Louis C.W. Pols and R.J.J.H. van Son

Institute of Phonetic Sciences, University of Amsterdam

Herengracht 338, 1016 CG Amsterdam, The Netherlands

{Louis.Pols, Rob.van.Son}@hum.uva.nl

ACCESSING THE IFA-CORPUS

Abstract

At the Institute of Phonetic Sciences (IFA) we have collected a corpus of spoken Dutch of 4 male and 4 female speakers, containing conversational as well as read speech, plus sentences, words and syllables taken from the transcribed conversation text, and then spoken in isolation. This pertains to about 5.5 hours of speech. All this material is segmented and labeled at the phoneme level. This information plus all meta data are stored in a database which makes all material highly accessible through SQL. Actually all information is freely available under the GNU General Public License for interested parties. This material will also be used in INTAS project 915, in which a comparison will be made of phonetic properties in Dutch, Finnish and Russian. As an initial result we will present some durational and spectral data of full and reduced phoneme realizations.

Introduction

In our region we are fortunate to be involved in a process of collecting about 1,000 hours of spoken Dutch (Pols 2001a). This Dutch-Flemish project (Spoken Dutch Corpus, CGN; for more details see Oostdijk (2000) and http://lands.let.kun.nl/cgn/home.htm) will result in a highly accessible abundance of speech material transcribed at various levels, from many adult speakers, in various age groups, at three education levels, and in a variety of speaking styles. However, the collection of *much* speech material from *single* speakers under *various* conditions, is not foreseen in this project. In the presently popular variable-units concatenative synthesis it is customary to collect much speech material from a single speaker, but this is most of the time application-specific and in one (read) style only. Since we were interested in studying various reduction and coarticulation phenomena as a function of speaking style, word stress, sentence accent, position in the word, word frequency, and position of the word in the sentence (Pols 2001b), we decided to collect our own IFA-corpus. However, it would of course be foolish not to make good use of all experiences collected so far. So, we followed the CGN protocols as much as possible and used available software to ease orthographic transcription, to derive a phonemic transcription and a syllable split (CELEX), to perform forced phoneme alignment before doing manual adjustment, and to automatically extract part-of-speech tags and lemmas. All speech material is accessible via the user-friendly and powerful speech signal processing package 'praat' that is developed at our institute and is freely available upon request (http://www.fon.hum.uva.nl/praat/). We also took great effort to put all non-speech data in an appropriate database structure, to make it easily and freely accessible via a WWW interface for data mining (http://fon.hum.uva.nl/IFAcorpus/). At the end of this short contribution we will give some examples of this facility and we will present some preliminary results. But first we will present some further details about the collected speech material, about the presently available label tiers, about the database structure itself and about the SQL querying possibilities. For more details see van Son et al. (2001).

Corpus content

Eighteen speakers (9 male and 9 female) participated in the recordings. Eight of them (4 male, 4 female) were selected for phonemic segmentation and constitute the present IFA-corpus. Eight speaking styles were distinguished, namely:

- informal story telling face-to-face to a colleague;
- retelling a previously read narrative story.

And reading aloud:

- a narrative story;
- a randomized list of all sentences of the narrative story;
- pseudo-sentences: replacing words in a sentence by randomly selected words with same POS tag;
- lists of selected words from the texts;
- lists of distinct syllables from the word lists;
- a collection of idiomatic (alphabet, numbers) and diagnostic sequences (V, hVd, VCV).

In Table 1 below the distribution of all segmented words per speaker and per speaking style is specified. All speech was recorded in a quiet, sound-treated room. For more details see van Son et al. (2001).

Table 1. Distribution of all segmented words per speaker and speaking style.

Speaker	sex	age	Informal	Retelling	Narrative	Sent.	Pseudo-S	Words	Syll.	Varia	All
N	F	20	660	385	2427	2850	412	262	292	356	7644
G	F	28	1850	1639	2761	2868	206	230	290	470	10314
L	F	40	885	465	2126	2078	423	239	274	387	6877
E	F	60	933	1178	2556	2765	215	261	313	432	8653
R	M	15	127	323	1348	1449	451	232	268	423	4621
K	M	40	538	435	1354	1346	-	248	275	415	4611
H	M	56	269	658	2005	2081	435	259	286	451	6444
O	M	66	-	1173	-	-	466	253	284	436	2612
All			5262	6256	14577	15437	2608	1984	2282	3370	51776

All audio-files were orthographically transcribed by hand according to the CGN protocol (Goedertier et al. 2000). The Dutch CELEX word list provided a pronunciation for most words as well as a syllable split-up, unknown words were hand-transcribed and added to the list.

The phonemic labeling and segmentation was a two-step process. An off-the-shelf phone-based HMM word recognizer was used first to time-align the speech files with the available phonemic transcription. These automatically generated phoneme labels and segment boundaries were then checked and adjusted by 7 student transcribers that got a thorough training in phoneme labeling according to the protocol. 64 Speech files were labeled twice to check consistency, for more details see van Son et al. (2001).

Apart from the meta data, presently the following levels of transcription (plus segment boundaries) are available on separate tiers and can thus be the basis for subsequent analyses:

- the orthography at the sentence level;
- the orthography at the word level
- the normative phonemic transcription at the word level;
- the syllable level, including lexical stress marks;
- the phoneme level.

Prominence marks as well as other prosodic transcriptions, via ToDI (http://lands.let.kun.nl/todi) or otherwise, will be added later.

SQL querying

With the implemented data structure and a powerful query language SQL, it is possible to answer rather intricate questions such as:

- what is the average articulation rate per sentence, expressed in number of syllables or phonemes per second, for these various speaking styles? See Table 2.
- what is the average duration of /m/ and /n/ in stressed syllables from spontaneous speech in initial, medial, and final position in the word, ignoring sentence boundaries? See Table 3.
- what is the corrected means duration of all intervocalic consonants in polysyllabic, non-high-frequent words, not at sentence boundaries, as a function of the within word position and the syllable stress, both in read as well as in spontaneous speech? See Table 4.
- what are the average vowel positions in the F1 - F2 space in different speaking style conditions? See Figure 1.

Table 2. Average articulation rate per sentence for the 8 different speaking styles.

	Informal	Retelling	Narrative	Sentences	Pseudo-Sent	Words	Syllables	Varia
Syllables/s	5.5	5.2	5.7	5.6	4.6	3.5	2.4	3.5
Phonemes/s	13.5	13.1	14.4	14.3	12.2	9.3	6.7	6.3

Somewhat to our surprise the articulation rates do not differ much between the first four communicative speaking styles, of which the first two represent conversational speech and the next two read speech. The final four non-communicative speaking styles indeed do show substantially lower rates.

Table 3. Average duration in ms of /m/ and /n/ in stressed syllables in initial, medial or final position in the word, for spontaneous speech.

	Initial	Medial	Final
/m/	71	72	87
/n/	63	66	78

Table 4. Corrected means duration in ms of intervocalic consonants (nasals, fricatives, stops, and glides), as a function of position in the word, syllable stress, and spontaneous or read speech. Between brackets the phoneme counts are given.

	Spontaneous		Read		Total
	Stressed	Unstressed	Stressed	Unstressed	phoneme counts
Initial	71 (202)	59 (96)	73 (715)	68 (285)	(1298)
Medial	63 (295)	61 (810)	69 (837)	63 (2586)	(4528)
Final	86 (20)	74 (94)	74 (75)	67 (317)	(506)

For the data in Table 4 a more complex analysis was required, we used a so-called corrected means analysis (van Santen 1992) which takes into account the unequal distribution of values in each cell. It is worth noting the long duration for the consonants in stressed syllables in word final position. However, unfortunately the number of observations is rather low for this cell. The durational measurements, as presented in the above three tables, could be derived directly from the segment boundaries. But of course also other parameters can rather easily be derived within 'praat', such as pitch, formant frequencies, intensity, or center of gravity. In Fig. 1 below we present the average vowel formant positions in F1-F2 for three speaking style conditions, namely:

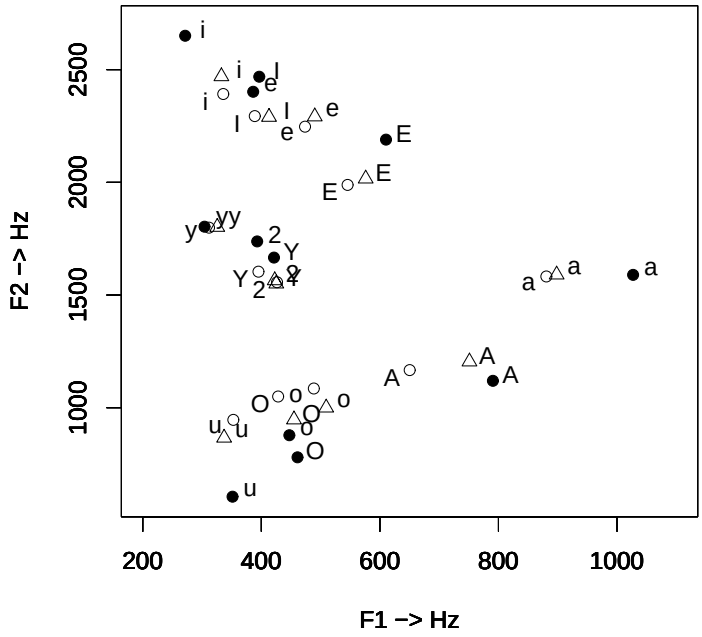

Fig. 1 Average vowel formant positions for one female speaker in three speaking style conditions. For more details, see text.

- at least 4 repetitions of clearly pronounced vowels in isolation or in spelled letters of the alphabet. See filled circles in Fig. 1;
- vowels taken from read sentences. See open triangles in Fig. 1;
- vowels taken from an informal story told by this female speaker face-to-face to an interviewer. See open circles in Fig. 1.

These data are from one of the four female speakers in this IFA-corpus. All vowel segments per condition are used for this analysis, but for the last two conditions only in multi-syllabic words and in lexically stressed position. The schwa was always excluded. The segment selection as well as the formant measurements (at the midpoint in each vowel segment) were done fully automatically. For large amounts of data this is the only possible way. However, unavoidably this might introduce some inconsistencies and errors. For instance, the average data in Fig. 1 are sometimes based on only 3 realizations (for the rare vowel /ø/, presented in the figure with the SAMPA symbol '2'), sometimes on as many as 127 (for the vowel /e/ in read sentences). Furthermore, not all formant measurements may be fully reliable. For instance, the standard deviation for the first formant measurements of the vowels /ɑ/ and /a/ in the informal speaking style is rather high, just as some of the second formant measurements for some other vowels, which may have to do with effects of reduction, coarticulation, diphtongization, or perhaps even labeling errors. But despite these imperfections, this figure nicely illustrates for 'real speech data' the large spread of the vowel space if the utterances are clearly spoken, as well as the substantially

reduced, but still easily recognizable, vowel triangle for more conversational speech. Actually we performed similar measurements for the unstressed realizations as well (not shown here), and found of course much more centralization in those conditions.

In the near future we will extend our analyses of this highly interesting speech material and we will compare the data for Dutch with those for Finnish and Russian. We will also add prosodic annotations to make this material even more useful.

Bibliography

1. Goedertier, W., Goddijn, S. & Martens, J.-P. (2000) Orthographic transcription of the Spoken Dutch Corpus, Proc. LREC-2000, Athens, Greece, Vol. 2, 909-914.
2. Oostdijk, N (2000) The Spoken Dutch Corpus. Overview and first evaluation, Proc. LREC-2000, Athens, Greece, Vol. 2, 887-894.
3. Pols, L.C.W. (2001a) The 10-million words Spoken Dutch Corpus and its possible use in experimental phonetics, Proceedings '100 Years of experimental phonetics in Russia', St.-Petersburg, Russia, 141-145.
4. Pols, L.C.W. (2001b) Acquiring and implementing phonetic knowledge, Proc. Eurospeech'2001, Aalborg, Denmark, Vol. 1, K-3-K-6.
5. Van Santen, J.P.H. (1992) Contextual effects on vowel duration, Speech Communication 11, 513-546.
6. Van Son, R.J.J.H., Binnenpoorte, D., van den Heuvel, H. & Pols, L.C.W. (2001) The IFA corpus: a phonemically segmented Dutch "Open Source" speech database, Proc. Eurospeech'2001, Aalborg, Denmark, Vol. 3, 2051-2054.