PHONETIC ANALYSIS OF VOWEL SEGMENTS IN THE PHONDAT DATABASE OF SPOKEN GERMAN

S. Heid, M.-B. Wesenick, Chr. Draxler
IPSK – Institut für Phonetik und Sprachliche Kommunikation, Munich, Germany
e-mail: {heid, draxler, wesenick}@phonetik.uni-muenchen.de
http://www.phonetik.uni-muenchen.de

ABSTRACT

This paper describes some characteristics of the PhonDat database of spoken German [9] and its use for empirical studies in phonetic research. As an example results of vowel duration and formant measurement are presented.

INTRODUCTION

The aim of this paper is to show how symbolic data related to speech signals can be made available in a well-structured way and how this information can be used to determine and extract the relevant signal fragments for an acoustic analysis. Presently methods for the investigation of very large speech corpora are being developed. As an example, we retrieve from the symbolic database information on vowel position and compute the duration of vowel classes in selected consonant contexts. We then use the position information to apply a semi-automatic formant extraction program to the signal fragments. Duration measurements of 9950 vowels with specified contexts and formants of 10629 vowels are presented.

The purpose of our investigations currently is to establish a basis for the empirical study of German phonetics and phonology.

SPEECH DATA

The PhonDat Database (PhDB) consists of two main corpora of which only one - the PhonDat II train enquiry corpus - was investigated for the present studies. It contains data of 16 speakers with 64 read sentences each from the domain of train enquiries. The speech signals of the utterances have been segmented manually using a broad phonemic transcription (SAMPA) relative to the given citation form. The PhDB strictly adheres to the Computer Representation of Individual Languages (CRIL) guidelines agreed upon at the IPA Kiel 89 convention. Data is represented on three different symbolic levels: orthography, citation form, and phonetic transcription with time marks. The PhDB is implemented in Prolog [2], using the persistent Prolog environment Eclipse [3]. Access to the data is possible via the symbolic data levels; the result of a query is a reference to a signal fragment, or again symbolic data. Most database queries can be formulated using the query toolbox with little Prolog knowledge (except for Prolog syntax: Variables begin with capital letters, constants with lower case letters; the „," is the logical AND, and „?-" initiates a query).

Example:

"Find the segmentation of the word "und" and display its labels in SAMPA".

```
?- word_canword(Id,und,_,_),
word_in_sentence(Id,_,SNb,WordPos),
segment_file(File,Spk,_,
  Sgr,SNb,_,Segs),
word_segments(WPos,Segs,WordSegs),
labels(WordSegs,WordLabels),
sampa_ipa(SampaLabels,WordLabels).
```

Complex applications combine the toolbox predicates with the standard control constructs of Prolog. The vowel duration analyses presented below are an example of a complex application; vowels are searched using a multi-level search pattern, e.g. "*vowel:, voiced-plosive". The code of the program proper is less than 20 lines, I/O and initialization require 15 lines each.

The PhDB contents are shown in table1

	number
word types	195
word tokens	676
segment files	5286
phonetic segments	238.769
reference segmentations	991
reference phonetic segments	39683

Table 1: PhonDat DB contents

The reference segmentations are the manual segmentations that have been selected for distribution within the PhonDat project. Phonetic segments do not contain para-phonetic (e.g. prosodic or syntactic) labels.

ANALYSES

Vowel Durations

Specific classes of speech sounds and the corresponding durations can be selected from the entire stored information by database queries only. For our investigations we chose the classic question about vowel duration and the influence of the following consonant and of vowel stress and length.

Stress and phonological length are factors that influence the duration of vowels [5]. In our analyses we compare the duration of stressed vs. unstressed vowels followed by a consonant and the duration of German long vs. short vowels in general and with voiced vs. voiceless following consonant.

Vowel duration is also very much influenced by the segmental context. In our study on 16 speakers we analyzed 9950 vowels that are followed by a consonant either within words or over word-boundaries. They are grouped according to the features stressed vs. unstressed; German central vowels /@/ and /6/ are looked at separately. Consonant classes are a) voiced/ voiceless, b) voiced/voiceless plosives, fricatives, and nasals. Our results of vowel duration measurements, which are shown in tables 2 - 6 provide further evidence to what is known from other investigations [6], [7]:

Phonologically long vowels have a longer duration than short vowels.

	number	duration
all V+C	9950	74
V long + C	3217	97
V short + C	6733	64

Table 2: number and average duration in ms of long vs. short vowels

Stressed vowels have longer duration than unstressed; the always unstressed "reduction" vowels /@/ and /6/ are shortest in duration.

	number	duration
stressed V + C	3710	96
unstressed V + C	6240	62
/@/, /6/ + C	1590	56

Table 3: number and average duration in ms of stressed vs. unstressed vowels and the German central vowel /@/ and /6/

Vowels before voiced consonants are longer than before voiceless consonants. This tendency is stronger with long vowels and not existent with short vowels.

	number	duration
V+C voiced	5595	76
V+C voiceless	4358	72
long V + C voiced	1334	115
long V + C voiceless	1883	84
short V + C voiced	4261	64
short V + C voiceless	1475	63

Table 4: number and average duration in ms vowels before voiced vs. voiceless consonants

When consonant classes are differentiated it appears that vowels have a longer duration before fricatives and nasals. This tendency is stronger when voiceless plosives and fricatives are considered separately.

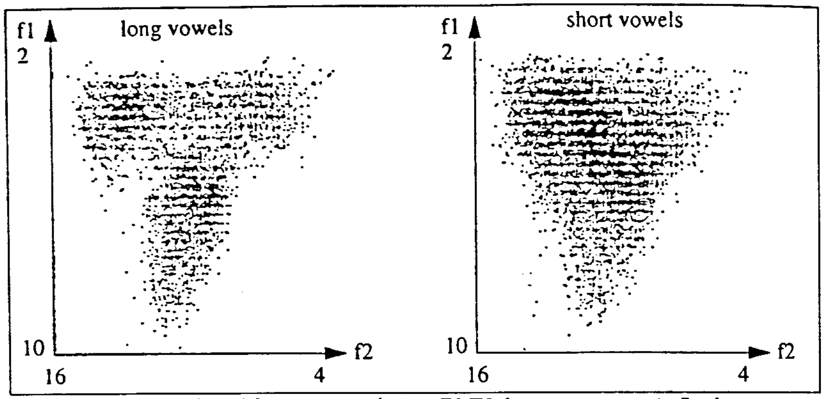

Figure 1: Scatterplot of German vowels in a F1/F2 frequency space in Bark

	number	duration
V + plosives	2586	84
V + fricatives	2557	67
V + nasals	3744	66

Table 5: number and average duration in ms of vowels before different classes of consonants

	V+C voiced		V+C voiceless	
	no	dur	no	dur
plosives	1056	89	1530	81
fricatives	88	90	2469	66

Table 6: number and average duration in ms of vowels before different classes of voiced vs. voiceless consonants

Formant Analysis

For the formant analysis we used the database tools to locate those sections in the speech signals that have been identified as vowels during the segmentation process. This information is used to apply a semi-automatic formant extraction program and measure F0 and the values of the first three formants of the vowels in PhDB. The program works as follows:

- The vowel segment plus 10 ms of the context is displayed in form of a spectrogram on a computer screen.
- A measurement point is proposed.
- F1, F2, F3 and F0 are extracted at the suggested time window in the signal.

The suggestion is based on the search for a local minimum of spectral variability as computed using cepstral difference coefficients [1]. The formant extraction is based on peak detection in a peak-enhanced 16 pole LPC-spectrum [8]. The F0-extraction is based on an autocorrelation PDA [4].

The formant values calculated by the program are marked and then checked by a phonetician who has the following options:

- accept the measurement
- select a new time location for the whole measurement by mouse click in the spectrogram
- correct the proposed formant values by mouse click at the preferred frequency in the spectrogram

The overall strategy to determine the measurement position was to find the target position of the vowel based on formant movement and energy. Altogether formant values of 10629 vowels have been extracted; reasonable F0 could be determined for 9228 vowels. Table 7 contains the average fundamental frequency and formant values of the analyzed vowels of PhDB.

Figure 1 shows the distribution of the spectral characteristics of long and short vowels in an F1/F2 scatterplot.

vowels	F0	F1	F2	F3
/i:/	185	324	2071	2698
/I/	175	369	1944	2698
/e:/	172	383	2076	2704
/E:/	176	443	2022	2660
/E/	169	486	1784	2633
/a:/	161	690	1339	2533
/a/	161	674	1362	2541
/O/	168	529	1162	2504
/o:/	165	416	927	2497
/U/	173	413	1093	2424
/u:/	199	328	946	2416
/y:/	174	341	1590	2298
/Y/	168	383	1541	2350
/2:/	177	395	1464	2249
/9/	168	477	1579	2291
/@/	178	449	1585	2570
/6/	173	535	1366	2490

Table 7: average F0 of 9228 vowels and formant values of 10629 German vowels (in Hz) for all 16 speakers of PhDB

Although the overall distribution is fairly similar, it appears by the degree of blackness that long vowels concentrate mainly in three regions: "back/round/high", "front/high" and "central/low". In contrast to this, the short vowels are distributed more regularly with a higher concentration in the centre of the vowel space.

CONCLUSION

Vowel duration and formant values of approx. 10000 German vowels have been measured. For duration measurements only the information stored in the Prolog database has been used. For formant measurement database information has been used to locate the exact position of vowels in the speech signal to apply a semi-automatic procedure for fundamental frequency and formant measurements.

Our findings of vowel duration and formant measurement support the results of earlier investigations [6], [7].

Our approach of combining symbolic database queries and acoustic analyses of speech signals has shown to be feasible and useful for phonetic research. The symbolic database has been extended with the formant data which is then available to further investigations.

The speech data of the PhonDat II train enquiry corpus is not phonetically balanced. The results we obtained may thus not be valid for spoken German in general. We plan to apply our methods to corpora with phonetically balanced data, and to larger speech corpora e.g. in cooperation with BAS [10].

REFERENCES

[1] van Bergem, D. R. (1993), "Acoustic vowel reduction as a function of sentence accent, word stress, and word class", *Speech Comm.*, vol. 12, pp. 1 - 23.

[2] Draxler, Chr. (1995), "Introduction to the Verbmobil-PhonDat Database of Spoken German", *Practical Applications of Prolog Conf. 95*, Paris.

[3] ECLiPSe (1992), *ECRC Common Logic Programming System*, User Manual, ECRC Munich.

[4] Hess, W. (1983): *Pitch Determination of Speech Signals*, Springer, Berlin.

[5] Lehiste, I. (1970), *Suprasegmentals*, Cambridge, Mass.: MIT Press.

[6] Maack, A. (1949), "Die spezifische Lautdauer deutscher Sonanten", *Zeitschr. f. Phon.*, vol. 3, pp. 190-232.

[7] Maack, A. (1953), "Die Beeinflussung der Sonantendauer durch die Nachbarkonsonanten, *Zeitschr. f. Phon.*, vol. 7, pp. 104 - 128.

[8] Markel, J.D., A., H. Gray (1976), *Linear Prediction of Speech*, Springer, New York.

[9] Pompino-Marschall, B. (1992), „PhonDat Daten und Formate", Institut für Phonetik, Uni München

[10] Tillmann, H. G. et al. (1995), "The Phonetic Goals of the New Bavarian Archive for Speech Signals", *Proc. of ICPhS*, Stockholm 1995.