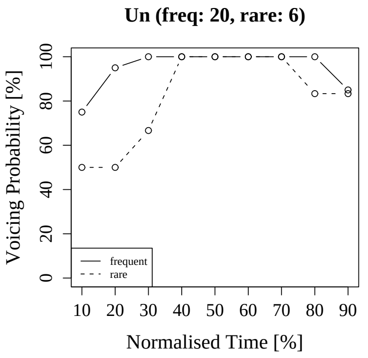

aIn (freq: 86, rare: 6)

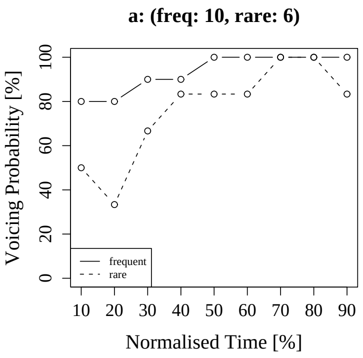

al (freq: 33, rare: 5)

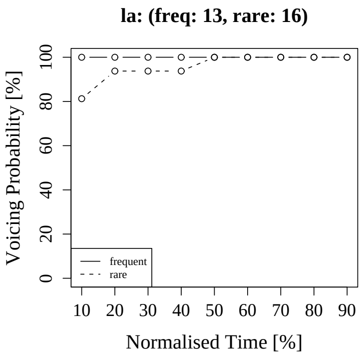

En (freq: 11, rare: 9)

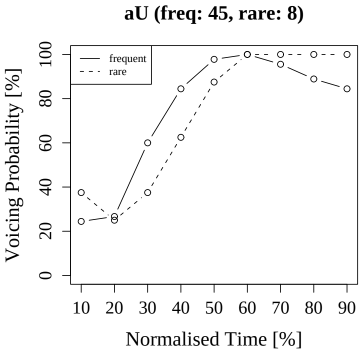

In (freq: 22, rare: 49)

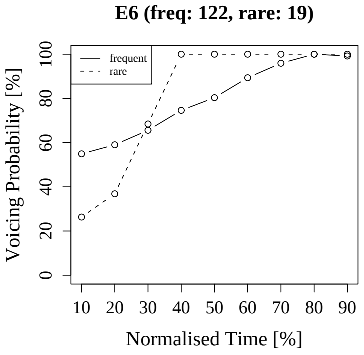

Um (freq: 25, rare: 6)

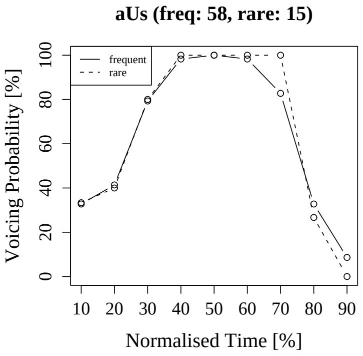

Un (freq: 20, rare: 6)

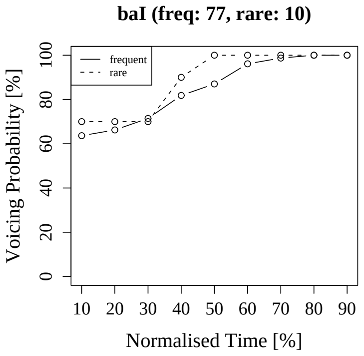

a: (freq: 10, rare: 6)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

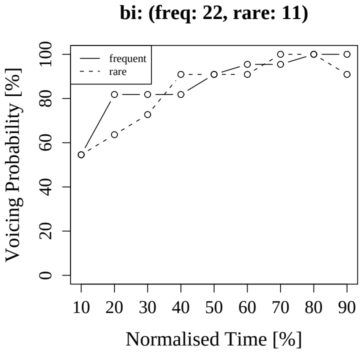

la: (freq: 13, rare: 16)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

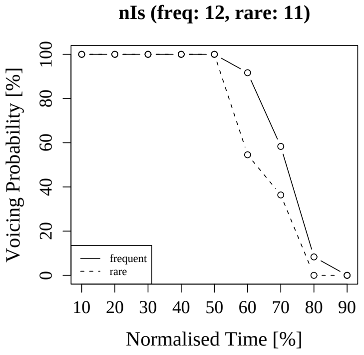

aU (freq: 45, rare: 8)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

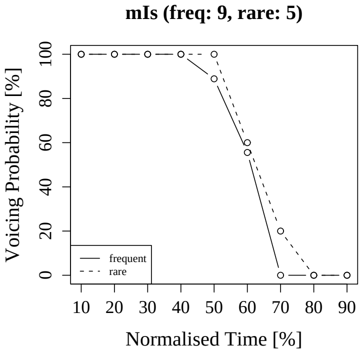

E6 (freq: 122, rare: 19)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

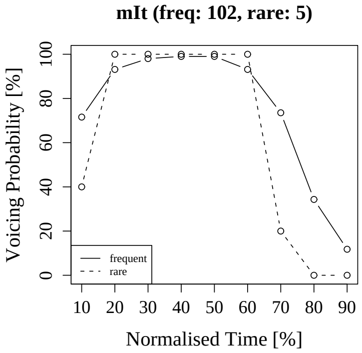

aUs (freq: 58, rare: 15)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

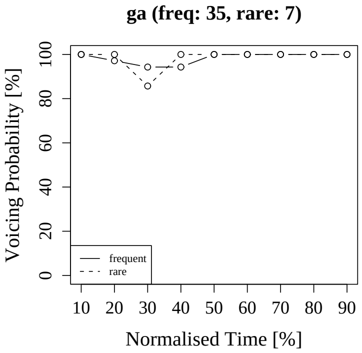

baI (freq: 77, rare: 10)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

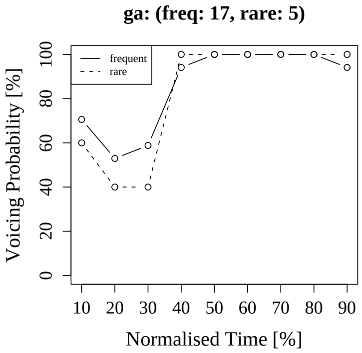

bi: (freq: 22, rare: 11)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

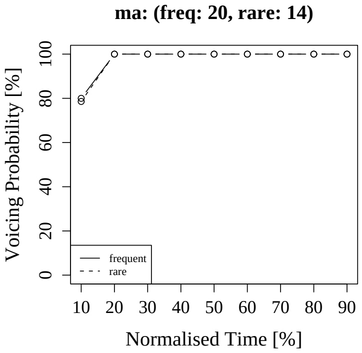

nIs (freq: 12, rare: 11)
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]
frequent
rare

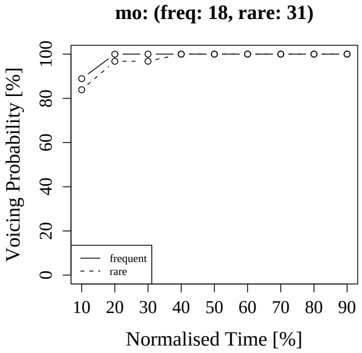

mIs (freq: 9, rare: 5)

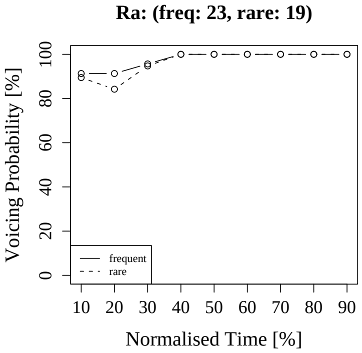

mIt (freq: 102, rare: 5)

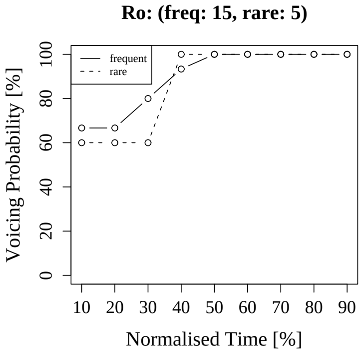

ga (freq: 35, rare: 7)

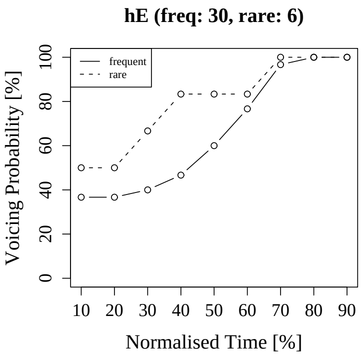

ga: (freq: 17, rare: 5)

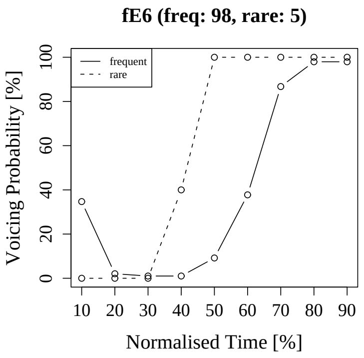

ma: (freq: 20, rare: 14)

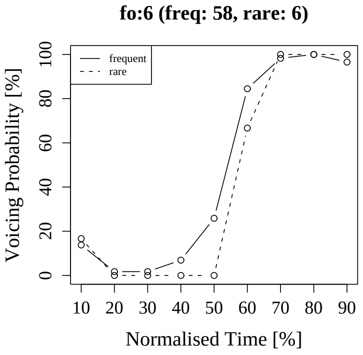

mo: (freq: 18, rare: 31)

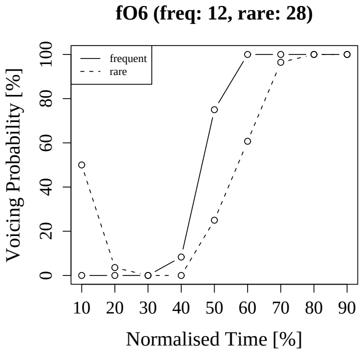

Ra: (freq: 23, rare: 19)

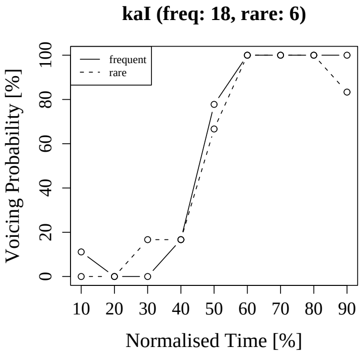

Ro: (freq: 15, rare: 5)

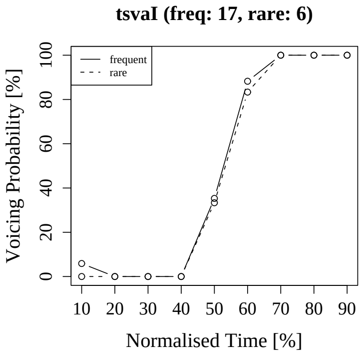

hE (freq: 30, rare: 6)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]

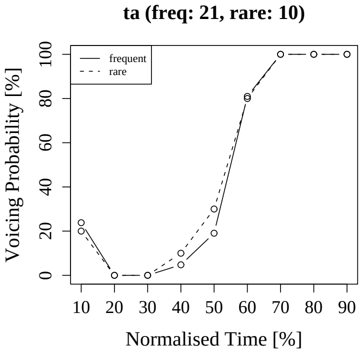

fE6 (freq: 98, rare: 5)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]

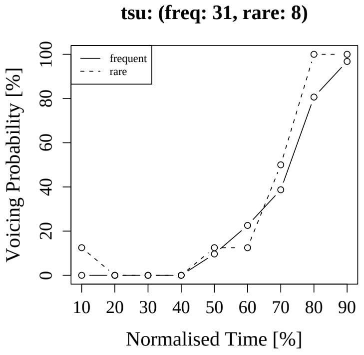

fo:6 (freq: 58, rare: 6)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]


fO6 (freq: 12, rare: 28)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]


kaI (freq: 18, rare: 6)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]

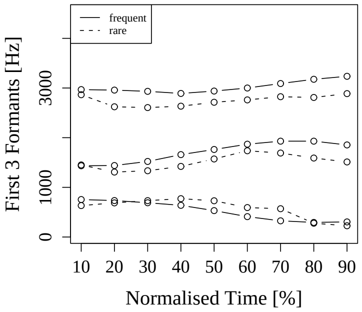

tsvaI (freq: 17, rare: 6)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]

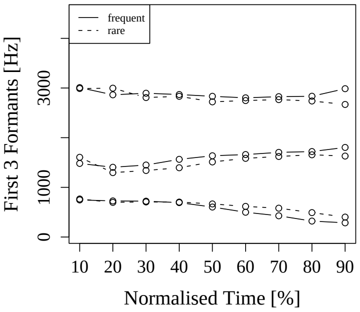

ta (freq: 21, rare: 10)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]

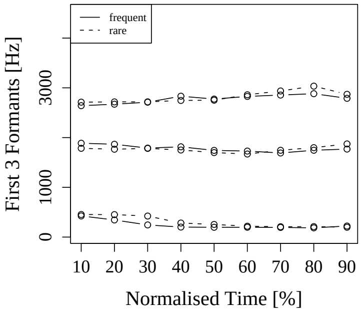

tsu: (freq: 31, rare: 8)
frequent
rare
Voicing Probability [%]
0
20
40
60
80
100
10
20
30
40
50
60
70
80
90
Normalised Time [%]

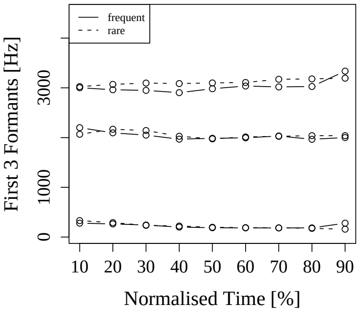
A.3 Formant transitions calculations of 30 minimal pairs

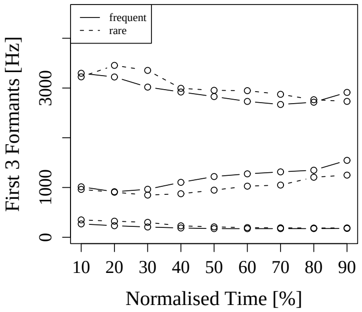
Here formant transition graphics of the 30 statistically most reliable minimal pairs (see 5.3.2) are given. Formant transitions are discussed in chapter 5.4.

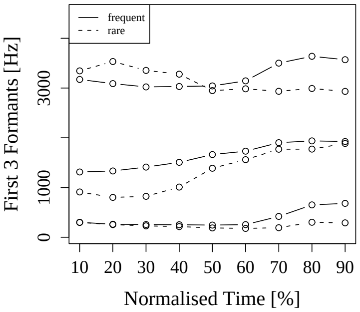

aIn (freq: 86, rare: 6)

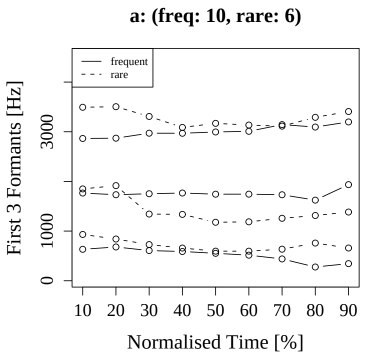

al (freq: 33, rare: 5)

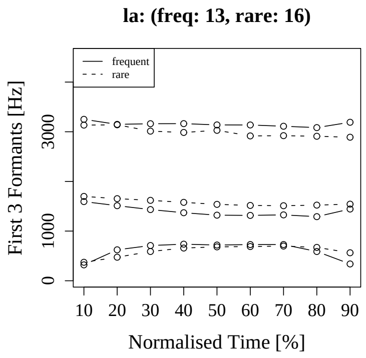

En (freq: 11, rare: 9)

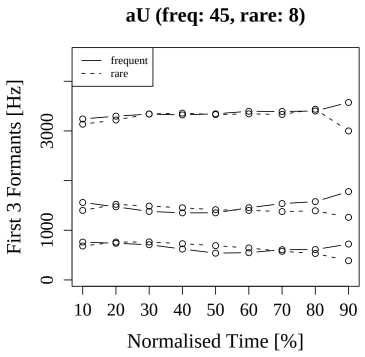

In (freq: 22, rare: 49)

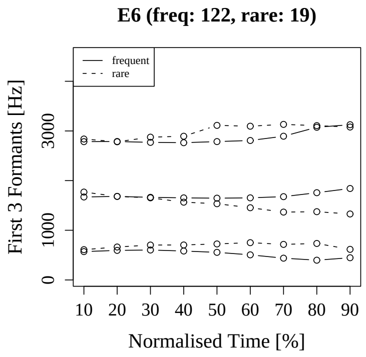

Um (freq: 25, rare: 6)

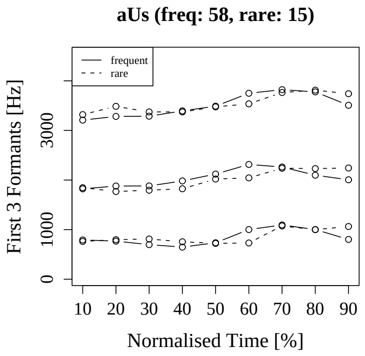

Un (freq: 20, rare: 6)

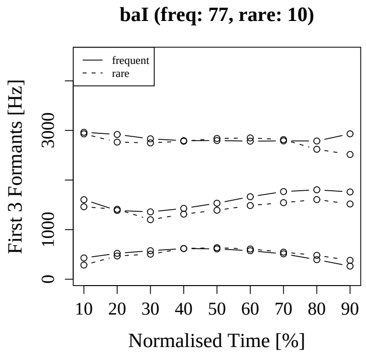

a: (freq: 10, rare: 6)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

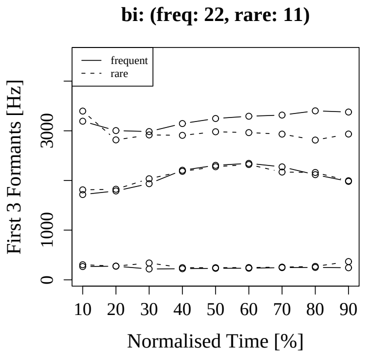

la: (freq: 13, rare: 16)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

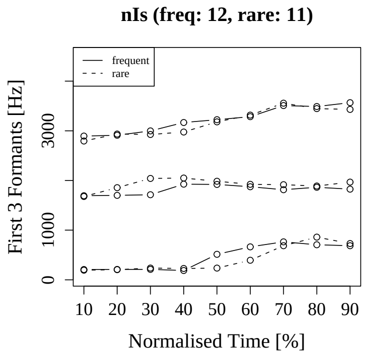

aU (freq: 45, rare: 8)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

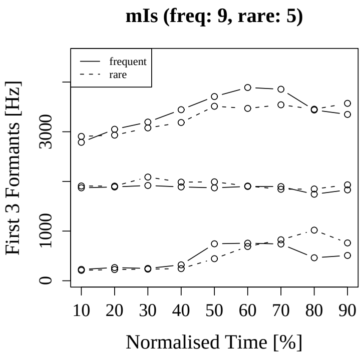

E6 (freq: 122, rare: 19)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

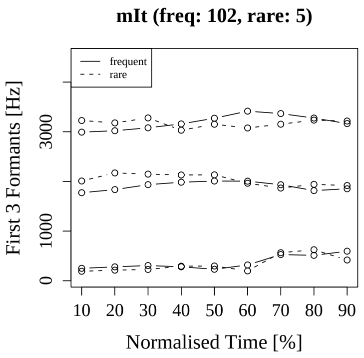

aUs (freq: 58, rare: 15)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

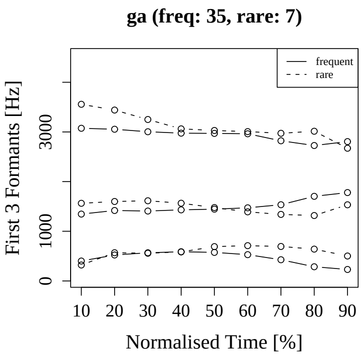

baI (freq: 77, rare: 10)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

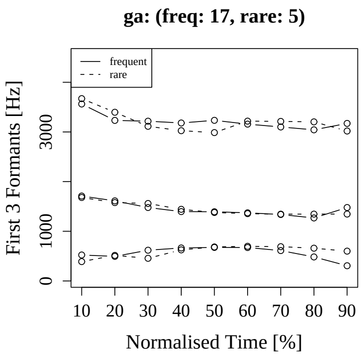

bi: (freq: 22, rare: 11)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

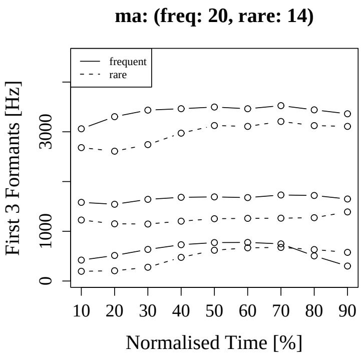

nIs (freq: 12, rare: 11)
frequent
rare
First 3 Formants [Hz]
0
1000
3000
10
20
30
40
50
60
70
80
90
Normalised Time [%]

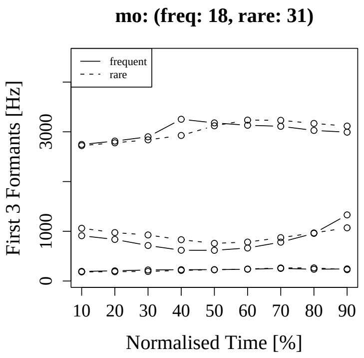

mIs (freq: 9, rare: 5)

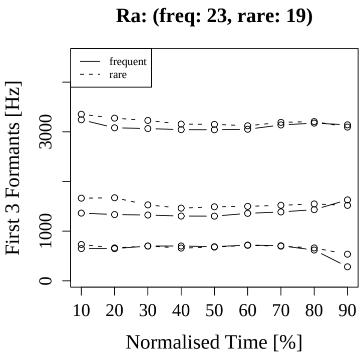

mIt (freq: 102, rare: 5)

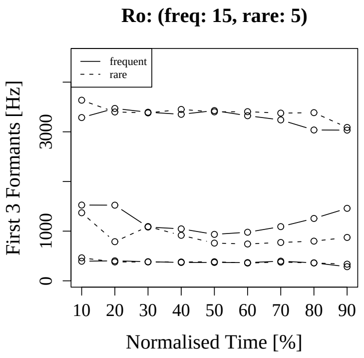

ga (freq: 35, rare: 7)

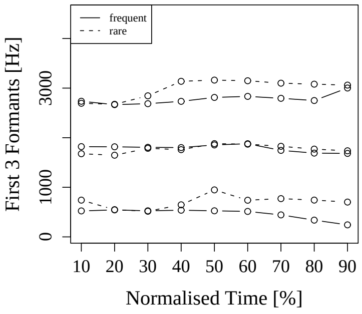

ga: (freq: 17, rare: 5)


ma: (freq: 20, rare: 14)

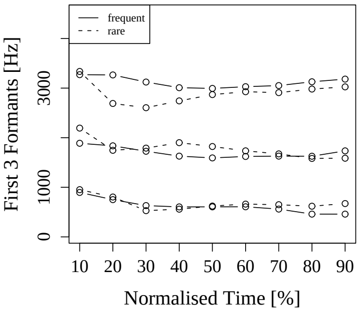

mo: (freq: 18, rare: 31)


Ra: (freq: 23, rare: 19)

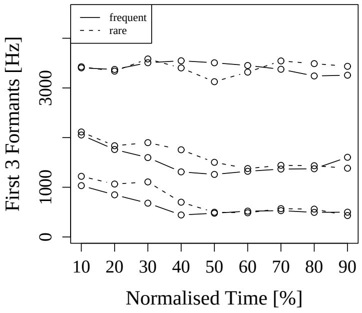

Ro: (freq: 15, rare: 5)



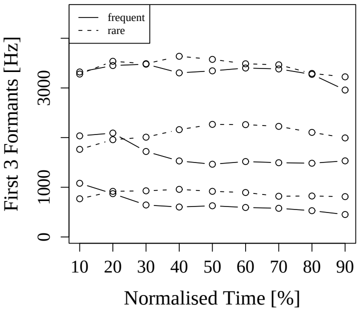
hE (freq: 30, rare: 6)



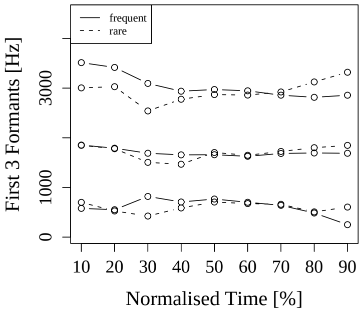
fE6 (freq: 98, rare: 5)



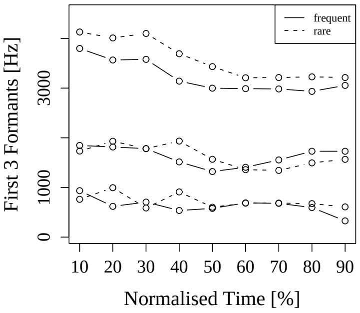
fo:6 (freq: 58, rare: 6)



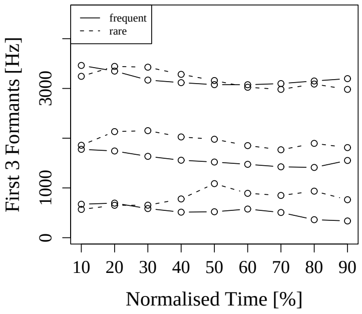
fO6 (freq: 12, rare: 28)



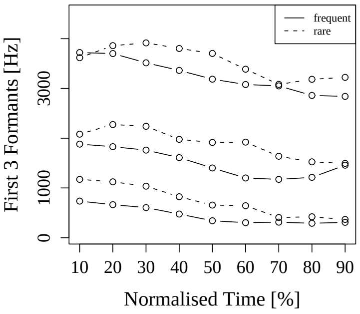
kaI (freq: 18, rare: 6)



tsvaI (freq: 17, rare: 6)



ta (freq: 21, rare: 10)



tsu: (freq: 31, rare: 8)