

PROFILE OF THE INSTITUTE OF PHONETIC SCIENCES UNIVERSITY OF AMSTERDAM

HISTORY

The Institute of Phonetic Sciences (in Dutch: Instituut voor Fonetische Wetenschappen, Amsterdam - IFA) belongs to the Faculty of Arts of the University of Amsterdam, the former municipal university. A short historical introduction is given below:

In 1922 the first courses in practical phonetics at the University of Amsterdam were given by Louise Kaiser, albeit in the physiological laboratory for the time being. In 1926 a chair in the Faculty of Arts was instituted and Louise Kaiser was appointed as a reader in Phonetics. The Physiological Laboratory, at the Jodenbreestraat, accommodated phonetics for several years, but in 1932 room was found for it in the building next to it. The new Laboratory for Experimental Phonetics was officially opened on November 30th 1933, some time after the First International Congress for Phonetic Sciences in 1932. It was thanks to the endeavour of Louise Kaiser that this congress took place in Amsterdam. On September 15th 1958, Louise Kaiser said farewell to the University of Amsterdam. In 1959 Hendrik Mol was appointed professor at this university while holding at the same time a chair at Leyden University. In 1967 he was appointed as a full professor in Amsterdam. Soon it became clear that the Phonetic Laboratory had to move from the Jodenbreestraat 72; one of the reasons being the construction of a road as an entrance to the tunnel under the river Y.

In 1963 the University bought a house on one of the famous canals in the center of Amsterdam (Herengracht 338). The building, dating from the end of the 17th century, was completely reconstructed. The conservation of the front and of the facade at the rear, as well as large parts of the interior, took much time. The interior of the building was brought up to date for experimental phonetic research including a small anechoic room. In the middle of 1966 the new building was delivered up and in 1967 the laboratory was officially opened by the Lord Mayor of Amsterdam, president of the board of governors of Amsterdam University. On that occasion the name of 'Phonetic Laboratory' was changed in 'Institute of Phonetic Sciences'. After the death of Prof. Mol in 1980, the chair in Phonetic Sciences is occupied by Louis Pols since October 1982.

TEACHING

Although it has been possible to choose phonetics as a major discipline since 1968 (as an extrastatutory course), phonetics was actually recognized as an autonomous discipline in the so-called Dutch Academic Statutes only since 1983, the year of the Tenth International Congress of Phonetic Sciences. Phonetics as a major study can only be chosen after one propaedeutic year in another discipline. This discipline can be a language, mathematics, a science, engineering, psychology or medicine; it reflects the multi-disciplinary character of the subject. And this leaves three more years for phonetic specialization in combination with a study in one of the related major studies such as General Linguistics or Alpha-Informatics. Recently a modular teaching system was adapted by the University of Amsterdam. Each of the three years consists of three

blocks of 14 course points. Most courses are seven points each and require therefore in principal work during seven weeks by the students. The first year mainly contains introductory courses, such as an introduction in phonetic sciences, an introduction in research methods and techniques, plus programming, mathematics and statistics. In the final two years one can choose between a specialization in 'speech development' or in 'speech communication and technology'. Upon conclusion of the study a 21-points Master's thesis has to be written, the research topic of which is frequently related to on-going research in the Institute.

Recently it has also become possible to study phonetics in the two above mentioned specialities, as a two-year program on top of a four-year (non-university) course for speech therapists or for electro-technical engineers.

Furthermore, a Ph.D. can be obtained upon completion of a doctoral dissertation, which is either a monograph, or a collection of related published articles.

The teaching program of our Institute is also intended to be incorporated in due time in the exchange possibilities within the scope of the ERASMUS-program "Phonetics and Speech Communication".

RESEARCH

The research in our Institute is concentrated in four research areas:

- (1) Speech analysis, perception and technology
- (2) Speech communication and computer science
- (3) Speech communication: development and pathology
- (4) Female voices studied phonetically

A complete description of the research projects is given in the section "Summary of research". Almost all research projects, incorporated in this summary, belong to a so-called faculty research area of major interest (FAG=Facultair Aandachtsgebied). This is in line with the present policy of government and university to concentrate (inter)-faculty research in a limited number of protected research VF-programs (VF="voorwaardelijk gefinancierd"). The Institute of Phonetic Sciences is, by means of the projects of research area (1), prominently represented in the FAG/VF-program "Perceptual and physical description of variation and invariance in speech". We also participate with the projects of research area (3) substantially in the FAG/VF-program "Language and speech: development, pathology and other forms of communication". These two FAG/VF-programs are incorporated in the Research Institute for Functional Studies of Language and Language Use (IFOTT).

Beyond that we also participate in the FAG/VF-program "Alpha-Informatics: grammars and data structures in language, speech and music", and in the FAG/VF-program "Women's studies in the Faculty of Arts", with one project each.

The research projects are either projects of permanent staff members (partly doctoral dissertations), or projects funded by the University of Amsterdam and executed by temporary junior research assistants, or projects externally funded and executed by temporary co-workers. Presently we receive grants from the following national and international organizations:

- Faculty Ph.D. program for researchers in training (AIO's)
- NWO, Netherlands Organization for Scientific Research
- SPIN-ASSP, Dutch Governmental Speech Program for Promoting Information Technology research
- Netherlands Prevention Fund
- Institute for the Deaf, Sint-Michielsgestel
- ESPRIT, European Strategic Program for Research and Development of Information Technology

These research grants permit us to establish and maintain scientific contacts with many national and international research laboratories, both in university as well as in industry. One of the outcomes was the organization of the ESCA "Tutorial Day and Workshop on Speech Input/Output Assessment and Speech Databases", held in Noordwijkerhout, The Netherlands, September 1989.

FACILITIES

The Institute is in the position to have a rather large, public scientific library in the interdisciplinary field of Phonetic Sciences. The library contains more than 3000 bands and many of them are of great historical value. We also have a subscription to about 45 journals and we receive many different progress and institute reports from all over the world, mainly on exchange basis. The Proceedings of the Institute of Phonetic Sciences Amsterdam is widely distributed; Proceedings nr. 14 became available by the end of 1990. Apart from papers in the open literature, we also produce internal publications in the form of Institute Reports. IFA Report nr. 107 recently got finished.

With respect to the technical facilities, we have the disposal of a small anechoic room (4.1 x 4.8 x 2.9m) hidden in the multi-stored canal house, as well as a mechanical and an electronical workshop. Furthermore, there are the common recording facilities including several reel recorders, video recorders for high quality registrations (very usefull in listening experiments), compact disk players, standard and portable cassette recorders, and DAT recorders. Since the middle of 1986 we have a DEC MicroVAX II computer under VMS, which has been extended recently by two VAXstations 3100, creating a multi-user Local Area VAX Cluster. PC-based stand-alone systems, such as a Compaq 286 and several others, are available for all kind of applications, e.g. stimulus generation, speech synthesis, and controlling listening experiments. Adequate facilities are also available for word processing, statistical analyses, graphical output and database collection, mainly on PC-base.

FACTS

In the following pages a survey is given of ongoing, and recently finished, research projects. Each projects is presented by a short description in which the aim of the research, some results, and future plans are indicated. Apart from the name of the principal investigator(s), some additional information is added, such as the project's internal code number, the funding organization, and the publications that have resulted from the project or are directly related to it.

SURVEY OF RESEARCH PROJECTS

AREA 1: Speech analysis, perception and technology

LET/FW008/10/1 **The processing of acoustic signals**

Drs. D.J.M. Weenink, Ir. D.G. Broeder, Drs. R. van Son

At the Institute of Phonetic Sciences in Amsterdam the analysis of acoustic signals is mainly performed on a multi-user Local Area VAX cluster consisting of a microVAX II and two VAXstations 3100. As secondary storage we have a total of 1.2 GB disk capacity. 12 bit A/D and D/A conversion is being done on a client-server basis, for which the microVAX II, because of the special hardware available for its Q-bus, functions as the server. The basic analysis software tool available is the LVS speech analysis system, a set of programs for analysis, (re)synthesis and modification of speech files developed at IPO Eindhoven (Vogten, 1985). Apart from this system, a number of programs have been developed at our institute by colleagues under SPIN-ASSP contract: a very powerful speech editor SESAM (Buiting, 1981; Broeder, 1990) and a very flexible program FORM, capable of editing and displaying speech parameters (van Son, 1991). Apart from speech processing capabilities on the VAX cluster, some smaller PC-based systems are in use. Amongst them there is a Compaq 286 with 16 bit A/D conversion and simple speech editing facilities. This is used as a stand alone stimulus generation machine for performing listening experiments, while special designed hardware can function as a real time filter bank. In the near future we plan to increase disk storage capacity as well as calculating power capacity in the cluster configuration. New programs will be written in the Xwindow environment.

- Buiting, H.J.A.G. (1981), 'SESAM Speech Editing System Amsterdam'. IFA-report 70, 15 pp. (in Dutch).
- Vogten, L.L.M. (1985), 'LVS-Speech processing programs on IPO-VAX 11/780'. Handleiding 67, Institute for Perception Research, Eindhoven, 107 pp.
- Weenink, D.J.M. (1986) 'QQ, a program for analysis, resynthesis, and recognition of vowel segments.' IFA-report 82, 23 pp. (in Dutch).
- Weenink, D.J.M. and Wempe, A.G. (1986). 'Communication between an Apple Iie and four Commodore Vic 20's'. IFA-report 83, 32 pp. (in Dutch).
- Horst, B. (1987). 'Manual for the creation of sampled data files on the VAX computer'. IFA-report 91, 13 pp. (in Dutch).
- Broeder, D.G. (1990). 'Analysis and synthesis of speech; CADDa-manual'. SPIN-ASSP report 23, 48 pp. (in Dutch).
- Weenink, D.J.M. (1988). 'Vowels: A computer program for generating vowel-like stimuli'. IFA-report 100, 32 pp. (in Dutch).
- Broeder, D.G. (1990). 'Analysis and synthesis of speech; SESAM manual'. SPIN-ASSP report 24, 82 pp. (in Dutch).
- Weenink, D.J.M. (1989). 'The signal processing computer facilities in the Faculty of Arts'. In: Mededelingen vakgroep Alfa-informatica 7, 15-23. (in Dutch).
- Bergem, D.R. van (1990). 'Pitch period estimation by filtering the fundamental frequency out of the speech waveform'. Proceedings of the Institute of Phonetic Sciences Amsterdam 14, 17-26.
- Son, R. van (1991). 'FORM: a program for the manipulation of analysis files'. In preparation.

LET/FW019/10/1 **Normalization of speaker variation**
drs. D.J.M. Weenink
dr. F.J. Koopmans-van Beinum, prof. dr. ir. L.C.W. Pols
NWO Project 300-161-026, 1982-1987

Although the physical differences between linguistically identical utterances produced by talkers with different vocal tract lengths are great, listeners do not seem to have great difficulty in perceiving these utterances as the same. Apparently the listener is able to extract the relevant features from the speech signal. Little is known about the perceptual strategies used to normalize for different talkers. The main topic of this investigation is to study the ability of the listener to adapt to different speakers. Listening experiments were performed in which the characteristics of short vowel-like stimuli were systematically manipulated in terms of fundamental frequency and spectral contents. Results of the listening experiments showed that, first, listeners are very well equipped to correctly identify vowel-like stimuli from different speakers (men, women, children) when presented to them in a completely randomized order, i.e. vowel *i* from speaker *m* could be followed by vowel *j* from speaker *n*. Secondly, identification results improved somewhat when listeners were presented with stimuli from the same speaker, i.e. vowel *i* from speaker *m* was followed by vowel *j* from speaker *m*. Concerning the normalization process it is probable that listeners use fundamental frequency information to roughly categorize speaker category.

In the near future we plan to complete this investigation with a Ph.D.-thesis.

- Dijk, J.S.C. van (1984). 'Conservation of vowel contrast in various speech conditions'. Proceedings of the Institute of Phonetic Sciences Amsterdam 8, 19-31.
- Weenink, D.J.M. (1984). 'Literature overview on perceptual and physical normalisation of speaker variation'. Proceedings of the Institute of Phonetic Sciences Amsterdam 8, 55-64.
- Pols, L.C.W. (1985). 'Similarity between schwa-like stimuli manipulated in terms of male, female, and child characteristics'. Proceedings of the Institute of Phonetic Sciences 9, 53-65.
- Weenink, D.J.M. (1985). 'Formant analysis of Dutch vowels from 10 children'. Proceedings of the Institute of Phonetic Sciences Amsterdam 9, 45-52.
- Weenink, D.J.M. (1986). 'The identification of vowel stimuli from men, women and children'. Proceedings of the Institute of Phonetic Sciences Amsterdam 10, 41-54.
- Weenink, D.J.M. (1987). 'Robust linear prediction analysis of vowels'. Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 79-88.
- Bergem, D.R. van, Pols, L.C.W. & Koopmans-van Beinum, F.J. (1988). 'Perceptual normalization of the vowels of a man and a child in various contexts'. Speech Communication 7, 1-20.

LET/FW/025/10/1 **A model for describing the structure of, and the dynamics within, vowel systems**

drs. L.F.M. ten Bosch

prof. dr. ir. L.C.W. Pols, drs. L.J. Bonder

NWO project 300-161-030, 1985-1988

In this project, a model has been developed for the description of the structure of vowel systems (VSS). A model aiming at such a structure description is denoted by a VSS-model. Vowel systems in natural languages show certain regularities that can be translated in so-called phonological 'universals'. The central question in this project is, whether these regularities on the phonological domain can be described by physical constraints on the phonetic domain. In the past decades, one has considered two constraint types, one based on speech production, the other on speech perception. The production constraint is formulated as an articulatory 'principle of minimal effort' which states that vowels are produced with least effort. The perceptual constraint deals with the 'principle of sufficient contrast' which states that vowels must preserve sufficient spectral distance in order to be perceived without confusion. Optimal vowel systems are those systems that optimally fulfill both principles simultaneously.

In this project we show that both principles account for the phonetic structure of vowel systems to a considerable extent, provided that the definitions of effort and contrast are carefully chosen. Finally, we investigated the optimal combination of both principles, as well as the implication of the results for the VSS-models in the literature.

Bonder, L.J. (1986). 'A prediction method for modal n-vowel systems'. Proceedings of the Institute of Phonetic Sciences Amsterdam 10, 73-90.

Bosch, L.F.M. ten (1987). 'About diphthongs. An implementation into vowel dispersion theory'. Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 1-14.

Bosch, L.F.M. ten, Bonder, L.J. & Pols, L.C.W. (1987). 'Static and dynamic structure of vowel systems'. Proceedings of the Eleventh International Congress of Phonetic Sciences, Tallinn, Vol. 1, 235-238.

Bosch, L.F.M. ten, Bonder, L.J. & Pols, L.C.W. (1987). 'A model for the description of the structure of, and dynamics inside, vowel systems'. Final report ZWO-project, 13 pp. (in Dutch).

Bosch, L.F.M. ten & Pols, L.C.W. (1989). 'On the necessity of quantal assumptions. Questions to the quantal theory'. Journal of Phonetics 17, 63-70.

LET/FW/027/10/1 **The importance of local context on the correct interpretation of acoustic-phonetic information in speech understanding**

drs. L.W.M. Damen

prof. dr. ir. L.C.W. Pols

NWO-project 300-161-032, 1983-1988

The aim of the project was to determine the contribution of optimal acoustic-phonetic information to speech processing. The main hypothesis of the research was that the correct interpretation of a speech segment will be influenced by the long-term and short-term physical characteristics of the adjacent speech signals. We opted for a gating task with words. In several listening experiments the intelligibility of the vowels (short and long vowels) and the consonants were measured in four conditions:

- a) words presented in isolation
- b) words presented in neutral carrier sentences
- c) words in contextual carrier sentences
- d) moreover, the contextual information was varied by using a slow speaking rate as well as a fast speaking rate.

The results were to draw several conclusions:

- The perceptual distinction between two vowels in short - long opposition is not only depending on durational differences but also on spectral differences.
- Speaking rate of the acoustic-phonetic context, i.e. neutral and nonsense carrier sentences, has no influence upon perception of vowels in short-long opposition.
- The intelligibility of a word or speech segment is negatively influenced by an isolated presentation.
- Increase of speaking rate in the carrier sentences has a negative effect upon a correct identification of the consonants.

Damen, L.W.M. (1987). 'Perception of the long-short vowel oppositions in Dutch: a gating study'. Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 35-46.

Damen, L.W.M. & Pols, L.C.W. (1987). 'The use of the gating paradigm for studying the long-short vowel opposition'. Proceedings of the Eleventh International Congress of Phonetic Sciences, Tallinn, Vol. 4, 116-119.

LET/FW/031/10/1 **Perceptual evaluation and description of features of voice and pronunciation**

drs. L.W.A. van Herpt

prof. dr. ir. L.C.W. Pols, dr. F.J.Koopmans-van Beinum

The project is conducted in two phases:

a) Construction and validation of a maximally reliable and at the same time efficient standard procedure for the description of voice and pronunciation (V&P) on an auditory basis by human listeners. Long-term speaker characterizing features of V&P have been identified using Osgood's semantic differential technique, and have been operationally realized in an instrument consisting of 14 (to 35) bipolar adjective scales typically describing five dimensions that might be called 'melodiousness', 'voice quality', 'articulation quality', 'pitch', and 'tempo'. This instrument is used in a variety of projects concerning normal and pathological human speech.

b) The relation between above mentioned speaker-specific features of V&P and, biological, regional and social speaker characteristics is investigated. The discrepancy between the attributions of listeners based on this verbal extra- and paralinguistic information and objective actual data, shows the importance of the listener's expectations, interpretations and biases as part of the measurement procedure. Experiments focused on listener/speaker interaction effects in relation to sex and age have been conducted.

Herpt, L.W.A. van (1986). 'Influence of rater's sex on voice and pronunciation assessment'. Proceedings of the Institute of Phonetic Sciences 10, 19-40.

Herpt, L.W.A. van & Hoebe, A.P. (1985). 'Attribution of age from perceived speech'. Proceedings of the Institute of Phonetic Sciences 9, 1-23.

Herpt, L.W.A. van (1987). 'Perception of paralinguistic cues of age and sex in manipulated speech: an exploration'. Proceedings of the Eleventh International Congress of Phonetic Sciences, Tallinn, Vol 1, 187-190.

Herpt, L.W.A. van (1987). 'Do men and women use a common semantic space to describe voice and pronunciation?'. Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 15-26.

LET/FW/033/10/1 **Speech quality evaluation of Dutch text-to-speech synthesis by rule**

dr. R.A.M.G. van Bezooijen
prof. dr. ir. L.C.W. Pols
SPIN-ASSP project 1986-1991

This is one of the many projects in a Dutch national cooperative research program on 'Analysis and Synthesis of Speech', funded by SPIN-ASSP. The various components, as well as the ongoing development of this advanced text-to-speech system for the Dutch language, are evaluated in a diagnostic and comparative way. The intelligibility at phoneme, word and sentence level has been measured. The overall speech quality is judged, whereas also separate components, such as the sentence accent algorithm, are tested. Results showed substantial progress for the allophone-based system, up to a level where it is comparable in performance to the diphone-based system, in as far as the phoneme intelligibility is concerned. Subsequent tests are planned at the so-called text level, where the overall quality of sentences generated from text are judged by listeners.

- Bezooijen, R. van (1987). 'Evaluation of two synthesis-by-rule systems for Dutch: Segmental intelligibility'. SPIN-ASSP report 1, 101 pp.
- Bezooijen, R. van (1987). 'Evaluation of two synthesis systems for Dutch: phonemes and consonant clusters'. Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 47-57.
- Bezooijen, R. van & Pols, L.C.W. (1987). 'Evaluation of two synthesis-by-rule systems for Dutch', Proceedings of the European Conference on Speech Technology, Edinburgh, Vol. 1, 183-186.
- Bezooijen, R. van (1988). 'Evaluation of two synthesis systems for Dutch: Intelligibility and overall quality of initial and final consonant clusters'. SPIN-ASSP report 3, 88 pp.
- Bezooijen, R. van (1988). 'Evaluation of the quality of consonant clusters in two synthesis systems for Dutch'. Proceedings of Speech '88, 7th FASE Symposium, Edinburgh, Book 2, 445-452.
- Bezooijen, R. van (1988). 'Evaluation of two synthesis systems for Dutch: Development and application of intelligibility tests'. SPIN-ASSP report 5, 78 pp.
- Bezooijen, R. van (1988). 'The relative importance of pronunciation prosody, and voice quality for the attribution of social status and personality characteristics'. In: R. van Hout & U. Knops (Eds.), Language attitudes in the Dutch Language area. Foris Publications, Dordrecht, 85-102.
- Pols, L.C.W. (1988). 'Improving synthetic speech quality by systematic evaluation'. Proceedings of the Institute of Phonetic Sciences Amsterdam 12, 19-27, and in: Proceedings of The Second Symposium on Advanced Man-Machine Interface Through Spoken Language, Nov. 1988, Makaha, Oahu, Hawaii, 17/1-17/9.
- Bezooijen, R. van (1989). 'Evaluation of an algorithm for the automatic assignment of sentence accents'. In: Bennis, H. & Kemenade, A. van (Eds.), Linguistics in the Netherlands, Dordrecht/Providence: Foris Publications, 6, 21-30.
- Bezooijen, R. van (1989). 'Evaluation of the suitability of Dutch text-to-speech conversion for application in a digital daily newspaper'. In: Pols, L.C.W., Heugten, L.J.P. van, Koopmans-van Beinum, F.J., Bezooijen, R. van & Steeneken, H.J.M. (Eds.), Proceedings of the ESCA Workshop on Speech Input/Output Assessment and Speech Databases, Noordwijkerhout, The Netherlands, 6.3.1-6.3.4.
- Bezooijen, R. van (1989). 'Evaluation of an algorithm for the automatic assignment of sentence accents in written text'. SPIN-ASSP report 9, 59 pp.
- Bezooijen, R. van (1989). 'Evaluation of text-to-speech conversion at the text level'. SPIN-ASSP report 15, 53 pp.
- Bezooijen, R. van (1989). 'Speech synthesis in The Netherlands: a snapshot'. Gramma 13, 97-112. (in Dutch).
- Bezooijen, R. van & Pols, L.C.W. (1989). 'Evaluation of text-to-speech conversion for Dutch: From segment to text'. In: Pols, L.C.W., Heugten, L.J.P. van, Koopmans-van Beinum, F.J., Bezooijen, R. van & Steeneken, H.J.M. (Eds.), Proceedings of the ESCA Workshop on Speech Input/Output Assessment and Speech Databases, Noordwijkerhout, The Netherlands, 3.4.1-3.4.4.

- Bezooijen R. van & Pols, L.C.W. (1989). 'Evaluation of a sentence accentuation algorithm for a Dutch text-to-speech system'. In: Tubach, J.P. & Mariani, J.J. (Eds.), Proceedings EUROSPEECH'89, Paris, CEP Consultants Ltd, Edinburgh, Vol. 1, 218-221.
- Bezooijen, R. van (1990). 'Evaluation of speech synthesis for Dutch: Comparison of synthesis systems, intelligibility tests, and scaling methods'. SPIN-ASSP report 22, 107 pp.
- Bezooijen, R. van & Pols, L.C.W. (1990). 'Evaluating text-to-speech systems: Some methodological aspects'. Speech Communication 9 (4), 263-270.

LET/FW/034/10/1 **Acoustic-phonetic aspects of automatic speech recognition**

drs. P. van Alphen
 prof. dr. ir. L.C.W. Pols, drs. J.S.C. van Dijk
 AIO project 1986-1990

The goal of this project is to specify and design an automatic speech recognition system, and to implement this system on a microVAX, with a digital signal processor as acoustic front-end. The analysis in the front-end consists of a 15-band filterbank, where some effort is made to simulate ear-characteristics (auditory modelling). The front-end yields spectral vectors, which are used as input for the actual recognition system. This recognition system is based on Hidden Markov Modelling (HMM). Being interested in a general method to recognize continuous speech, we constructed HMM's for phonemes. If we want to recognize words, these words have to be constructed by concatenation of the phoneme-models. It is also possible to include explicit word duration and grammar knowledge in the recognition algorithm. In this way it is possible to reduce the search-space, and enable accurate word and/or sentence recognition. At this moment we have build a system that is capable of recognizing speech from one male speaker. The spoken sentences have a vocabulary of 240 words, but due to the grammar, the number of words to be recognized was 110, 60 and 20 (=perplexity). The recognition results for these perplexities are 81.7%, 89.1% and 96.6% respectively (word accuracy). The only thing that remains to be done in this project is to finish a Ph.D.-thesis.

- Alphen, van P., Horst, B. & Pols, L.C.W. (1988). 'A fast algorithm for a FIR-based filterbank, designed for the acoustic front-end of a speech recognizer'. Proceedings of Speech '88, 7th FASE Symposium, Edinburgh, Book 2, 677-682.
- Horst, B. (1988). 'The design of a filterbank using changes in sampling frequency for the purpose of energy estimates of band-limited signals'. IFA-report 98, 38 pp. (in Dutch).
- Alphen, P. van & Bergem, D.R. (1989). 'Markov models and their application in speech recognition'. Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 1-26.
- Alphen, P. van & Pols, L.C.W. (1989). 'A real-time FIR-based filterbank, as the acoustic front end of a speech recognizer'. In: Tubach, J.P. & Mariani, J.J. (Eds.), Proceedings EUROSPEECH'89, Paris, CEP Consultants Ltd, Edinburgh, Vol. 1, 621-624.
- Alphen, P. van (1990). 'Phone recognition in continuous speech'. J. Acoust. Soc. Am. Suppl. 1, Volume 87, Spring 1990, S49 (A).
- Alphen, P. van (1990). 'Hidden Markov Models and their application in speech recognition'. In: Reader Colloquium Signaalanalyse en Spraak, IPO report 765, contribution 10, 6 pp.

LET/FW/035/10/1 **Spectro-temporal analysis and perception of vowel formant contours**

drs. R.J.J.H. van Son

prof. dr. ir. L.C.W. Pols, dr. F.J. Koopmans-van Beinum

SPIN-ASSP project 1986-1991

In a previous experiment we have found that vowel formant target frequencies do not change (or at least not in the way predicted by current articulation theories) when speaking rate is higher than in normal speech. Target frequencies are a stationary measure of vowel articulation and give no information on the dynamical behaviour of formants. Therefore we performed an experiment to examine the influence of vowel duration and speaking rate on vowel formant track shape (i.e. dynamics).

We analyzed all acoustic realizations of 7 Dutch vowels ($n > 500$) in one long text read by a single male speaker at two different speaking rates, fast and normal. It was investigated how the contours of the first two formants of the vowels changed under these different speaking rate conditions. The formant track shape was assessed on a point-to-point basis, using 16 samples at the same relative positions in the vowels. Global measures of track shape that measured up to fourth order features of complete formant tracks were used. No differences between speaking rates were found that could be described as spectral vowel reduction or spectral undershoot. Only a uniform rise in first formant frequency, independent of vowel identity, could be proven. Within both speaking rates a small, but statistically significant, correlation between vowel duration and second order component of the first formant tracks was found. This implies a decrease in the excursion size within the track with shorter durations. The size of the effect makes a simple cause-effect relation with duration unlikely, phonemic context caused a far bigger and rate-consistent effect. The second order component of the vowel formant track shape shows to be an independent guide to vowel identity.

From the above mentioned experimental results it can be concluded that in speech production the parameters that describe vowel formant tracks, i.e. target frequency and track shape or curvature (scaled for duration), are relatively independent from the absolute duration of the vowel realization. This raises the question of how listeners perceive differences in these parameters and durations, i.e. in what way is the perceived identity of a vowel realization affected by changes in formant target frequency, track shape, and duration. In a perception experiment, we assessed the relation between and relative importance of formant target frequency, formant track shape (i.e. different parabolic curvatures), and the absolute duration in the identification of synthetic vowels. Vowel realizations were synthesized with varying duration and second order (i.e. parabolic) curvature of the formant tracks, but with identical formant target frequencies. The range in which these parameters were varied were taken from the production data (see above). A group of 29 listeners (both naive and trained listeners) were asked to identify each of the individual stimuli as one of the Dutch vowels.

Son, R. van (1987). 'Automatic slope measurement on formant tracks'. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 11, 67-78.

Son, R.J.J.H. van (1988). 'Automatic segmentation and stylisation of parameter contours'. In: *Reader Colloquium Singaalanalyse en Spraak, Den Haag*, 57-60. (in Dutch).

Son, R.J.J.H. van & Pols, L.C.W. (1988). 'Differences in formant values of Dutch vowels due to speaking rate'. *Proceedings of Speech'88, 7th FASE Symposium, Edinburgh, Book 1*, 313-320.

Son, R.J.J.H. van & Pols, L.C.W. (1989). 'Comparing formant movements in fast and normal rate speech'. In: Tubach, J.P. & Mariani, J.J. (Eds.), *Proceedings EUROSPEECH'89, Paris, CEP Consultants Ltd, Edinburgh, Vol. 2*, 665-668.

Son, R. van (1990). 'Formant frequencies of Dutch vowels in a text, read at normal and fast rate'. *SPIN-ASSP report 16, Utrecht*, 22 pp.

Son, R.J.J.H. van & Pols, L.C.W. (1990). 'Formant frequencies of Dutch vowels in a text, read at normal and fast rate'. *J. Acoust. Soc. Am.* 88 (4), 1683-1693.

LET/FW/036/10/1 Perception of dynamic speech-like signals

drs. A. van Wieringen

prof. dr. ir. L.C.W. Pols, dr. M.E.H. Schouten, prof.dr. E. de Boer

AIO-project 1989-1994

Due to the dynamic nature of the speech signal - caused by context effects such as speaking rate and speaking style - acoustic segments of the speech signal hardly ever correspond uniquely to the percept. In order to understand how listeners make use of the acoustical variation, the relevant importance of stationary and dynamic variables, resembling contextual conditions in natural speech, may be understood from various perceptual tasks. By means of different methodological paradigms and carefully manipulated synthetic stimuli, the physical correlates of the vowels and consonants, i.e. the duration(s) and frequency(ies) of the steady-states, and the duration(s), rate(s) and frequency range(s) of the transition(s), will be combined such that the processing of speech-like sounds, presumably void of higher-level, linguistic, knowledge, can be investigated in various contexts.

In the mutual interaction of quasi-stationary and dynamic variables many cues are present for each phonetic contrast, so that the perception of one cue is not independent of the information delivered by other perceptual attributes. Instead of measuring independent cues under certain conditions, the interaction of dynamic and stationary cues, obviously also governed by the context and the task in which they are presented, is tested perceptually in search of perceptual compensation of one parameter for another. Apart from getting insight into the process of speech perception, the perceptual dimensions for speech sounds should also be represented in terms of the transformation of the acoustic signal by the peripheral auditory system.

The theoretical importance of this project lies in the increasing fundamental knowledge of the processing of spectrally-complex sounds. From a practical point of view speech synthesis may benefit from the findings of the perception experiments. To date, three preliminary triad comparison experiments with one-formant synthetic CVC-like stimuli, were performed. The first two sets were designed such that there was a perceptual trade-off between three transition durations and three target frequencies. From these tests it was concluded that two of three spectrally-complex sounds may sound alike, despite different target frequencies, provided that the stimulus having the highest (or in exp. 2 lowest) target frequency is compensated by a longer/slower trajectory. In a third experiment it was found that much of the perceptual equivalence is based on transition rate. Ongoing research will focus on compensation effects of various spectro-temporal changes, tested by means of different paradigms, such as matching, triadic/paired comparison, etc.

Pols, L.C.W., Boxelaar, G.W., and Koopmans-van Beinum, F.J. (1984). 'Study of the role of formant transitions in vowel recognition using the matching paradigm'. *Proceedings of the Institute of Acoustics*, 6(4), 371-378.

Schouten, M.E.H. (1985). 'Identification and discrimination of sweep tones'. *Percep. Psychophys.* 37 (4), 369-376.

Schouten, M.E.H. (1986). 'Three-way identification of sweep tones'. *Percep. Psychophys.* 40 (5), 359-361.

Schouten, M.E.H. and Pols, L.C.W. (1986). 'Identification and discrimination of single formant sweeps'. *PRIPU*, vol. 11, 2, 21-33.

Pols, L.C.W. (1987). 'Chairman report of general discussion of session 4: Timbre'. In: M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception*, Martinus Nijhoff Publishers, Dordrecht, 328-330.

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. (1989). 'Studying speech perception: From simple stationary stimuli to more and more complex speech(-like) signals'. Proceedings 28th Conf. on Acoustics, Strebeké Pleso, High Tatras, Czechoslovakia, 27-34.
- Wieringen, A. van (1989). 'Discrimination in densely and sparsely spectralled harmonic complexes', extended version. IFA-report 109, 82 pp.
- Schouten, M.E.H. & Pols, L.C.W. (1989). 'Identification and discrimination of sweep formants', *Percep. Psychophys.* 46 (3), 235-244.
- Wieringen, A. van & Pols, L.C.W. (1989). 'Discrimination in densely and sparsely spectralled harmonic complexes', Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 27-43.
- Wieringen, A. van & Pols, L.C.W. (1990). 'Transition rate-dependent processing of one-formant speech-like stimuli', Proceedings of the Institute of Phonetic Sciences Amsterdam 14, 1-16.

LET/FW/038/10/1 **The predictability of vowel reduction**

drs. D.R. van Bergem

dr. F.J. Koopmans-van Beinum

AIO-project 1987-1992

In this project the phenomenon of vowel reduction is studied in relation to style of speech and the linguistic factors word class (the distinction between function words and content words), word stress and sentence accent. Up till now two major investigations were undertaken. In the first one, vowels from the free conversation of a male subject were compared with vowels from a text that was read aloud by the same subject. This text was a (grammatically polished) written copy of this free conversation. This study revealed a significant loss of vowel quality in the conversational style (smaller duration and spectral reduction). In the second investigation the influence of the above mentioned linguistic factors on vowel pronunciation was examined in sentences that were read aloud by 15 male subjects. In order to exclude coarticulatory effects, identical CVC syllables were used in all experimental conditions. It was found that all linguistic factors had a significant effect on both duration and spectral quality of the vowels. In future research we also intend to investigate the perceptual effect of vowel reduction in relation to these linguistic factors.

- Bergem, D.R. van (1988). 'The first step to a better understanding of vowel reduction'. Proceedings of the Institute of Phonetic Sciences Amsterdam 12, 61-75.
- Bergem, D.R. van (1989). 'Phonetic and linguistic aspects of vowel reduction'. Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 97-105.
- Bergem, D.R. van & Koopmans-van Beinum, F.J. (1989). 'Vowel reduction in natural speech'. In: Tubach, J.P. & Mariani, J.J. (Eds.), Proceedings EURO-SPEECH'89, Paris, CEP Consultants Ltd, Edinburgh, Vol. 2, 285-288.
- Bergem, D.R. van (1990). 'The influence of linguistic factors on vowel reduction'. Proceedings of Linguistics and Phonetics '90, Prague (in press).
- Bergem, D.R. van (1990). 'In defense of a probabilistic view on word recognition'. Proceedings of the Institute of Phonetic Sciences Amsterdam 14, 51-62.

LET/FW/040/10/1 **A model for the sound structure of natural language**
drs. P.P.G. Boersma
prof. dr. ir. L.C.W. Pols, drs. L.F.M. ten Bosch
AIO-project 1989-1993

The sound systems of languages are thought to have been built on these two principles: that the ease of articulation of words and the perceptual distinctions between words should be as high as possible. Either principle has to be evaluated in the context of the language: phonologies are structured to a much greater extent than could be expected from universal gauges of articulatory effort and perceptual contrast being applied to individual words. These linguistic organizations of speech production and perception are topics of a different branche of science, however. The phonetician is after what remains if structure is accounted for. Restricting drastically the number of possible sounds allows us to see beyond structure into the physics of speech. At this level, the possibilities of speech production are the movements of the speech organs, and the perceptual space is not at variance with the acoustic space. For example, if we restrict the set of possible consonants to the non-nasal labial consonants {p,ph,b,f,v}, we could ask: if a language would allow only three of them, which would it choose? I have argued (Boersma, 1989) that there can be circumstances under which there will allways exist a better system than the current one. Instead of focusing on manner features, as in the example, we could consider taking looks at small sets of sounds that differ only in, say, place of articulation or tone. In order to prevent, however, that these investigations will soon meet their boundaries, structure will be looked for in the motoric organization of the movements of the speech organs.

Boersma, P.P.G. (1989). 'Modelling the distribution of consonant inventories by taking a functionalist approach to sound change'. Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 107-124.

Boersma, P. (1990). 'Modelling the distribution of consonant inventories'. Proceedings of Linguistics and Phonetics '90, Prague (in press).

LET/FW/041/10/1 **Speech Processing Expertise Center (SPEX)**
prof. dr. ir. L.C.W. Pols
1989-1993

This project is funded by the Dutch Ministry of Science and Education. Its aim is to create in four years time the infrastructure (hardware and software) for recording, storing, retrieving, managing, and processing a variety of speechdata(bases). The Center is actually situated at the PTT Research Neher Laboratory in Leidschendam. Our Institute of Phonetic Sciences is more than just interested in this initiative. Prof. Pols is actually chairing the Board of SPEX. It will be clear that SPEX can and will play an active role both in the ESPRIT-SAM project (LET/FW/026/28/2), as well as in the SPIN-ASSP project (LET/FW/033/10/1).

LET/FW/043/10/1 Speech input/output assessment and speech databases

prof. dr. ir. L.C.W. Pols

dr. F.J. Koopmans-van Beinum, dr. R.A.M.G. van Bezooijen,

ir. L.J.P. van Heugten, ing. H.J.M. Steeneken

This project is not a research project as commonly defined, although it took much time and energy to prepare and organize the four-days ESCA Tutorial Day and Research Workshop (ETRW) on 'Speech Input/Output Assessment and Speech Databases' in September 1989 in Noordwijkerhout, The Netherlands. An international audience of 61 people participated in the Tutorial day, whereas 110 people actively participated in the subsequent workshop. The proceedings of both Tutorial Day (7 papers) and Workshop (59 papers) were available at the start of the ETRW.

A selection of 13 papers, all revised and some of them extended and updated, was published recently in the International Journal Speech Communication.

Pols, L.C.W., Heugten, L.J.P. van, Koopmans-van Beinum, F.J., Bezooijen, R. van & Steeneken, H.J.M. (Eds.) (1989). 'Proceedings of the ESCA Tutorial Day and Workshop on Speech Input/Output Assessment and Speech Databases', Noordwijkerhout, The Netherlands, 20-23 September 1989.

Pols, L.C.W. (1990). (Ed.) 'Special issue on speech input/output assessment and speech databases'. Speech Communication 9 (4), 263- 388. Editorial/Biography, v-vi.

LET/FW/044/10/1 Phonetic and linguistic validation of ratings of speaking performances

drs. A. van Gelderen

prof. dr. ir. L.C.W. Pols, dr. G.C.W. Rijlaarsdam

NWO-project 300-167-012, 1989-1993

In previous studies a rating scheme has been constructed and partly validated for the evaluation of performances of pupils of 11 to 12 years of age on several oral tasks. The scheme consists of four dimensions, presumably expressing the main functions speakers generally perform, namely Reference (informative function), Articulacy (identifiability of speech), Delivery (appropriateness and credibility) and Fluency (continuity of speech). In the present study the two dimensions mentioned last, will be validated by correlating jury-ratings on a narrative task with phonetic and linguistic measures of the pupils' performances. Predictors of Delivery are measures for pitch and intonation variation, variation in intensity and the use of specific lexical items that are associated with the register of narrative production. Predictors of Fluency are measures of tempo and pausing as well as a classification of dysfluencies according to the reason for their occurrence and an analysis of errors that are not repaired.

Gelderen, A. van (1986). 'The validation of analytic ratings of oral performances'. Tijdschrift voor Taalbeheersing 8 (3), 204-221. (in Dutch).

Bergh, H. van den & Gelderen, A. van (1986). 'Functional oral tasks in national performance surveys and their possible applications in education'. SLO-katernen 8, 29-35. (in Dutch).

Gelderen, A. van (1987). 'Language measures; construction of detailed rating procedures for speech and writing in support of performance surveys. Part 1: rating of oral performances'. SCO-report 138, Amsterdam. (in Dutch).

Gelderen, A. van (1987). 'Differential rating of performances on oral tasks in a large scale survey in the Netherlands'. International Oracy Convention, Norwich, ERIC Document Reproduction Service ED 300851.

LET/FW/045/10/1 **Speech database**
drs. L.W.M. Damen
prof. dr. ir. L.C.W. Pols
SPIN-ASSP project 1989-1991

The aim of the project is to develop a speech database of one professional speaker. All segments will be labeled and stored in a database. This project is part of the Dutch SPIN-ASSP program "Analysis and Synthesis of Speech" and is performed in collaboration with the University of Leyden.

LET/FW/048/10/1 **Spectro-temporal aspects of reduction and expansion in spontaneous speech and read text**

drs. G.P.M. Laan
dr. F.J. Koopmans-van Beinum
AIO-project 1991-1996

The aim of this project is to study the communicatively important aspects in spontaneous speech as compared to read texts, and the function of reduction and expansion in speech perception. In order to investigate this, we are studying global and local spectral, temporal, and intonational aspects of focus words (i.e. words bearing the highest load of semantic information) and non-focus words (the remaining part) in spontaneous speech and in the same texts, re-read after orthographic transcription.

For this purpose use is made of the speech material produced by a number of professional speakers, both males and females. The voice and pronunciation of one of these speakers served as a model for the diphone-based component of the Dutch national speech synthesis program. Moreover, this speaker has also been one of the subjects in our previous research on vowel reduction in various conditions.

Results will be of importance for basic research and in speech technology. Especially in man-machine interaction and in spoken dialogues these functional aspects of speech are of high importance.

Koopmans-van Beinum, F.J. & Bergem, D.R. van (1989). 'The role of "Given" and "New" in the production and perception of vowel contrasts in read text and in spontaneous speech'. In: Tubach, J.P. & Mariani, J.J. (Eds.), *Proceedings EUROSPEECH'89*, Paris, CEP Consultants Ltd, Edinburgh, Vol. 2, 113-116.

Koopmans-van Beinum, F.J. (1990). 'Spectro-temporal reduction and expansion in spontaneous speech and read text: the role of focus words'. In: *Proceedings of the 1990 International Conference of Spoken Language Processing*, Kobe, Japan, Vol. 1, 21-24.

AREA 2: Speech communication en computer science

LET/FW/026/28/2 **ESPRIT-SAM Speech assessment methods**

prof. dr. ir. L.C.W. Pols
ESPRIT-project 1984-1992

The full title of this European cooperative ESPRIT project SAM is 'Multilingual speech input/output assessment methodology and standardization'. In this project our institute works together with the TNO Institute for Perception in Soesterberg and the PTT Research Neher Laboratory in Leidschendam, The Netherlands, as well as with many other laboratories all over Europe. The aim is to develop methods for evaluating the performance of speech technology systems. Prof. Pols coordinates one of the three Working Groups on Speech Output Assessment. In this group attention varies from phoneme and word intelligibility, to prosodic quality, system acceptability, listener variability, and other aspects. The generally agreed use of a PC-based SAM workstation allows for extensive software exchange, universal testing methods, and comparable test results in as far as that is possible over languages. Pols is also a member of the SAM Management Team and was asked by Brussels to act as an external reviewer of Polyglot, another ESPRIT speech project.

- Pols, L.C.W. (1987). 'Quality evaluation of text-to-speech synthesis systems', Deliverable of ESPRIT-SAM project no. 1541, IFA-report 94, 38 pp.
- Pols, L.C.W. (1987). 'Speech technology and corpus linguistics'. In: W. Meijs (Ed.), *Corpus linguistics and beyond*. Proceedings of the seventh international conference on English language research on computerized corpora. RODOPI, Amsterdam, 285-294.
- Pols, L.C.W., Lefèvre, J.-P., Boxelaar, G.W. & Son, N. van (1987). 'Word intelligibility of a rule synthesis system for French', *Proceedings of the European Conference on Speech Technology*, Edinburgh, 1987, Vol. 1, 179-182.
- Son, N. van & Pols, L.C.W. (1987). 'Evaluation of a French diphone based synthesis system', Deliverable of ESPRIT-SPIN project no. 64 'Speech interface at office workstation', IFA-report 93, 66 pp. + app.
- Pols, L.C.W. (1989). 'Assessment of text-to-speech synthesis systems'. In: Fourcin, A., Harland, G., Barry, W., and Hazan, V. (Eds.), *Speech input and output assessment. Multilingual methods and standards*, Ellis Horwood Ltd., Chichester, 1989, Ch. III, 53-81 plus 251-266.
- Pols, L.C.W. & ESPRIT-SPIN partners (1989). 'Speech Interface at Office Workstation: Final report'. ESPRIT-SPIN deliverable, 181 pp.
- Pols, L.C.W. (1990). 'Assessing the speech quality of text-to-speech synthesizers'. *Proceedings VERBA '90, International Conference on Speech Technology*, Rome, Italy, 295-302.
- Pols, L.C.W. (1990). 'Does improved performance also contribute to more phonetic knowledge'. *Proceedings of the ESCA Tutorial Day on Speech Synthesis*, Autrans, France, 49-54.
- Pols, L.C.W. (1990). 'How useful are speech databases for rule synthesis development and assessment?'. In: *Proceedings of the 1990 International Conference of Spoken Language Processing*, Kobe, Japan, Vol. 2, 1289-1292.
- Pols, L.C.W. (1990). "'Standardized" synthesis evaluation methods'. In: *Proceedings International Workshop on International Coordination and Standardization of Speech Database and Assessment Techniques for Speech Input/Output*, Kobe, Japan, 53- 60.

AREA 3: Speech communication: development and pathology

LET/FW/029/14/3 **The influence of an oral plate on speech development and interaction in the first years of life of infants with a cleft lip and palate**

drs. J.M. van der Stelt, drs. K. Jansonius-Schultheiss

dr. F.J. Koopmans-van Beinum

Praeventiefonds-project 28-1042, 1984-1990

The effects of an oral plate applied after birth, in infants with a complete cleft or cleft palate only, were examined in two ways: longitudinally in a small group of children (from 0;0 to 2;0), as well as once-only in a group of forty children at the age of 2;0 years. As for the longitudinally followed children (12 cleft palate children and 6 normal born infants) four-weekly video recordings were made at the infants' home, in naturalistic setting and in interaction with their mother. Three of the normal born infants were recorded bi-weekly during their first year of life. One of the major topics in the project was to develop an adequate methodology in describing and analyzing the development of speech production and interaction. This was done in order to determine how *speech* communication develops from a process of *total* communication.

The effect of the oral plate in the group of 40 children of 2;0 years (30 cleft palate children, viz. 13 with and 17 without a plate, and 10 normal born children) was based upon the speech material collected by video registrations during 20 minutes in a laboratory situation in which mother and child played together. The phonetic, phonological, and grammatical capacities of the child were investigated to see whether the oral plate promoted a better speech and language production as well as a better speech interaction. Within the group of two-year-old cleft-palate children it became obvious that children with an oral plate reached a better speech motor control than those without an oral plate. The appliance of the oral plate was also positively related to the verbal interaction between mother and child. It turned out that, with relation to speech production and interaction, the group of children provided with an oral plate, approached the results of the normal group of children substantially more than the children without an oral plate, although they all are clearly removed from the level of the normal group.

As a result of the deviant mechanism, the development of speech motor control of the cleft-palate children differs from that of normal children: during the first stages the rate of development in both groups of children is almost identical. There is some difference with regard to the production of single articulatory movements, backward in the oral cavity. In cleft-palate children these sounds are acquired in a later phase. With respect to the development of interaction (with the mother), normal and cleft-palate children first show the same tendencies, but as soon as speech emerges the defective speech motor control in the cleft-palate children may cause retardations in the interaction process.

After the termination of the grant the project is continued in two Ph.D. studies (project LET/FW/050/14/3 by drs. J.M. van der Stelt, and project LET/FW/051/14/3 by drs. K. Jansonius-Schultheiss).

Koopmans-van Beinum, F.J., Stelt, J.M. van der & Jansonius-Schultheiss, K. (1984). 'Speech development and interaction in the first years of life of normal and cleft-palate babies'. Proceedings of the Institute of Phonetic Sciences 8, 65-69.

- Jansonius-Schultheiss, K. (1985). 'Is there a relation between eating and learning to speak. An outline of research on mouth behaviour in infancy'. *Proceedings of the Institute of Phonetic Sciences* 9, 93-101.
- Jansonius-Schultheiss, K. and Stelt, J.M. van der (1985). 'Annual Report 1984-1985 to the Praeventiefonds'. Cleft-palate Research Project nr. 28-1042. IFA-report 80, 10 pp. (in Dutch).
- Stelt, J.M. van der (1985). 'Speech development and the single subject research design'. *Proceedings of the Institute of Phonetic Sciences* 9, 109-115.
- Feenstra-Ikink, G.G.J. & Jansonius-Schultheiss, K. (1986). 'Developments in the treatment of cleft lip and palate children'. *Logopedie en Foniatrie*, 58, 175-181 (in Dutch).
- Jansonius-Schultheiss, K., Stelt, J.M. van der & Koopmans-van Beinum, F.J. (1986). 'Relation between mouth functions and onset of babbling stage, oral aspects in feeding and early speech development in normal and cleft lip and palate infants'. *Logopedie and Foniatrie*, 58, 238-241 (in Dutch).
- Jansonius-Schultheiss, K.(1986a). 'Mouth movements and mouth behaviors: a dialogue between dentists and speech pathologists'. *Nederlands Tandartsenblad*, 41, 50-54 (in Dutch).
- Jansonius-Schultheiss, K.(1986b). 'Dysphasia in children'. *Logopedie en Foniatrie*, 58, 102-107 (in Dutch).
- Reichart, A. (1986). 'Speech in relation to surgery in cleft palate infants'. IFA-report 84, 14 pp. (in Dutch).
- Koopmans-van Beinum, F.J. (1987). 'Cleft lip/palate'. In: C.R. Reynolds & L. Mann (Eds.), *Encyclopedia of Special Education*. John Wiley & Sons, New York, 347-348.
- Jansonius-Schultheiss, K., Coppenolle, L. van & Beyaert, E. (1987). 'Deviant mouth-habits and speech. Introduction, assessment, and therapy'. 4th revised edition. ACCO, Leuven, 240 pp. (in Dutch).
- Jansonius-Schultheiss, K. (1987). 'Not in touch: An example of a communication disorder in development'. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 11, 119-128.
- Overes, A.C.M. (1987). 'Anne. Verbal interaction in the first six month of a child'. IFA-report 90, 46 pp. (in Dutch).
- Stelt, J.M. van der (1987). 'Facial and gaze behaviour in mother-infant interaction'. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 11, 111-118.
- Jansonius-Schultheiss, K. (1988). 'Mother-child interaction concerning two-year-old cleft-palate children'. In: Janssen, P. (Ed.), *Afscheidssymposium Prof. Dr. P.H. Damsté, Department of Phoniatics, AZU, Utrecht*, 21-23. (in Dutch).
- Jansonius-Schultheiss, K. (1988). 'How parents and their baby (with or without a cleft palate) jointly learn to speak'. *BOSK-nieuws* no. 5/11, 20-23. (in Dutch).
- Dijkstra, A. (1989). 'Manual for Protin'. IFA-report 106, 45pp. (in Dutch).
- Jansonius-Schultheiss, K. (1989). 'Oral perception and communicative interaction in speech development: feeding a cleft lip and palate infant'. In: O. Kriens (Ed.), *What is a cleft lip and palate? A multidisciplinary update. Proceedings of an Advanced Workshop, Bremen 1987*, 76-79.
- Jansonius-Schultheiss, K. (1989). 'Speaking together: the communication development of a mother and her cleft lip and palate child, during the first two years of life'. In: Aufdemkampe, G., Beyaert, E.M.M.L., Bour, H.L.M.G. & Praag, W. van (Eds.), *Effectonderzoek van fysiotherapie en logopedie. Utrecht*, 184-198. (in Dutch).
- Jansonius-Schultheiss, K. (1989). 'The communicative development of two children with a cleft lip, alveolus, and palate'. *Logopedie-Foniatrie* 61, 348-352. (in Dutch).
- Stelt, J.M. van der (1988). 'Sensitive stages in face-to-face interaction in the first two years of life'. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 12, 41-59.
- Stelt, J.M. van der (1989). 'Learning to communicate in the wink of an eye'. *Bulletin 1988-1989 Netwerk Eerste Taalverwerving*, 18. (in Dutch).
- Stelt, J.M. van der (1987). 'Facial and gaze behaviour in mother-infant interaction'. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 11, 111-118.
- Koopmans-van Beinum, F.J., Jansonius-Schultheiss, K. & Stelt, J.M. van der (1990). 'The influence of an oral plate on the speech development and interaction in the first years of life of cleft-palate infants'. Final report to the Netherlands Prevention Fund concerning project 28-1042. Report 110 of the Institute of Phonetic Sciences Amsterdam , 158 pp. (in Dutch).
- Jansonius-Schultheiss, K. & Stelt, J.M. van der, 'Literature Review for the research project: The influence of an oral plate on the speech development and interaction in the first years of life of cleft-palate infants'. Final report to the Netherlands Prevention Fund concerning project 28-1042. Report 111 of the Institute of Phonetic Sciences Amsterdam , 68 pp. (in Dutch).
- Jansonius-Schultheiss, K. & Stelt, J.M. van der, 'Parents' Book: Questions on the development of your baby'. Final report to the Netherlands Prevention Fund concerning project 28-1042. Report 113 of the Institute of Phonetic Sciences Amsterdam , 47 pp. and 3 appendices. (in Dutch).

Stelt, J.M. van der & Jansonius-Schultheiss, K., 'Code book for mother-infant interaction from 0 to 24 months'. Final report to the Netherlands Prevention Fund concerning project 28-1042. Report 112 of the Institute of Phonetic Sciences Amsterdam , 31 pp. and appendix. (in Dutch).

LET/FW/037/14/3 Child and contrast: imitation, voicing and assimilation in speech development

drs. C.T.L. Kuijpers
dr. F.J. Koopmans-van Beinum
AIO-project 1987-1992

As has been indicated in the title, the project includes three topics. Firstly, we examine the structure and change of imitative speech forms of children between the ages 2;3 and 3;3 in a mother-child interaction. A methodology for classification of imitations has been developed. In the children's utterances evidence could be found of a progressive change from imitation towards a more adult-like dialogue.

Secondly, the study concentrates upon the question how young children control for the voiced-voiceless distinction in intervocalic position and how they grow towards the adult norm. Therefore, four different age groups participated in the experiment (four-year-olds, six-year-olds, twelve-year-olds and adults). A controlled set of two-syllabic words containing the intervocalic plosives /p,b/, /t,d/ and /k/ preceded by a short or a long vowel were recorded in a naturalistic setting. Results indicate that the relative length distinction of short and long vowels has been established in the youngest age group and improves with age. The difference of voiced and voiceless plosives in terms of closure duration also shows a clear growth towards the adult norm. The temporal coordination of vowel and consonant sequence, however, i.e. a shortening of the vowel before voiceless plosives and a lengthening before voiced plosive, is absent in the four-year-olds and only emerges in the six-year-olds to become more and more refined. It seems that the (adult) temporal compensation mechanism only matures with age. To date, we are preparing a listening experiment with children in the same age groups to study the role of closure duration and preceding vowel duration in the perception of voiced and voiceless plosives.

Thirdly, we examine the assimilation of voice in speech of the same children as described above and measurements are analyzed and interpreted within soon. We hope to contribute to a better understanding of the development of the speech production and perception mechanism.

Kuijpers, C.T.L. (1987). 'Imitations in speech development: a literature overview'. Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 103-110.

Haaren, H. van (1988). 'Imitation in the verbal mother-child interaction. A comparative study concerning six mother-child pairs with three of the children born with a cleft palate'. IFA-report 102, 76 pp. (in Dutch).

Kuijpers, C.T.L. (1989). 'Strategies in developmental dialogue system'. In: Tubach, J.P. & Mariani, J.J. (Eds.), Proceedings EUROSPEECH'89, Paris, CEP Consultants Ltd, Edinburgh, Vol. 2, 593-596.

Kuijpers, C.T.L. (1989). 'The voiced-voiceless distinction in four-year-old children'. Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 59-75.

Kuijpers, C.T.L. (1990). 'The voiced-voiceless distinction in young children'. In: Abstracts of the Fifth International Congress for the study of child language, Budapest, Hungary, 111.

LET/FW/039/14/3 **Objective evaluation of the pronunciation of words by deaf speakers**

ir. M.J. Bakkum
prof. dr. ir. L.C.W. Pols
1987-1991

In this project, which is a continuation of NWO-project 300-161-038, the relation is studied between subjective quality of pronunciation as judged by listeners, and objective, physical properties of CVC-type words as spoken by deaf speakers and normal-hearing speakers. We also try to compare the judgements at segmental level (vowels and consonants) with more general judgements at sentence level. Objective, physical, measures are based on stationary and dynamic bandfilter spectra. This project is sponsored by the Institute for the Deaf, Sint-Michielsgestel and runs at the Free University of Amsterdam. Pols is one of the grant applicants.

Bakkum, M.J., Plomp, R. & Pols, L.C.W. (1989). 'Objective evaluation of word pronunciation by filter-band analysis'. In: Tubach, J.P. & Mariani, J.J. (Eds.), Proceedings EUROSPEECH'89, Paris, CEP Consultants Ltd, Edinburgh, Vol. 1, 425-428.

LET/FW/042/14/3 **The development of pitch and temporal structure starting from the early lingual period**

dr. E.A. den Os
dr. F.J. Koopmans-van Beinum
NWO-project 560-256-035, 1987-1991

This project deals with the development of prosody in relation to language development. We studied utterances of one child who was one year and three months at the beginning of the projects and three years old at the end. We monthly made audio and video recordings of mother-child interaction in a free-play situation at home. Twenty minutes per recording were transcribed and the utterances were analysed on syntactic and pragmatic structures, on intonation and duration. All data were stored in a database. Concerning the intonational part, we concentrated on the final contour direction of the child utterances. Till the age of one year and ten months the child did not use rises in an adult way: there were relatively many rises in non-question situations. We now investigate whether the mother was also using relatively many rises to the child, possibly as a part of her 'motherese'.

Concerning the durational part, we found no clear signs of shortening of mean syllable- or mean phoneme duration with growing age and growing grammatical complexity. On the contrary, at the age of two years and four months, the child spoke very slowly. And after this month (2;4 years of age) we saw a clear increase of utterance length (measured in number of syllables), a decrease of standard deviation of mean syllable duration, and an increase in the number of hesitations. These findings point in the direction of a restructuring of articulation at about this age.

Os, E.A. den & Koopmans-van Beinum, F.J. (1989). 'Measuring pitch patterns within the scope of early speech development'. Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 89-96.

Os, E.A. den (1990). 'The development of prosody in relation to language acquisition'. NET-bulletin 1990, 6-8. (in Dutch).

Os, E.A. den (1990). 'The development of prosody'. In: Abstracts of the Fifth International Congress for the studie of child language, Budapest, Hungary, 56.

Os, E.A. den (1990). 'Development of temporal properties in the speech of one child between one and three years of age'. Proceedings of the Institute of Phonetic Sciences Amsterdam 14, 39-52.

LET/FW/046/14/3 **Acoustical aspects of a model of early speech development**

dr. F.J. Koopmans-van Beinum
1990-1993

In this project it is investigated how infants learn to use their sound production mechanism as a speech instrument, within the developmental process towards adult speech communication. Initially a neonate's vocal activities are limited to crying and producing vegetative sounds, but in the course of the first months sound production develops considerably. Since neonates normally are born with a complete sound production mechanism, a transcription system is designed that is based on the source-filter model of speech production. Within this system a breath group is used as a segmentation unit, whereas for the source-filter description use is made of aspects of phonation and articulatory movements within the framework of the anatomical and physiological capacities of the infant. The use of this transcription system results in finding six developmental stages in the infant's sound productions in the first year of life, focussing mainly on syllabic aspects. For each of these stages a number of acoustical parameters are studied in order to create a possibility to compare normal and deviant developmental processes and to relate early sound productions to adult use of speech.

- Koopmans-van Beinum, F.J. (1987). 'Early capacities for speech communication'. Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 89-102.
- Koopmans-van Beinum, F.J. (1989). Book review of A.M. Schaerlaekens & S. Gillis, 'De Taalverwerving van een kind' (Groningen: Wolters-Noordhoff, 1987). In: First Language, Vol. 9, 325-326.
- Wieringen, A. van & Koopmans-van Beinum, F.J. (1989). 'Acoustic analyses of the vocalisations of a hearing-impaired infant: exploring methods'. Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 77-87.
- Koopmans-van Beinum, F.J. (1990). 'A database for infant speech: wish-dream or possibility in future?' NET-bulletin 1990, 38-42. (in Dutch).
- Koopmans-van Beinum, F.J. (1990). 'The source-filter model of speech production applied to early speech development'. Proceedings of the 1990 International Conference of Spoken Language Processing, Kobe, Japan, Vol. 2, 1321-1324.

LET/FW/047/14/3 **The development of an instrument to analyse the effect of the clinical treatment on voice quality: the relation between perceptual, acoustical and clinical parameters.**

drs. I.M. de Leeuw

dr. F.J. Koopmans-van Beinum, drs. L.W.A. van Herpt

AIO-project 1990-1995

To date studies to analyse the quality of voice of patients with early glottic cancer before and after radiotherapy are seldomly done. Also there exists uncertainty about the optimal radiation dose delivered to these small glottic tumours. Optimal radiation dose should be based upon dose response curves for tumour control and the complications of the radiation. One of these complications is the decrease of voice quality. Voice quality can be investigated by means of perceptual, acoustical and clinical parameters. The purpose of the study is to develop an instrument to analyse the effect of the clinical treatment, i.e. the radiotherapy, on voice quality. Therefore the relation between perceptual, acoustical and clinical parameters will be analysed. Perceptual parameters includes judgements of listeners on semantic 7-point scales. The acoustical parameters are measures of pitch, tempo, spectral noise etc. The clinical parameters are tests related to aerodynamics, visual inspection of the larynx and others. These parameters will be gathered from patients with early glottic cancer before radiotherapy, and six months, and two years after radiotherapy as well as from a control-group of speakers.

Leeuw, I. de (1990). 'The relation between perceptual and clinical parameters of voice quality of patients with early glottic cancer before and after radiotherapy and normal speakers'. Proceedings of the Institute of Phonetic Sciences Amsterdam 14, 27-38.

LET/FW/050/14/3 **Early mother-infant interaction and their development to speech communication**

drs. J.M. van der Stelt

dr. F.J. Koopmans-van Beinum, prof. dr. ir. L.C.W. Pols

1990-1992

This Ph.D. study is a direct continuation of project LET/FW/029/14/3. In the present project two longitudinally followed mother-infant pairs are studied with regard to their development to speech communication during the first two years of the children's life. Subjectively judged these two pairs differ with regard to their quality of interaction. Their interaction is also described micro-analytically by means of a multi-channel coding system for movements and sound productions. Visual and intentional communication is studied specifically, regarding the development as a dynamic control system.

Stelt, J.M. van der (1989). 'Learning to communicate in the wink of an eye'. Bulletin 1988-1989 Netwerk Eerste Taalverwerving, 18. (in Dutch).

Stelt, J.M. van der (1990). 'Mother-child communication by eye and by ear'. In: Van Horen en Zeggen 31 (2), 51-56. (in Dutch).

Stelt, J.M. van der (1990). 'Are intentional movements basic to speech development'. In: Abstracts of the Fifth International Congress for the study of child language, Budapest, Hungary, addendum, Volume Abstracts, 13.

LET/FW/051/14/3 **Speech development in dialogue**

drs. K. Jansonius-Schultheiss

prof. dr. ir. L.C.W. Pols, prof. dr. A. Mills, dr. F.J. Koopmans-van Beinum
1990-1993

This Ph.D.study is a direct continuation of project LET/FW/029/14/3. In the present project the communicative development of twelve babies with and six without a cleft lip and/or palate (0;2 - 2;0 years of age) is studied. Use is made of four-weekly video registrations of the children in the home situation, during twenty minutes, in interaction with their mother. The focus of this study lies upon the longitudinally studied speech development and -interaction in each single case (of that specific mother-child pair), as well as in comparison with the other cleft-palate and normal born babies, and the comparison of the cleft-palate group with the group of normal born babies.

Questions to be answered are:

- what is the course of the speech development in terms of speech motor development;
- what is the course of the speech development in terms of speech interaction between mother and child;
- in what way do medical intervention and/or other psycho-social variables enhance the speech development;
- in what way can we describe, based upon the speech results, the transfer from infant speech movements into phonemes, embedded in words; from physiologically into physiologically-linguistic skills;
- in what way, given an established retardation or disturbance in the speech development, is the child able to overcome its problems by itself and/or by mother-child interaction and/or by medical intervention; based upon these results as well as upon the literature, suggestions will be given for an early intervention program.

LET/FW/052/14/3 **Development of early vocalisations in severely hearing-impaired children**

dr. F.J. Koopmans-van Beinum, dr. E.A. den Os

In order to investigate the role of speech sound input during the infants first year of life, in the developmental process towards adult speech communication, we recently designed a research project concerning the early sound productions of hearing-impaired children compared to normally hearing infants. This research design is now submitted for acquiring a grant-in-aid.

AREA 4: Female voices in phonetics

LET/FW/020/21/4 **Women can speak too...!**

drs. M.J.T. Tielen

dr. F.J. Koopmans-van Beinum, drs. L.W.A. van Herpt

AIO project 1986-1991

The aims of this project were to study experimentally sex-specific voice and pronunciation characteristics and to establish whether or not some presumed characteristics for either female or male speech are indeed apparent. Acoustic measurements have been performed with respect to the perceptually most prominent difference between male and female voices, viz. pitch. On the average, a difference of about 10 semitones was found for the fundamental-frequency level between male and female speakers. A larger range and variability for female voices, as mentioned in literature, could not be found, however.

Intelligibility experiments have been performed in order to determine the relative CVC and phoneme intelligibility of men and women in a number of interfering noise conditions. Despite of substantial differences with respect to intelligibility between individual speakers, no clear differences were found between the male and female speaker groups.

Also, an experiment with respect to perceptual voice and pronunciation characteristics of men and women has been performed. A sample of 60 male and female speakers was taken from three different profession categories, in order to study the influence of Social Economic Status on speech behaviour. The results by means of semantic scale scores are being analysed at this moment. Although many aspects of male and female voice and pronunciation have been left, only a small perception experiment with respect to pitch will be performed, before the project has to be concluded with a Ph.D.thesis.

- Koopmans-van Beinum, F.J. & Tielen, M.T.J. (1987). 'Are female voices more intelligible than male voices?' *J. Acoust. Soc. Am., Supp 1, Vol. 82, S85(A)*.
- Tielen, M.T.J. (1987). 'The relative intelligibility of male and female voices'. *Proceedings of the Institute of Phonetic Sciences Amsterdam 11, 27-33*.
- Tielen, M.T.J. (1987). 'A sociophonetic research in women's voices'. In: R. Knorringa, A. Burger & J. Bossinade (Eds.), *Uitgesproken niet uitgepraat. Lezingenbundel Vrouwenstudies Letteren. Amsterdam, 127-140. (in Dutch)*.
- Tielen, M.T.J. & Koopmans-van Beinum, F.J. (1988). 'The relative intelligibility of male and female voices'. *Proceedings of Speech '88, 7th FASE Symposium, Edinburgh, Book 1, 277-282*.
- Tielen, M.T.J. (1989). 'Intelligibility of male and female voices under a few noise conditions'. In: Tubach, J.P. & Mariani, J.J. (Eds.), *Proceedings EURO-SPEECH '89, Paris, CEP Consultants Ltd, Edinburgh, Vol. 2, 127-130*.
- Tielen, M.T.J. (1989). 'Fundamental frequency characteristics of middle aged men and women'. *Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 45-58*.
- Tielen, M.T.J. (1990). 'Perception of the voices of men and women in relation to their profession'. In: J. Laver, M. Jack & A. Gardiner (Eds.), *Proceedings of the Tutorial and Research Workshop on Speaker Characterization in Speech Technology, Edinburgh, 192-197*.

NOT INVOLVED IN RESEARCH AREA 1-4:

LET/FW/000/00/0 **Models of the cochlea**

drs. J.S.C. van Dijk

The spectral contents of sophisticated signals such as speech and music varies as a function of the time. In order to make this spectro-temporal message accessible, we first applied a recursive Fourier transform. As a result of this procedure, the spectral contents of a running window is depicted as a function of the time. Actually, this method defines a simple discrete dynamical system which does its duty in the time domain.

This kind of methods has intrinsic imperfections. Firstly, the signal enclosed by the running window is considered as one period of an infinitely periodic signal. Secondly, there is a lack of attention especially to rapid variations of a signal within the window. Model studies clearly show the background of some well-known psychophysical observations. For instance, results of masking experiments with complex sound stimuli can be simulated very well. Moreover, our simulations support the hypotheses which are basic in Goldstein's theory concerning the formation of pitch in complex signals.

In most cochlear models, the sharpness of the cochlear process of filtering and the frequency selectivity along the basilar membrane are closely related to each other.

Often, sharpness is modelled as an effect of a relatively low rate of the damping which contributes to the definition of the frequency selectivity along the membrane. In consequence of this, the response time of a filter is always too long in order to give an adequate representation of rapid variations in speech and music. Therefore, we changed the description of the cochlear filters considerably. Firstly, we raised the damping parameters to a more natural value. Secondly, we introduced additional weighting functions to the admittance of the membrane. These functions can be ascribed to the influence of boundary conditions on the motion of the basilar membrane. The resulting model shows that the main properties of frequency selectivity have been conserved and lead to well-defined filter characteristics with acceptable values of the response time.

Ultimately, we transformed the model to a dynamical system consisting of parallel filters which have been coupled to each other. As a first application, several speech utterances have been coded in terms of intensity and/or deflection patterns along the basilar membrane as a function of the time. A second application consists of studying the physical background of combination tones.

Dijk, J.S.C. van (1989). 'Notes on cochlear filtering'. Proceedings of the Institute of Phonetic Sciences Amsterdam 13, 133-143.

Dijk, J.S.C. van (1990). 'On dipoles and the radiating membrane'. In: Lecture Notes in Biomathematics, Springer-Verlag, New York. (In press, 8 pages).

Dijk, J.S.C. van (1990). 'Simulation of cochlear filtering'. In: J. Cornelis and S. Peeters (Eds.), Proceedings North Sea Conference on Biomedical Engineering 1990, Topic 8. Modelling and Simulation, 87-92.

TO BE ADDED TO AREA 1:

LET/FW/051/10/1 Knowledge implementation in stochastic recognition models

Drs. L.F.M. ten Bosch, prof. dr. ir. L.C.W. Pols

This project serves as a follow-up of a previous AIO-project (LET/FW/034/10/1) on automatic speech recognition ASR by means of Hidden Markov Modelling. Basically two approaches to ASR exist. One approach is based on 'template matching' and the concept of 'minimal distance'. Another approach is based on the statistical properties of the speech signal. Speech is then described as the output of a Markov system. The Markov approach is characterized by a learning phase as well as by a recognition phase. In the learning phase, the Markov system learns by iterative adjustment of many Markov parameters. This iterative procedure is not based on linguistic knowledge. In general incorporation of explicit knowledge is difficult.

In this project it is our aim to explicitly incorporate knowledge about one phonetic parameter, viz. the (segmental) duration. Recent theoretic research suggests that this is a realistic goal.