Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 26 (2005), 7-23.

SAMENSPRAAK. ACQUIRING AND IMPLEMENTING PHONETIC KNOWLEDGE

Louis C.W. Pols

Abstract

This is the slightly adapted English version of my farewell lecture that I presented in Dutch on June 3rd, 2005 when I formally retired from the University of Amsterdam. It gives a good overview of the cooperative work I did between 1982 and 2005 with my colleague staff members, post-docs and Ph.D. students. *Samenspraak*, the first word in my title, is a Dutch word implying both speaking as well as working together. The emphasis in this contribution is on acquiring phonetic knowledge via phonetics research, via speech databases, as well as in studying phonetics as a student, in order to apply and implement that knowledge in efficient speech communication, in basic research and in specific applications, as well as to improve efficient training in phonetics.

This farewell lecture on June 3rd, 2005 was presented in Dutch in the Aula of the University of Amsterdam for a big audience consisting of representatives from various organizations, colleague professors and many other colleagues, as well as students, family and friends. Amongst the guests were the chairman of the Board of the University of Amsterdam, the dean of the Faculty of Humanities (who chaired the session), as well as the president and several (former) Board members of the International Speech Communication Association ISCA. After the lecture dr. Sijbolt Noorda, the chairman of the Board of the University of Amsterdam, handed over the honorary medal for special services to the university (‘toegekend wegens bijzondere verdiensten voor de universiteit’). Later that day Cecilia Odé and Rob van Son, as guest editors, symbolically presented a special issue of the international journal Speech Communication (Odé & van Son, 2005), that was actually published in September 2005 as volume 47, issues 1-2, containing 17 reviewed papers by Dutch and international colleagues and friends.

1 Introduction

Children as well as eager adults continuously *acquire* much *knowledge* that they can profitably use, either consciously or unconsciously, in everyday life. For university students as well as for active academics, and hopefully also for retiring professors, this knowledge acquisition and the implementation of it, is almost a full-time job. In this farewell lecture I will zoom in on acquiring *phonetic* knowledge and on implementing that specific knowledge both in everyday life as well as in specific applications. This way I hope to be able to give structure to the various topics that I want to discuss. Clearly, I will look back at, as well as extrapolate from, all those

topics that I was involved in since October 1982 (when I was appointed as a full professor in Phonetic Sciences at the University of Amsterdam) and that continue to fascinate me.

For the non-phoneticians in my audience I will first of all globally summarize my *field of interest*. While doing so, I will not hesitate to interpret this field of phonetic sciences as well as the neighboring fields in a broad sense. After all, phonetics is very much an *interdisciplinary science* that has as its fields of interest the speech signal and the spoken language as well as speech communication in general, but it is also about phonemes and prosody. It is about speaking and listening and about mental storage and retrieval. It is about early speech acquisition and about speech pathology, but also about (language and) speech technology and speech databases, whether or not multimodal or audiovisual. It is about languages of the world and their dialects, and furthermore about all those topics that phoneticians appear to be good at, such as laboratory phonology, evaluating the use of cochlear implants, or designing speaking web avatars.

The Chair of Phonetic Sciences, which from 1958 until his death in 1980 was occupied by my predecessor prof. Hendrik Mol, is now part of the department of Linguistics (opleiding Taalwetenschap) within the section of Linguistics and Literature (afdeling Taal- en Letterkunde) of the faculty of Humanities (faculteit der Geesteswetenschappen) at this University of Amsterdam. But within another university structure, of which there are several examples in the world, we could also have been part of Psychology, or Informatics, or Medicine. Actually, also here in Amsterdam there are active contacts with all these faculties, be it in teaching and/or in research.

Speech appears to be a fantastically informative and efficient acoustic signal for allowing linguistic communication. This does not just concern communication between humans (be it children or adults), but also between humans and machines. In all these cases speech is not just text-transformed-into-sound, but speech also contains all kinds of information about the speaker (e.g. sex, age, weight), his or her health and emotional situation (e.g. having a cold or being nervous or angry), about the acoustical environment (e.g. resonant room or free field) or about the communication channel that is being used (e.g. mobile phone or direct sound) (Steeneken, 1992). Furthermore, the speech signal also contains information about the local and global speaking rate, about the amount of emphasis or reduction of syllables or words, about the actually realized sentence accent, about the sentence being a question or a statement, about whether the speaker is a native or a dialect speaker, about whether this is perhaps machine speech or natural speech, and about many more aspects.

A striking difference between phoneticians and most other linguists is that phoneticians consider all these variations in the speech signal to be fascinating and challenging, whereas most other linguists like to abstract from them as much and as soon as possible. Nevertheless, also for phoneticians it is very important to disentangle how this extremely variable speech signal can be transformed into normative observations in terms of phonemes, syllables, words, sentences and dialogues and how young children and non-native speakers can learn to speak and to understand.

Also the way in which the variability within and between speakers is handled by *listeners*, and even more so in *speech technology*, is a research problem that still has not been solved. It is also fascinating to realize that on the one hand very subtle details in the speech signal are essential, e.g. to distinguish between ‘hij spreekt die kerel soms’ (single) or ‘hij spreekt die kerels soms’ (plural) (Kemps, 2004), or, to mention something totally different, to break in at the right moment into a dialogue with a so-called turn. While on the other hand *highly reduced speech* not at all has to

lead to misunderstandings (e.g. by saying [x@mEnt] rather than 'op een gegeven moment'). All that underlying phonetic knowledge about the process of speaking and listening and about the speech signal itself, has to be acquired by first and second language learners, language users and speech and language technological systems.

This brings me back to the main theme of my lecture, the acquisition of phonetic knowledge and its implementation. For an overview see Table 1.

Table 1. Overview of the topics to be discussed

acquiring phonetic knowledge in	*implementation in/for*	*see section*
early speech development	efficient linguistic communication	2
speech pathology	deviant communication	2 & 3
phonetics research	basic knowledge or in applications	4
speech databases	rules or otherwise	5
studying phonetics	educational purposes	6

2 Early speech development and speech pathology

How does a child acquire phonetic knowledge? My guess would be that a child learns by experience, although of course education and training also play a role. In our Chair there has always been a lot of attention for early speech acquisition, especially in the first year of life. This is exactly the period for which most linguists have little attention because grammatical structure, word forms and word meaning are then not yet easily recognizable. However, it is exactly the period in which the bases for prosody, word and sentence structures and linguistic communication are created via mother-child interaction and exploration.

Apart from early speech acquisition for normal children (van Gelderen, 1992; van der Stelt, 1993), much research has been done in our group also with children with a physical and/or auditory handicap, such as cleft palate (Jansonius-Schultheiss, 1999), as well as with hard-of-hearing children (Clement, 2004). Almost all this work has been supervised in a very efficient way by my colleague and associate professor Florien van Beinum, who retired in June 2004, on which occasion a Symposium was held with the appropriate title "Spraak in the maak" (Speech in the making). Therefore I prefer to concentrate here on another aspect in which *my own* specific knowledge about signal and data processing fits nicely with the need to describe and illustrate in an insightful way the growth of these children to adult speech production. Baby vocalizations cannot simply be labeled as phonemes or words, but at best only in terms of respiration, phonation and articulation. Those acoustic realizations frequently show neither a clear periodic character nor do they have a clear formant structure (formants represent the resonances of the mouth, nose and throat cavities). This poorly defined formant structure is mainly caused by the high pitch of these voices, which can rise to 400 Hz and higher. Manual measurement of the formants of such sound segments is then easily influenced by the supposed vowel quality that the researcher believes to hear in the signal. In the early sixties my colleagues and I at TNO (Plomp et al., 1967) developed a spectral analysis method that is based on the whole spectrum as measured with some 20 bandfilters, and is followed by some form of data reduction to a small number of dimensions. This method is highly objective and reproducible and appears to show great similarity with a formant representation (Pols et al., 1973), whereas it also can very well describe the perceptual similarities and confusions between sounds (Pols et al., 1969). This is less surprising than it might

be since the ear is much more a bandfilter analyzer than a formant analyzer (van Dijk, 2001). However, also this bandfilter analysis method experiences difficulties in handling periodic signals with high F0. Fortunately Ton Wempe has developed a procedure that does allow measuring the envelope spectrum (Wempe & Boersma, 2003).

With this approach unlabeled vocalizations of young children at an age of several months up to two years can be measured automatically and can be displayed in a formant-like representation. In this way we can nicely illustrate how a limited vowel space for five months old children, with mainly /i/- and /u/-like vocalic sounds, grows to a rather complete vowel space at two years of age. Because at two years of age, words and sounds can already be identified rather well, we can easily add to this collection of *unlabeled* segments *labeled* vowel segments for reference. In a similar way the development within and between *normal hearing* children can be compared with that of *hearing-impaired* children (van der Stelt et al., 2003). Because of the good contacts that Jeannette van der Stelt already has with research groups in the United States, France, Hungary and Russia, and because of the fact that these groups have similar children speech databases, also the similarities and dissimilarities between languages can be studied very well (Zajdó et al., 2005).

So far we have limited ourselves to a maximum of 10 isolated measuring points per utterance, without giving much attention to the temporal structure *within* utterances. But, just as we can represent a single spectral measurement as a point in a space, we can also represent subsequent measurements as a trace of points in that space, which may give us insight in articulatory dynamics as well.

3 Adult speech pathology (dyslexia, laryngectomy)

Additional to the early development of child language, our group also has shown a strong interest in the production and perception of pathological and normal speech of growing-up children and of adults (Bakkum et al., 1993; Clement, 2004; Jansonius-Schultheiss, 1999; Tielen, 1992; Verdonck-de Leeuw, 1998). In this section I will limit myself to two totally different forms of pathology, namely dyslexia and laryngectomy.

Especially because of the involvement of Florien van Beinum in the Steering Group of the NWO-program Dyslexia, we have searched for the most adequate stimuli for the early detection of dyslexia via listening experiments. For this we based ourselves upon the Ph.D. work of Cecile Kuijpers (1993). For children between 4 and 12 years of age she looked at the *perception* of the voicing contrast in manipulated stimuli (the child heard a manipulated stimulus that sounded like ‘patto’ or ‘paddo’ and then had to point to one of two dolls with the learned names Patto or Paddo). She also looked at the *production* of voice assimilation in Dutch words like ‘leesboek’ or ‘groot beest’.

For the dyslexia program Caroline Schwippert (2000) created continua from ‘bak’ to ‘dak’, starting from a natural utterance ‘bak’. Identification and discrimination experiments were performed with these stimuli with adult dyslectic subjects as well as with a control group. Only under critical conditions of small step size and short duration of the formant transition, the performance of the adult dyslectics were less good. Nevertheless these signals seem to be adequate for early detection with young children (van Beinum et al., 2005).

If for people that suffer from throat cancer the vocal cords have to be taken away surgically, they can no longer talk, because on the one hand the trachea no longer ends in the throat but in a stoma in the neck, whereas on the other hand the voice

source in the form of vocal folds is no longer available. Fortunately nowadays already during surgery a so-called speech valve is installed between trachea and esophagus, through which air from the lungs can stream via the esophagus into the vocal tract as long as the stoma is kept closed. Whether or not with some operative help, this air can be made to vibrate a stricture in the esophagus, called the neo-glottis. This simple neo-glottis, that has very little in common with the complex control mechanism of the vocal folds, can nevertheless function as an alternative voice source, which allows most patients to communicate in a rather intelligible way. After the pioneering work of Corina van As (2002) about the voice quality of these patients, Petra Jongmans now continues this research with an emphasis on phoneme and word intelligibility and on speech naturalness (Jongmans et al., 2005). She will also study the physical processes of this alternative way of voicing. Of course one of the goals is to further optimize speech training. My colleague prof. Frans Hilgers (2004) presented much more information about laryngectomy during his inaugural lecture on September 23rd, 2004. Let me just point out here that it is almost a miracle how good this speech frequently sounds, since much of the flexibility to phonate, to vary pitch and to shift between voiced and voiceless is lacking.

4 Specific phonetics research

Since my appointment in 1982 I was fortunate to have the opportunity to perform much specific phonetics research, together with my Ph.D. students and post-docs and to report about that at conferences and in the open literature. Below I will present in a number of subsections some salient points of that research as well as possible consequences for future research.

4.1 jnd, reduction, coarticulation, and prominence

A speaker speaks as sloppily as the listener allows him to do. Similarly the speaker tries to adapt as much as possible to environmental conditions, such as a reverberant church or a telephone connection. It appears that high-frequent and/or more-predictable words are pronounced less clearly than low-frequent and/or less-predictable words (van Son & Pols, 2003). Furthermore, there appears to be a, mainly articulatory-based, neighbor interaction or co-articulation that for instance implies that a ‘k’ in ‘koe’ sounds different from a ‘k’ in ‘kie’, and that the ‘ie’ in ‘dief’ sounds different from that in ‘dier’. There has been much discussion in literature whether co-articulation originates from articulatory limitations (a speaker can only move his articulators at a certain speed from one place to another), or is more a matter of laziness on the speaker side, or simply a matter of efficiency. Both Rob van Son and I strongly believe in the latter, since at normal- and even at fast-rate conversations the speaker almost never meets the limitations of the movability of his articulators (Pols & van Son, 1993; van Son, 1993; van Son & Pols, 1992). This also implies that, according to us, the Lindblom model (Lindblom & Studdert-Kennedy, 1967), based on articulatory undershoot compensated by perceptual overshoot, does not hold. Phoneme and word recognition thus cannot simply be template matching but requires that the listener takes into account local and global context.

In analytical listening tasks using artificial stimuli it is frequently shown how sensitive the listener is to small differences in the signal, as expressed in the so-called just noticeable difference or jnd. For more speech-like and thus more complex stimuli, this jnd quickly becomes larger. Astrid van Wieringen has shown that for

short formant transitions around 1000 Hz, the jnd is more than 200 Hz (van Wieringen, 1995; van Wieringen & Pols, 1995), thus diminishing the need for a very precise perception of each formant frequency. This is probably more efficient after all, how else would we be able to categorize phonemes with such a variable input, especially with good ears?

Dick van Bergem (1995) did pioneering work on *vowel reduction*. He showed that such a reduction does not simply imply a more central location of the vowel, but rather involves a substantial adaptation to the consonant environment. So, for instance the reduced vowel /I/ in the first syllable of the Dutch word 'miljoen' physically is not so much a /@/ but rather an /u/ because of the /m-l/ environment. He also showed that for instance the F2-transition in the schwa of /C1@C2VC/ words like 'patat' /p@tAt/ can be modeled and thus predicted by taking into account the effects of C1=/p/, C2=/t/ and V=/A/. Almost 80% of the variance in the central position of F2 can be explained this way. I consider this to be an excellent example of *computational phonetics* and I would like to see more of our phonetic measurement results being presented this way.

It should be no surprise that, next to *vowel* reduction, there is also something like *consonant* reduction. In the post-doc project of Rob van Son this is exactly what was studied systematically for the first time (van Son & Pols, 1997, 1999a).

The opposite of reduction is *prominence*, this involves *emphasizing* a syllable because the pronunciation of the word requires that (one does not say 'uit*spraak*' but '*uit*spraak' in Dutch), or, because for instance the presence of new information in a sentence makes that desirable. Barbertje Streefkerk studied up to what level this desirable prominence can be predicted from text on paper, can be detected in a spoken text, and can be optimized in speech synthesis (Streefkerk, 2002; Streefkerk et al., 2001). Apparently there is ample room here for new research questions.

4.2 phoneme perception and vowel systems

In the initial years of my appointment in Amsterdam I have given much attention to vowel and consonant perception with natural and with manipulated stimuli (Bakkum et al., 1993; Bloothooft, 1985; Klaassen-Don, 1983; Pols, 1983; Pols & Schouten, 1987; Pols & van Son, 1993; Schouten & Pols, 1983; van Son & Pols, 1999b) and to the structure of vowel systems (ten Bosch, 1991). Most of this work was done in cooperation with Florien van Beinum and Bert Schouten and with several Ph.D. students. However, since I cannot cover all topics extensively in this lecture, I refer for more information about these interesting themes to the references given above.

4.3 speaker normalization

Most automatic speech recognition systems lack robustness, this implies that their performance immediately degrades under conditions that somewhat deviate from the training data. One aspect of such a deviant condition concerns variation *between* speakers. Generally, *speaker-independent* systems can only perform properly with a *small* vocabulary, whereas *speaker-specific* systems can only handle a *large* vocabulary once extensively trained.

Also the human listener is somewhat sensitive to this phenomenon: If a series of vowel stimuli is presented in a *blocked* way, implying the presentation of subsequent vowels all from the same speaker, then recognition is significantly better than with a so-called *mixed* presentation, when the speaker varies from stimulus to stimulus.

Still, human listeners are extremely good at quickly adapting to ever changing speaker signals, implying that in terms of physical characteristics the same phonemes and words vary considerably over speakers. The most extreme example of that is probably the communication between a father and a young child: The father with a heavy, low voice and a big mouth, the child with a high voice and a tiny head.

In the soon to be completed Ph.D. project of David Weenink various ways in which this speaker normalization might work, have been systematically studied. A big problem is to find an optimum between on the one hand quick learning, adapting and generalizing on the basis of a *small amount* of new information, while on the other hand avoiding to forget too quickly already derived knowledge (Pols & Weenink, 2005).

4.4 discourse information structure and turn-taking behavior

Research in phonetic sciences has quickly developed in the past 23 years of my appointment. Next to the speech technological developments and the use of large speech databases (more about this below), this is most clearly visible in the much bigger attention for natural speech and for communicative processes, such as prosodic aspects of discourse, the Ph.D. project of Monique van Donzel (1999) and turntaking behavior in dialogues, the new Innovational Research Incentives (VIDI) project of Rob van Son (Wesseling & van Son, 2005).

Van Donzel studied the information structure in short stories retold by four male and four female Dutch speakers, both in terms of various overall acoustic-prosodic measures that the speakers use (such as pitch accents and pauses), as well as in terms of the listeners' interpretation of that (such as perceived prominence and discourse markers). For practical reasons this extensive study had to be limited to pausal and intonational features. Other important aspects, such as spectral and segmental information, could not be taken into account. The same was true for the linguistic structure of the message, in terms of for instance word order or grammatical structure. Furthermore, this project was limited to monologues without interrupts from another speaker. Clearly, on all these points extensions are possible.

These aspects will at least partly be covered in van Son's VIDI-project about dialogue turn-taking behavior. The *phonetic* approach that is also chosen in this project is evident from the experimental procedure used. In order to understand better the acoustic and linguistic aspects that are the earliest indications of a possible turn to be taken, subjects are asked to indicate as quickly as possible when they would like to interrupt while listening to the dialogue. They do this by saying 'ah' while a laryngograph registers their vocal cord activities. First results indicate that the earliest vocal cord activities take place surprisingly early in time, which suggests that apparently listeners in an early phase are already planning and thinking ahead. Which acoustic, phonetic and linguistic dialogue aspects are most responsible for this, is of course point of further research. One of the ways to study this, is by manipulating the speech signal in such a way that this becomes totally unintelligible while at the same time the information about pitch and pauses is fully maintained. It is also planned to evaluate what information is contributed to the dialogue via additional video recordings of the speaker. The results of this study will hopefully lead to more efficient and more natural human-computer dialogues.

4.5 speech technology (recognition, synthesis evaluation)

Within the present structure of our faculty of Humanities it is hardly realistic to suppose that at an Institute of Phonetic Sciences there will be sufficient means and expertise available to provide an adequate technological contribution to the development and the operationalization of complex language- and speech-technological systems. Nevertheless I do believe that Phonetics is indispensable for providing essential contributions to *specific* aspects of such systems. The brute force probabilistic approach that is based on large amounts of highly variable training data, is still very popular and rather successful (van Alphen, 1992). Still, there are specific aspects of phoneme and word modeling and there is specific knowledge of the speech signal, that do make a phonetic input potentially highly profitable.

For example, specific information about the *duration* of individual phonemes can be a valuable contribution to recognition, although right now this only contributes in the form of a normal distribution of phoneme duration in which all possible conditions are put together. The fact that vowels in stressed syllables, or in words in utterance-final position, or when speaking at a slower speaking rate, are systematically longer in duration, is not used at all. Xue Wang (1997) showed in his doctoral thesis that some profit in performance can be achieved in a post-processing phase, by selecting from the N-best set of sentences, that sentence for which the actual durational information fits best.

It is still the ideal of many phoneticians to include also other *prosodic knowledge* in the automatic speech recognition process. Unfortunately the necessary specific and preferably error-free information can not yet simply be derived from the speech signal (Reetz, 1996).

Already for over 30 years the parametric representation of the speech signal in speech technological systems is almost exclusively based on linear predictive coding, or variants of this such as mel-frequency-based cepstral coefficients. Already in my own doctoral thesis (Pols, 1977) I have indicated that a principal components analysis based on bandfilter data is almost synonymous to MFCC.

It might be interesting to point out here that the first American commercial application of LPC analysis/resynthesis by Texas Instruments was not a computer application, but a nicely designed toy called Speak & Spell that, when it came on the market in 1978, was not only immediately embraced by my children, but was also used in 1982 by Steven Spielberg in his famous movie ET to simulate extra-terrestrial communication by children on earth.

Partly within the framework of the Dutch SPIN-ASSP program and the European programs SAM and SPIN, I have spent some 15 years together with Renée van Bezooijen and others on evaluating speech synthesis systems (e.g. van Bezooijen & Pols, 1991; Pols, 1991, 1994, 2000; Pols & Olive, 1983; Pols & SAM-partners, 1992).

I see the Synthesis Workshop in Jenolan Caves, Australia in 1998 as a conclusion of all that work (van Santen et al., 1998). There, for the first time an on-site evaluation took place of some 40 systems in eight different languages (American- and British-English, German, French, Dutch, Spanish, Chinese, and Japanese) with three representative types of texts, namely newspaper texts, semantically unpredictable sentences such as 'the plane closed the fish that lived' (Benoît et al., 1996) and telephone book information. Good initial speech quality is nowadays achieved by the popular concatenative synthesis, however, the problem is now shifted to things like how several voices and speaking styles can be achieved efficiently, how emotion can be implemented (Mozziconacci, 1998) and how naturalness can be improved.

4.6 languages and dialectse

Most phonetic research methods are language independent and most phoneticians prefer to do research on their own mother tongue or on English. Still we do have several language-specific projects. Partly because of the presence of Cecilia Odé in our group since 2002, we now study Russian intonation and its communicative function (Odé, 2003). Through our participation in the European INTAS program we have executed a comparative study on a number of phonetic properties of Finnish, Russian and Dutch (van Son et al., 2004). Within the NWO Endangered Languages program, Odé applied for a new project about Tundra Yukagir, which was unfortunately not honored. Recently another NWO-project, a so called small-scale interrelated program presented by Ben Hermans, Marc van Oostendorp and Paul Boersma, about 'Tone and intrasegmental structure in West-Germanic Dialects', such as Limburgian, was honored. Linguistic variation was also at stake in a recently completed MA-project by Jelske Dijkstra (Dijkstra et al., 2004) about Frisian text-to-speech synthesis, that won the biennial MA-thesis prize of the Fryske Akademy. For several other language-specific projects like this, I refer to Botha (1996), Elgendy (2001), Kovačić et al., (2003), Ölander et al. (1999), and Rechziegel (2001).

4.7 phonological aspects of speech acquisition

Without pretending at all that my present Chair involves Phonology as well, it is unmistakable that with the doctoral thesis of Paul Boersma (1998) about functional phonology, and even more so with the approval of his project 'Adequacy and acquisition of functional constraint grammars' within the NWO Innovational Research Incentives Scheme in 2002, high quality phonological research using Optimality Theory has entered our Chair. It is remarkable and inspiring to see how adequate his Gradual Learning Algorithm (Boersma & Hayes, 2001) is able to properly model the phoneme acquisition of second language learners (Escudero & Boersma, 2004), as well as the early acquisition of word stress in languages such as Latin (Apoussidou & Boersma, 2003) and Pintupi, and hopefully soon also for Dutch.

5 Speech databases

A long time ago I started my academic career at the TNO Institute for Perception (Instituut voor Zintuigfysiologie) in Soesterberg at the Speech and Hearing department under the inspiring leadership of prof. Reinier Plomp. Later on that institute was called Human Factors (Technische Menskunde), and after a recent re-organization of TNO, everything changed again. My first serious publication in The Journal of the Acoustical Society of America in 1967 (Plomp et al., 1967) was about the spectral analysis of 12 Dutch monophthongeal vowels spoken by 10 males in words of the type 'hot', 'haat' and 'hoot'. That small set of 120 words was my very first speech database. Later that set was extended to a vowel set of 50 male (Pols et al., 1973) and 25 female speakers (van Nierop et al., 1973). Even today one regularly refers to these bandfilter and formant analysis data of the Dutch vowels and to the related perceptual data (Pols et al., 1969). In my doctoral thesis (Pols, 1977) the word set was extended to 270 CVC words of the type 'pan', 'heer' and 'nu', while also manipulative synthesis was introduced and later also automatic speech recognition (Pols, 1979). This type of analysis and data representation is now very popular again

for our analyses of child vocalizations, for laryngectomized speech, for Polder Dutch, as well as for speaker normalization.

In the many years of cooperation with Ron van Son, now post-doc, several other annotated speech databases have been created with his substantial support, such as a text spoken at a normal and at a fast rate by the well-known radio and movie commentator Philip Bloemendal, in order to study the possible processes of acoustical undershoot and perceptual overshoot (van Son & Pols, 1992). Or, the so-called IFA-corpus, in which some eight speakers spoke texts in a variety of ways, from isolated syllables, words and sentences, to read and spontaneous speech. In his latest VIDI-project attention will be given not just to a selection of existing dialogues from the CGN-corpus, but also to new dialogues recorded audiovisually. Special attention will be given in this project, jointly executed with the Ph.D. student Wieneke Wesseling, to turn-taking behavior.

In overviewing this list, it is unmistakable that we go from highly structured but rather unnatural speech material to rather unstructured but much more natural conversations. It is a real challenge to make all this material accessible via transcriptions, to extract specific knowledge from it and to ask the right questions.

The existence of the IFA-corpus made it possible for us to participate in a sensible way, but without too much additional effort, in an international cooperation with Finnish and Russian colleagues, INTAS 915, that was meant to compare various phonetic properties over three totally different languages (van Son et al., 2004). Meanwhile, the IFA-corpus with about five hours of speech has been included in the Spoken Dutch Corpus (CGN), which consists of almost 1000 hours of speech from Dutch and Flemish speakers (Oostdijk et al., 2002). It was my pleasure as a member of the CGN Steering Committee to contribute my share to this mega project. All available annotations and all audio files are now put together on 33 DVD's. Meanwhile at various places including IFA, this material has been made directly accessible via a network server by storing everything on a large disk. In this way this corpus can for instance also be used for the Polder Dutch project, a Ph.D. project of Irene Jacobi with Jan Stroop as her first supervisor (Jacobi et al., 2005).

Table 2. Frequency of occurrence and average duration of various realizations of the Dutch expression '*op een gegeven moment'* in the CGN-corpus.

utterance type	*count*	*perc.*	*av. dur. (sd) in ms*
Op @n x@x'ev@(n) mom'Ent	0	0	-
Op @ (x@)x'ev@ mOm'En(t)	19	4	947 (185)
(Op) (@) x'ev@m'En(t)	210	44.3	548 (126)
Op (@) xem'En(t)	178	37.6	394 (90)
Op (@) xev@	11	2.3	362 (117)
inappropriate segments	56	11.8	-
Total	474	100	495 (170)

As an illustration for this lecture, I have searched within the CGN-corpus for a clear example of phoneme and syllable reduction. During my student courses I have regularly claimed that the seven-syllable expression '*op een gegeven moment*' is frequently reduced in conversational speech to something like [Op xev@m'Ent] with just three or four syllables. With the exploration program COREX it is relatively easy to search the CGN-corpus for such sequences, whereas Rob van Son developed additional software to make this and the IFA-corpus better accessible via SQL. The expression *'op een gegeven moment'* occurs about 1200 times in the CGN-corpus, of

which about 475 times in spontaneous conversations recorded by microphone between two or more Dutch speakers. Table 2 shows the frequency of occurrence, as well as the average duration of the various varieties. It will be clear that all kinds of rather extreme forms of reduction are quite common, whereas the full form never happens!

Next to these rather recently acquired speech corpora, it is worth mentioning another interesting development. That is making available a historical collection of all kinds of speech recordings that were stored under rather poor conditions on the top floor of our institute's building below the tiled roof. These recordings contain more than 200 hours of speech and date back to the twenties of the previous century. They have been recorded on such diverse carriers as wax rolls, glass plates, aluminum, vinyl and polyester records, as well as paper tape and magnetic tape. With funding, acquired by Marian Schilder of the University Museum via the so-called Nuis-gelden, a large part of these recordings have been digitized at the Meertens Institute. So far all these audio files have been stored in a not very efficient way on more than a thousand CD's. Soon all these files will be transferred to one big disk on our server. Unfortunately the metadata are very incomplete, as frequently is the case with historical material, so a systematic search is rather difficult. Also funds are lacking to transcribe parts that seem to be interesting, not to speak about transcribing everything. Despite all this, several interesting parts have already been discovered. This allowed my colleague prof. Frans Hilgers to play during his inaugural address (Hilgers, 2004) unique recordings on wax rolls dating back to 1925 of spoken and sung texts of a laryngectomized patient. This well-documented case in the literature almost certainly concerns the oldest recording of its kind. Furthermore, Ton Wempe en Paul Boersma could trace several old recordings of speakers from Venlo, which may be an interesting addition to the just started research project on present-day Limburgian.

6 Educational training in phonetics

Even more so than in *research*, specific aspects of knowledge acquisition and implementation, are important in *educational training*. Phonetic knowledge is an indispensable aspect of linguistic knowledge, and it is therefore appropriate that three phonetic modules 'Speech', 'Phonetic Transcription' and 'Speech processing by computer' are part of the present bachelor program in Linguistics in Amsterdam. Although our Chair is part of the department of Linguistics (opleiding Taal-wetenschap), this is most certainly not the only framework in which we operate. No one will be surprised that, because of the interdisciplinary nature of Phonetics, there are also good teaching connections with Informatics and Medicine whereas cooperation with Psychology and Cognition has been explored. At the interfacultary department of Artificial intelligence we offer the bachelor module 'Speech recognition and speech synthesis', as well as the modules 'Spoken Language Generation' and 'Speech Technology project' in the master specialization 'Language and Speech'. We expect that actually more students may show an interest in this specialization than in the specialization 'Speech Technology' within the MA Linguistics in our own faculty. So far almost all students with a proper phonetic training could easily find a job in research or in the society.

7 Future

Finally a word about the future. Unfortunately, it is quite common these days, both here in Amsterdam as well as elsewhere in Holland, that the underlying feelings at a farewell lecture are rather gloomy, because the firm desire to continue the specialization at stake is not clearly expressed by the board of the faculty because of a chronical money shortage or new ideas about the curriculum. In my case there is fortunately much less reason to be gloomy, because the board of the faculty together with the board of the university, made the timely decision to appoint Paul Boersma as my successor. Since this appointment became effective already at June 1st, 2005, Paul Boersma could gladfully join the cortege of professors at my farewell lecture. Because he is one of my doctoral students (Boersma, 1998) with a great international reputation, I can only be very happy and grateful about these developments. Because Boersma is not just a capable phonetician, but also a renowned phonologist, and furthermore is also responsible together with David Weenink for the highly appreciated speech signal processing software package *Praat* with more than 10,000 satisfied users all over the world, there is every reason to believe that the Chair of Phonetic Sciences within the department of Linguistics will foresee a bright future under his leadership.

I am pleased to be able to leave the Chair with four externally financed post-docs and four doctoral students. Worrisome is the fact that because of early retirement the permanent staff will be seriously reduced by the end of 2005, which will have consequences for the educational program that can be offered. However, for these future developments I am not responsible anymore, I just have to properly summarize this presentation.

I have shown clear trends in the past 25 years as far as phonetics research is concerned: attention is shifted more and more from simple stationary signals to complex dynamic signals, from laboratory speech to natural speech (Plomp, 2002), from rather artificial tasks for the listeners to tasks that are more and more communicatively relevant. Because of much faster computers and much more storage capacity, the speech material that is used these days for analyses has risen exponentially in size, diversity and complexity. A phonetician can no longer limit himself to the speech material as such, but has to be aware of its linguistic, communicative and multimodal aspects as well. The easily available software has become much more powerful and user friendly. Partly because of that we no longer have to limit ourselves to incidental observations, but phenomena can be described much more systematically and we can apply *computational phonetics* (Pols, 2004a). Present-day scientific research stands much closer to industrial, clinical and educational applications.

Next to all these new developments I cannot avoid establishing the fact that many old problems remain. One year after my appointment I was greatly inspired by the MIT symposium in 1983 entitled ‘Invariance and variability in speech processes’ (Perkell & Klatt, 1986; Pols, 1986). This problem of large variability in production next to necessary forms of invariance for perception and recognition, still keeps us busy.

The present curriculum of the broad bachelor and the short specialized master do not allow to pay much attention to the latest developments in our field. The student with a BA or MA certificate will still have to acquire much additional knowledge and experience before becoming useful in a specific research project or in society. Within the newly proposed sequel of 3 years BA, then 2 years research MA, plus 3 years for a doctoral thesis, it is questionable whether the high demands made so far on the quality of doctoral theses can be maintained in the future, the more so since a timely

completion becomes more and more mandatory. In the past three years still about seven doctoral theses have been completed in the Netherlands in the phonetics domain (Pols, 2004b), but I am afraid that this number will decline drastically in the future because the reservoir of properly educated phoneticians will dry up.

8 Thanks

In this last section I want to express my thanks. The first word in the title of my lecture was *samenspraak*. Especially for phoneticians this is a nice word in the Dutch language. It properly expresses how I have tried to function in my group with the permanent and temporary staff, now and in the past. I cannot function *alone* and I have to work *together* with others to achieve joined targets. I hope that as a supervisor I was able to stimulate and motivate people to reach their goals. I myself have learned a lot from the enthusiasm, the perseverance and the ingenuity of the permanent staff and the supporting personnel, as well as the doctoral students, post-docs and guest researchers with whom I had the privilege to work since 1982.

It was also a pleasure to function in various cooperative communities over the years, such as the former working community on Phonetics and the NWO Foundation on Linguistics, the speech expertise center SPEX, the Foundation on Speech Technology, the research institute IFOTT (now ACLC), the Steering Committees of such national programs as SPIN-ASSP, CGN and IMIX, and the active participation in European programs such as SAM, SPIN, Eagles, and EuroCocosda. Especially interesting and excellent for networking was my participation in the NWO Council for the Humanities, my membership of the Royal Holland Society of Arts and Sciences and my role in the Senate of this university.

I got my training as a scientist from my teacher Reinier Plomp, head of the department of Audiology when I came to the Institute of Perception in Soesterberg in 1965. He was also my later promotor at the Free University of Amsterdam. He taught me the arts of reasoning, experimenting and publishing. Already in 1970 he took me to the United States to visit speech laboratories and conferences, and he always strongly supported international contacts. This led, among other things, to my intensive involvement in the founding in 1987 and the later growth of the European Speech Communication Association ESCA, later ISCA, and in the workshops organized by that association, amongst them the one in Noordwijkerhout in 1989, as well as the big international conferences initially called Eurospeech and now Interspeech.

A second person to whom I deserve much gratitude is my comrade for many years at the IFA and my travel mate during many jointly visited scientific meetings, namely Florien van Beinum. Together most of the time we could handle the diversity of topics in educational teaching and in research.

Furthermore I am devoted to all my doctoral students, because each completed dissertation almost always marked a long period of intensive cooperation with a tangible result. Three of them I would like to mention specifically, because the cooperation with them took especially long and the joint effort in experimenting, data processing and publishing was especially substantial, namely Louis ten Bosch, Rob van Son and Astrid van Wieringen.

Running a ‘vakgroep’, or as it is called now a ‘leerstoelgroep’, within a complex university structure with departments, sections and teaching and research institutes, as well as national and international connections, is a fascinating and challenging job, but it is also time consuming and sometimes tiresome. It is then especially important that your family at home supports you as much as possible, but also demands

attention for totally different matters. I was glad to always have had substantial support from my wife Anneliese and we are the happy parents of two charming daughters Edith en Mirjam that now are adult ladies but continue to provide us much joy with their families with meanwhile four grandsons. Of course they were present during the June 3rd festivities, but I am still exalted by the fact that our little family was also complete when I was honored in Aalborg during Eurospeech 2001 by receiving the ISCA medal for scientific achievement.

In short, I am a happy person that looks forward to all the nice things that lie ahead.

References

Alphen, P. van (1992), *HMM-based continuous-speech recognition. Systematic evaluation of various system components*, Ph.D. thesis Univ. of Amsterdam: 216 pp.

Apoussidou, D. & Boersma, P. (2003), "The learnability of Latin stress", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 25:101-148.

As, C. van (2002), *Tracheoesophageal speech. A multidimensional assessment of voice quality*, Ph.D. thesis Univ. of Amsterdam: 211 pp.

Bakkum, M.J., Plomp, R. & Pols, L.C.W. (1993), "Objective analysis versus subjective assessment of vowels pronounced by native, non-native, and deaf male speakers of Dutch", *J. Acoust. Soc. Amer.* 94(4): 1989-2004.

Beinum, F.J. van, Schwippert, C.E., Been, P.H., Leeuwen., T.H. van & Kuijpers, C.T.L. (2005), "Development and application of a /bAk/–/dAk/ continuum for testing auditory perception within the Dutch longitudinal dyslexia study", *Speech Communication* 47 (1-2): 124-142.

Benoît, C., Grice, M. & Hazan, V. (1996), "The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences", *Speech Communication* 18(4): 381-392.

Bergem, D. R. van (1995), *Acoustic and lexical vowel reduction*, Ph.D. thesis Univ. of Amsterdam: 195 pp.

Bezooijen, R. van & Pols, L.C.W. (1991), "Evaluating text-to-speech systems: Some methodological aspects", *Speech Communication* 9(4): 263-270.

Bloothooft, G. (1985), *Spectrum and timbre of sung vowels*, Ph.D thesis Free Univ. of Amsterdam: 169 pp.

Boersma, P. (1998), *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives*, Ph.D. thesis Univ. of Amsterdam: 493 pp.

Boersma, P. & Hayes, B. (2001), "Empirical tests of the Gradual Learning Algorithm", *Linguistic Inquiry* 32: 45-86.

Bosch, L. F. M. ten (1991), *On the structure of vowel systems. Aspects of an extended vowel model using effort and contrast*, Ph.D. thesis Univ. of Amsterdam: 190 pp.

Botha, L. (1996), "Towards modelling acoustic differences between L1 and L2 speech: The short vowels of Afrikaans and South-African English", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 20: 65-80.

Clement, C.J. (2004), *Development of vocalizations in deaf and normally hearing infants*, Ph.D. thesis Univ. of Amsterdam, LOT 100: 264 pp.

Dijk, J. van (2001), *Mechanical aspects of hearing*, Ph.D. thesis Univ. of Amsterdam: 211 pp.

Dijkstra, J., Pols, L.C.W. & Son, R.J.J.H. van (2004), "Frisian TTS, an example of bootstrapping TTS for minority languages", *Proc. 5th ISCA Synthesis Workshop*, Pittsburgh: 97-102.

Donzel, M. E. van (1999), *Prosodic aspects of information structure in discourse*, Ph.D. thesis Univ. of Amsterdam: 194 pp.

Elgendy, A. M. (2001), *Aspects of pharyngeal coarticulation*, Ph.D. thesis Univ. of Amsterdam: 312 pp.

Escudero, P. & Boersma, P. (2004), "Bridging the gap between L2 speech perception research and phonological theory", *Studies in Second Language Acquisition* 26(4): 551-585.

Gelderen, A. J. S. van (1992), *De evaluatie van spreekvaardigheid in communicatieve situaties. Globale beoordeling en gedetailleerde analyse van spreekprestaties van 11- en 12-jarigen*, Ph.D. thesis Univ. of Amsterdam: 265 pp.

Hilgers, F. (2004), "Van verbazing naar vanzelfsprekendheid. From astonishing to predictable speech", Oratie Univ. van Amsterdam, Vossiuspers UvA, Amsterdam: 64 pp.

Jacobi, I., Pols, L.C.W. & Stroop, J. (2005), "Polder Dutch: Aspects of the /Ei/-lowering in standard Dutch", *Proc. Interspeech 2005*, Lisbon: 2877-2880.
Jansonius-Schultheiss, K. (1999), *Twee jaar spraak en taal bij schisis*, Ph.D. thesis Univ of Amsterdam: 277 pp.
Jongmans, P., Hilgers, F.J.M., Pols, L.C.W. & Van As-Brooks, C.J. (2005), "The intelligibility of tracheoesophegeal speech: First results", *Proc. Interspeech 2005*, Lisbon: 1749-1752.
Kemps, R.J.J.K. (2004), *Morphology in auditory lexical processing. Sensitivity to fine phonetic detail and insensitivity to suffix reduction*, Ph.D. thesis Univ. of Nijmegen: 159 pp.
Klaassen-Don, L. E. O. (1983), *The influence of vowels on the perception of consonants*, Ph.D. thesis Univ. of Leiden: 153 pp.
Kovačić, G., Boersma, P. & Domitrović, H. (2003), "Long-term average spectra in professional folk singing voices: A comparison of the *klapa* and *dozivački* styles", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 25: 53-64.
Kuijpers, C. T. L. (1993), *Temporal coordination in speech development. A study on voicing contrast and assimilation*, Ph.D. thesis Univ. of Amsterdam: 165 pp.
Lindblom, B. & Studdert-Kennedy, M. (1967), "On the role of formant transitions in vowel recognition", *J. Acoust. Soc. Amer.* 42: 830-843.
Mozziconacci, S. (1998), *Speech variability and emotion: Production and perception*, Ph.D. thesis Technological Univ. of Eindhoven: 210 pp.
Nierop, D.J.P.J. van, Pols, L.C.W. & Plomp, R. (1973), "Frequency analysis of Dutch vowels from 25 female speakers", *Acustica* 29(2): 110-118.
Odé, C. (2003), "Description and transcription of Russian Intonation (ToRI)", In: J. Schaeken, P. Houtzagers & J. Kalsbeek (Eds.), *Dutch contributions to the thirteenth international congress of Slavists*, Studies in Slavic and General Linguistics 30, Rodopi, Amsterdam: 279-288.
Odé, C. & Van Son, R.J.J.H. (2003), "Note from the Guest Editors", *Speech Communication* 47(1-2): 1-2.
Ölander, A., Perridon, H. & Pols, L.C.W. (1999), "Length acquisition of stressed syllables in Swedish by Dutch L2 learners", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 23: 67-76.
Oostdijk, N., Goedertier, W., Eynde, F. van, Boves, L., Martens, J.P., Moortgat, M. & Baayen, H. (2002), "Experiences from the Spoken Dutch Corpus project", *Proc. 3rd LREC*: 340-347.
Perkell, J.S. & Klatt, D.H. (Eds.) (1986), *Invariance and variability in speech processes*, LEA, Hillsdale NJ: 604 pp.
Plomp, R. (2002), *The intelligent ear. On the nature of sound perception*, LEA, Mahwah, NJ: 174 pp.
Plomp, R., Pols, L.C.W. & Geer, J.P. van der (1967), "Dimensional analysis of vowel spectra", *J. Acoust. Soc. Amer.* 41: 707-712.
Pols, L.C.W. (1977), *Spectral analysis and identification of Dutch vowels in monosyllabic words*, Ph.D. Thesis Free Univ. of Amsterdam: 152 pp.
Pols, L.C.W. (1979), "Real-time recognition of spoken words", *IEEE Trans. Comp.* C20: 972-977.
Pols, L.C.W. (1983), "Three-mode principal components analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation", *Speech Communication* 2(4): 275-293.
Pols, L.C.W. (1986), "Variation and interaction in speech", In: J.S. Perkell & D.H. Klatt (Eds.), *Invariance and variability in speech processes*, Lawrence Erlbaum Ass., Publishers, Hillsdale, N.J.: 140-154.
Pols, L.C.W. (1991), "Quality assessment of text-to-speech synthesis-by-rule", In: S. Furui & M.M. Sondhi (Eds.), *Advances in Speech Signal Processing*, Marcel Dekker Inc., Chapter 13: 387-416.
Pols, L.C.W. (1994), "Speech technology systems: Performance and evaluation", In: R.E. Asher (Ed.), *Encyclopedia of Language & Linguistics*, Pergamon Press, Oxford: 4289-4296.
Pols, L.C.W. (2000), "Even better synthesis", In: W.F. Sendlmeier (Ed.), *Speech and signals - Aspects of speech synthesis and automatic speech recognition*, Hector, Frankfurt am Main: 15-25.
Pols, L.C.W. (2004a), "Current developments in phonetics", In: F. Slifka, S. Manuel and M. Matthies (Eds.), *Proc. 'From Sound to Sense: 50+ Years of Discoveries in Speech Communication'*, MIT Cambridge: B52-B61.
Pols, L.C.W. (2004b), "Expanding Phonetics", In: H. Quené and V.J. van Heuven (Eds.), *On speech and language: Studies for Sieb G. Nooteboom*, LOT Occasional Series 2: 141-148.
Pols, L.C.W., Kamp, L.J.Th. van der & Plomp, R. (1969), "Perceptual and physical space of vowel sounds", *J. Acoust. Soc. Amer.* 46: 458-467.
Pols, L.C.W. & Olive, J.P. (1983), "Intelligibility of consonants in CVC utterances produced by dyadic rule synthesis", *Speech Communication* 2(1): 3-13.
Pols, L.C.W. & SAM-partners (1992), "Multi-lingual synthesis evaluation methods", *Proc. ICSLP'92*, Banff, Canada, Vol. 1: 181-184.

Pols, L.C.W. & Schouten, M.E.H. (1987), "Perception of tone, band, and formant sweeps", In: M.E.H. Schouten (Ed.), *The psychophysics of speech perception*, Martinus Nijhoff Publishers, Dordrecht: 231-240.
Pols, L.C.W. & Son, R.J.J.H. van (1993), "Acoustics and perception of dynamic vowel segments", *Speech Communication* 13: 135-147.
Pols, L.C.W., Tromp. H.R.C. & Plomp, R. (1973), "Frequency analysis of Dutch vowels from 50 male speakers", *J. Acoust. Soc. Amer.* 53(4): 1093-1101.
Pols, L.C.W., Wang, X. & Bosch, L.F.M. ten (1996), "Modelling of phone duration (using the TIMIT database) and its potential benefit for ASR", *Speech Communication* 19: 161-176.
Pols, L.C.W. & Weenink, D.J.M. (2205), "Vowel recognition and (adaptive) speaker normalization", *Proc. SPECOM 2005*, Patras, Vol. 1: 17-24.
Rechziegel, A. (2001), "Consonants in contact: On assimilation and cross-language contrast", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 24: 103-115.
Reetz, H. (1996), *Pitch perception in speech: A time domain approach. Implementation and evaluation*, Ph.D. thesis Univ. of Amsterdam: 236 pp.
Santen, J.P.H. van, Pols, L.C.W., Abe, M., Kahn, D., Keller, E. & Vonwiller, J. (1998), "Report on the Third ESCA TTS Workshop Evaluation Procedure", *Proc. Third ESCA TTS Workshop*, Jenolan Caves, Australia: 329-332.
Schouten, M.E.H. & Pols, L.C.W. (1983), "Perception of plosive consonants. The relative contributions of bursts and vocalic transitions", In: M.P.R. van den Broecke, V.J. van Heuven & W. Zonneveld (Eds.), *Sound Structures: Studies for Antonie Cohen*, Foris Publications, Dordrecht: 227-243.
Schwippert, C. (2000), "Perception of a b-d contrast by adult dyslexic and control subjects", *IFA internal report*.
Son, R. J. J. H. van (1993), *Spectro-temporal features of vowel segments*, Ph.D. thesis Univ. of Amsterdam: 195 pp.
Son, R. J. J. H. van, Binnenpoorte, D., van den Heuvel, H. & Pols, L.C.W. (2001), "The IFA corpus: a phonemically segmented Dutch 'open source' speech database", *Proc. Eurospeech'01*, Aalborg, Denmark, Vol. 3: 2051-2054.
Son, R.J.J.H. van, Bolotova, O., Lennes, M. & Pols, L.C.W. (2004), "Frequency effects on vowel reduction in three typologically different languages (Dutch, Finnish, Russian)", *Proc. Interspeech 2004 - ICSLP*, Jeju Island, Korea, Vol. II: 1277-1280.
Son, R. J. J. H. van & Pols, L. C. W. (1992), "Formant movements of Dutch vowels in a text, read at normal and fast rate", *J. Acoust. Soc. Amer.* 92(1): 121-127
Son, R. J. J. H. van & Pols, L. C. W. (1997), "The correlation between consonant identification and the amount of acoustic consonant reduction", *Proc. Eurospeech'97*, Rhodes, Vol. 4: 2135-2138.
Son, R. J. J. H. van & Pols, L. C. W. (1999a), "An acoustic description of consonant reduction", *Speech Communication* 28(2): 125-140.
Son, R. J. J. H. van & Pols, L. C. W. (1999b), "Perisegmental speech improves consonant and vowel identification", *Speech Communication* 29(1): 1-22.
Son, R.J.J.H. van & Pols, L.C.W. (2001), "Structure and access of the open source IFA-corpus", *Proc. IRCS Workshop on Linguistic Databases*, S. Bird, P. Buneman & M. Liberman (Eds.): 245-253.
Son, R. J. J. H. van & Pols, L. C. W. (2003), "Information structure and efficiency in speech production", *Proc. Eurospeech'03*, Geneva, Vol. 1: 769-772.
Steeneken, H. J. M. (1992), *On measuring and predicting speech intelligibility*, Ph.D. thesis Univ. of Amsterdam: 165 pp.
Stelt, J. M. van der (1993), *Finally a word: A sensori-motor approach of the mother-infant system in its development towards speech*, Ph.D. thesis Univ. of Amsterdam: 226 pp.
Stelt, J.M. van der, Wempe, A.G. & Pols, L.C.W. (2003), "Progression in infants' vowel space: An analysis of deaf and hearing infants' sounds", *Proc. 15th ICPhS*, Barcelona, Vol. 3: 2225-2228.
Streefkerk, B.M. (2002), *Prominence. Acoustic and lexical/syntactic correlates*, Ph.D. thesis Univ. of Amsterdam: 168 pp.
Streefkerk, B. M., Pols, L. C. W. & ten Bosch, L.F.M. (2001), "Up to what level can acoustical and textual features predict prominence", *Proc. Eurospeech '01*, Aalborg, Denmark, Vol. 2: 811-814.
Tielen, M. T. J. (1992), *Male and female speech. An experimental study of sex-related voice and pronunciation characteristics*, Ph.D. thesis Univ. of Amsterdam: 180 pp.
Verdonck-de Leeuw, I. M. (1998), *Voice characteristics following radiotherapy: the development of a protocol*, Ph.D. thesis Univ. of Amsterdam: 137 pp.
Wang, X. (1997), *Incorporating knowledge on segmental duration in HMM-based continuous speech recognition*, Ph.D. thesis Univ. of Amsterdam: 190 pp.

Wempe, A.G. & Boersma, P. (2003), "The interactive design of an F_0-related spectral analyser", *Proc. 15th ICPhS*, Barcelona, Vol. 1: 343-346.
Wesseling, W. & Son, R.J.J.H. van (2005), "Timing of experimentally elicited minimal responses as quantitative evidence for the use of intonation in projecting TRP's", *Proc. Interspeech 2005*, Lisbon: 3398-3392.
Wieringen, A. van (1995), *Perceiving dynamic speechlike sounds. Psycho-acoustics and speech perception*, Ph.D. thesis Univ. of Amsterdam: 256 pp.
Wieringen, A. van & Pols, L.C.W. (1995), "Discrimination of single and complex consonant-vowel- and vowel-consonant-like formant transitions", *J. Acoust. Soc. Amer.* 93(3): 1304-1312.
Zajdó, K., Van der Stelt, J.M., Wempe, T.G. & Pols, L.C.W. (2005), "Cross-linguistic comparison of two-year old children's acoustic vowel spaces: Contrasting Hungarian with Dutch", *Proc. Interspeech 2005*, Lisbon: 1173-1176.