Fifty years of Phonetic Sciences in The Netherlands

Vincent J. van Heuven & Toni Rietveld

The General Society of Linguistics (Algemene Vereniging voor Taalwetenschap, AVT) in the Netherlands was founded 50 years ago. On the occasion of its 50th anniversary we were asked to look back and assess the achievements of the discipline of Phonetics, with special emphasis on the contribution made in the Netherlands in the last five decades, and speculate on developments that may take place in the next 50 years. Citing a well-known aphorism of disputed origin, predictions are extremely hazardous, especially when they are about the future, so that we are reluctant to engage in this second part of the assignment. Henry Ford, when introducing his first T-Ford model in 1928, told the assembled press: When fifty years ago people asked themselves how to travel faster, they could only think of breeding faster horses. Although Ford almost certainly never said anything of the sort, the message is clear enough. It is impossible to predict sudden breakthroughs in science and how they change the discipline. Therefore, we opt for the easier way out, and ask ourselves what ideas were around in phonetics at the end of the 1960s, and if anyone made any useful prediction at the time of what the field would be like by 2020.

In 1926 the University of Amsterdam appointed medical doctor Louise Kaiser reader in Phonetics. She retired in 1958 to be succeeded by Hendrik Mol, an electrical engineer specializing in microphones, who was employed by the Netherlands Telephone Company PTT, and who held a part-time professorship in Phonetics in Leiden. Once in Amsterdam Mol assembled a team of mathematicians and craftsmen who developed mathematical models of vowel acoustics and built hardware twin tubes to test and demonstrate the adequacy of the models. In the early 1970s the universities of Groningen and Nijmegen followed the Amsterdam example, and appointed engineers as professors of phonetics, i.e. Donald Graham Stuart and Wilhelm ('Felix') Vieregge, who in turn hired mathematically inclined lecturers, i.e. astronomer Tjeerd de Graaf in Groningen and Louis Boves and Toni Rietveld (both of whom had specialized in phonetics under Mol in Amsterdam) in Nijmegen. Meanwhile the Phillips Electronics Company had co-founded the Institute for Perception Research (IPO) as an annex to the Philips Physics Lab and appointed Antonie Cohen, a former structuralist phonologist, as the head of a speech research group which targeted the phonetics of speech prosody, a then seriously understudied aspect of spoken language. After eight years at Eindhoven, Cohen accepted a chair at Utrecht University, first as professor of English and from 1976 onwards as professor of Phonetics. In 1981 Sieb Nooteboom, Cohen's successor in Eindhoven (and later also in Utrecht), was appointed (part-time) professor of Phonetics in Leiden.

The hallmark of phonetic research as it was (and still is) practiced in the Netherlands is that theoretical accounts of human speech production and perception should be explicit enough to allow an engineer to build a machine (or write a computer program) that would essentially simulate the human speech process or part thereof. The hardware twin-tube models built by the Amsterdam group (Mol 1970) are one example of this, the work on talking computers (speech synthesis with correct temporal organization and melody) at IPO Eindhoven is another. The quality (and underlying simplicity) of the IPO synthesis of speech melody was internationally acclaimed ('t Hart et al. 1990, Ladd 1996).

In 1983 the Ministry of Science and Education recognized Phonetics as an independent scientific discipline but at the same time dictated that only one university in the country would

be accredited for the master diploma. The accreditation was awarded to Utrecht University. Competitors Amsterdam and Nijmegen then promptly started non-accredited specialization programs under deviant names such as Language and Speech Technology, Language and Speech Pathology, and similar. The covert animosity between 'the engineers' (Amsterdam, Nijmegen, Groningen) and 'the linguists' (Utrecht, Leiden, Eindhoven) was ended abruptly by a joint action of the Ministries of Science and Education and of Economic Affairs in the mid 1980s, when a M€ 6 research subsidy was granted contingent on national cooperation of the phonetic research groups. Hatchets were buried, pipes were smoked, and the Analysis and Synthesis of Speech Program (ASSP, Van Heuven & Pols 1993) was launched. Program leader Cohen was convinced that automatic speech recognition was principally impossible but at the same time had high hopes that top-quality speech synthesis (e.g. a reading machine for the blind) was within reach if engineers would avail themselves of insights that could only be obtained by linguists and phoneticians. Five years later intelligible text-to-speech conversion was achieved but the conversion was slow and the speech output still sounded non-human.

From our present vantage point, 30 years later, we may observe, that speech technology has indeed made enormous progress. Unrestricted text-to-speech conversion is almost impossible to distinguish from a human reader, thanks to variable unit concatenation based on huge databases of pre-stored spoken sentences. Virtually error-free automatic speech recognition is at our finger tips on every mobile phone, tablet or laptop computer for a wide range of languages, including Dutch. Talk shows on television can be automatically subtitled in real time, again virtually error-free. Voice conversion programs are now seen as a threat. These applications steal a person's voice so that we can make anyone say anything such that it takes an expert to prove the fabrication. Interpreting telephony allows an American speaker in New York to present a lecture in China online over the internet where his English speech is recognized, converted to text, automatically translated into Mandarin, and shown on screen as subtitles in Chinese characters in synchrony with highly intelligible Mandarin speech output. We mention these achievements with some trepidation because these technological advances have come about without much help from linguists and phoneticians. All these systems were developed basically by computer engineers and are driven by mathematical and statistical models that were extracted from enormous databases of human speech and language use. The role of linguists and phoneticians in this number-crunching (or 'brute-force') approach has been relatively minor and was limited to assisting the engineers in finding adequate inventories of linguistic units (e.g. sounds, syllables, morphemes, parts of speech) and coding systems for labelling ('tagging') these units. A major problem with the engineers' self-learning algorithms (e.g. Hidden Markov Models, Neural Networks) is that their internal structures do not map onto human speech production and perception in an insightful manner. The models do not tell us how the human mind works. Nevertheless, in pure research the algorithms are quite useful as heuristic tools. Whenever the self-learning algorithm performs its task better than an implementation of linguistic or phonetic human performance models, we know that essential knowledge is still missing. Then, by eliminating specific properties from the input to the self-learning algorithms, we may try to narrow down the (class of) properties that constitute the missing ingredient in our theoretical account. From this perspective the roles of pure phonetic/linguistic research and of technology have been reversed: it is no longer pure research that informs technology but rather the other way around.

At the first International Congress in The Hague (1928) Phonology was set up as a separate (sub-)discipline to study the structural regularities underlying sound systems rather than the myriad measurable differences between sounds, which part was left to phonetics. For

decades an iron curtain divided the two phonic disciplines. Although it was said on many occasions that abstract Phonology ignored Phonetics at its own peril, (e.g. Lehiste, 1970: vi) it was not until the end of the 1980s (note the parallel with the political "Wende" that took place in Eastern Europe at the same time) that serious attempts were made to reunite the two disciplines in what was alternately called Laboratory Phonology (Kingston & Beckman 1990), Phonetically Driven Phonology (Hayes, 1996) or Functional Phonology (Boersma, 1995).

Another landmark is the establishment of the Corpus of Spoken Dutch (Corpus Gesproken Nederlands, see the CGN-website). The aim was to document Standard Dutch, as spoken in Belgium and the Netherlands around the turn of the millennium. Ten million words of spoken text (~1000 hours) were recorded, transcribed and annotated at all relevant linguistic levels, i.e. syntax, morphology, phonology (including prosody) and orthography. This now publically available corpus was set up to serve the needs of speech technological applications (Spyns & Odijk, 2013) as well as fundamental (socio- and psycho-)linguistic research (Oostdijk, 2000). The recordings are synchronized with the annotated text and can be listened to and acoustically analyzed. One important speech-technological tool which has been interfaced with the corpus is Praat ('A system for doing phonetics by computer', Boersma & Weenink 1996), which by now is the most widely used software package for phonetic and phonological research.[1]

Rather than trying to predict the future of the discipline, we end this contribution by asking ourselves when linguistics (or phonetics) would be finished. Ultimately, linguistics should be able to understand and explain how human beings (and possibly other species as well) learn language(s) and how they use language to communicate, whether by speech, writing, signing or other means. The neuro-physiological aspects of speech communication have played an essential role in the study of speech and language pathology. We believe that within the coming decades abstract views of language and speech processes will be replaced by neuro-physiologically informed models. Such models will be implemented in simulations of human language learning and language behavior. Probably, such implementations will be called 'robots'. These should be able to learn any language (or even multiple languages) they are exposed to, as if they were infants – but preferably faster – and then be able to extract meaning from language input or convert intentions to linguistic output. These 'robots' will fulfill the role of the language engineering applications in the past decades but on a comprehensive scale covering the full linguistic capabilities of the human species.

References

Kingston, John. & Beckman, Mary E. (1990*). Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.

---

[1] The number of citations to Praat (10,424 in Google Scholar, as of 1 March 2019) exceeds that to any other linguistic publication produced in The Netherlands. Praat was conceived by Paul Boersma in the late 1980s as a graphics tool for drawing illustrations in his dissertation but was since then extended to include, among other things, a large number of acoustic measurement techniques, speech synthesis, neural networks, a scripting language, experimentation programs and an optimality constraint ranking module. The development was financially supported by NWO and the Speech Technology Foundation on several occasions. Libraries with Praat scripts have been disseminated over the web throughout the 'Praat community' and many speech researchers have written tutorials explaining how to use the package. Praat can be downloaded for free from *www.praat.org*.

Boersma, Paul (1998). *Functional Phonology: Formalizing the Interactions Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.

Boersma, Paul & Weenink, David (1996). Praat, a system for doing phonetics by computer. *Report nr. 136,* Institute of Phonetic Sciences, University of Amsterdam*.*

Hart, J. 't, Collier, R. & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Hayes, Bruce P. (1996). Phonetically Driven Phonology: The Role of Optimality Theory and inductive grounding**.** *Proceedings 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics.* [Rutgers Optimality Archive 158, http://ruccs.rutgers.edu/roa.html]

Heuven, Vincent J. van & Pols, Louis (eds., 1993). *Analysis and synthesis of speech. Strategic research towards high-quality text-to-speech generation*. Berlin/New York: Mouton de Gruyter.

Ladd, D.Robert (1996).*Intonational Phonology*. Cambridge: Cambridge University Press

Lehiste, Ilse (1970). *Suprasegmentals*. Cambridge, MA: The MIT Press.

Mol, Hendrik (1970). *Fundamentals of Phonetics*. The Hague/Paris: Mouton.

Oostdijk, Nelleke (2000). The Spoken Dutch Corpus. Overview and first evaluation. In: M. Gavralidou, G. Carayannis, S. Markantonatou, S. Piperidis, & G. Stainhaouer (eds.), *Proceedings of the Second International Conference on Language Resources and Evaluation* (pp. 887–893), Paris: ELRA.

Spyns, Peter & Odijk, Jan (eds., 2013). *Essential Speech and Language Technology for Dutch: Results by the STEVIN programme.* Berlin: Springer Verlag.