THREE AREAS OF COMPUTATIONAL PHONETICS

Hans G. Tillmann
Institute of Phonetics and Speech Communication, University of Munich

ABSTRACT

We propose to consider three areas of computational phonetics. The first area deals with speech signals transmitting phonetic information from the effectors to the receptors of natural nervous systems. The second one deals with the categories of phonetic facts accessed in large speech databases. The third one is devoted to computing time functions of given phonetic categories and to predicting the categories evoked by phonetic time functions.

INTRODUCTION

The term "computational phonetics" could be given many meanings. Is it the brother (or sister) of Computational Linguistics? This would restrict the meaning of the term to those parts of phonetics which make use of computers in a specific way. Indeed, many examples of this type of computational phonetics could be described, since computers, equipped with analog-to-digital and digital-to-analog converters, have become the most important, or even, in a number of cases, the only instrument of instrumental phonetics.

We also could look back into the history of instrumental phonetics where we will find clear instances which likewise could fall under the term computational phonetics. There are at least two prominent examples which we would like to mention.

The first one is probably the earliest instance of computational phonetics. We find it in the appendix of Scripture's "Elements of Experimental Phonetics" where the author, nearly 100 years ago, showed how the Fast Fourier Transform of a voiced speech signal can be computed using paper and pencil, after the amplitude of a pitch-period from a graphically recorded oscillogram has been optically magnified and manually sampled into equidistant discrete amplitude values.

Another very important historical example which must be mentioned here was presented in 1957 in an article entitled "Die Vokalartikulation als Eigenwertproblem" by Meyer-Eppler and Ungeheuer in Zsch. f. Phonetik. The authors made use a second order homogeneous differential equation (Webster's Horn equation) and showed how the three-dimensional geometry of the vocal tract can be reduced to the two-dimensional area function in order to compute the resonance frequencies of the human vocal tract during the production of vowels. Thus the values of formant frequencies became uniquely predictable by a mathematically formulated physical theory (cf. also Ungeheuer 1957, 1962 and Fant 1960).

It seems to make sense to restrict the term computational phonetics to theories that take values of some given type and compute new values which are not trivially available in the computed form. This restriction is useful only to exclude what shall not fall under a computational theory. So we need further criteria to determine which types of phonetic theories shall be part of computational phonetics. In the following we would like to argue that there are exactly three interesting areas of computational phonetics which should be particularly considered and metatheoretically investigated in more detail. It could well be the case that the future of phonetics as one of the speech sciences will depend on further theoretical developments in just these three areas.

(I) THE AREA OF CAUSAL SPEECH THEORIES: THE TRANSMISSION OF PHONETIC INFORMATION

During any real speech act many physical processes which remain transphenomenal to the normal speakers and listeners can be measured and represented in the form of digitally sampled discrete time functions. As soon as a speaker conducts an act of speech these physical processes are assumed to run in the brains of speakers and listeners as well as in all channels that connect the involved individual nervous systems; they synchronously accompany any naturally perceived speech utterance.

These processes are extremely complex time functions inside the verbally communicating neural systems and still widely unknown with respect to their specific segmental and prosodic neural form. However, at the very periphery of the communicating systems they become fairly simple [15] and can be easily represented by well manageable AD-converted discrete time functions. This is the reason why causal theories of speech transmission allow us to fill the gap between the communicating neural systems. The articulatory CVCVC-actions of the motor system of the speaker cause speech movements which are discretely mapped to the proximal receptors of the speaking system and are also indirectly transformed into the acoustic output which transmits all relevant phonetic information to the auditory receptors (also the visible speech input plays a role).

The theory of Meyer-Eppler and Ungeheuer has been further developed by Schroeder (1967) and others; so it is no problem today to compute the acoustic output from a given articulatory time function digitally representing the speech movements of a given speech utterance [1,12]. If we know the impulse response of the acoustic system and the given excitation, the output can be simply derived by a convolution (in the time domain) or by a multiplication of the z-Transforms (in the frequency domain). Thus in the first area of computational phonetics the transmission of phonetic information is made explainable by showing how the peripheral actions of the motor system cause a mathematical mapping to the receptors of the sensory systems of speakers and listeners.

(II) THE AREA OF CATEGORICAL SPEECH THEORIES: THE REPRESENTATION OF PHONETIC FACTS

The world of phonetic facts that are relevant to the speakers and listeners of a language can be satisfyingly represented by logically oriented programming languages (such as Prolog). They allow computation with these facts in a very effective way as soon as there is access to a database large enough to contain all possible instances of those facts. And if the utterances in the database are represented according to the CRIL conventions of the IPA (concerning the Computer Representation of Individual Languages [4]), the lexical items they are composed of can not only be identified by their orthographic form, but also automatically compared as to their canonic citation forms and the actual and factual realisations (varying in a Lindblomian H-H-space [9]).

This type of computationally approach to phonetic facts has been systematically developed by Christoph Draxler in the German Verbmobil-PhonDat-project [2]. The experiences up to now show two things: Categorical representation of phonetic facts on the 3rd CRIL-level, aligned to individual

sound segments, is still somewhat critical in the case of reduced words in spontaneous speech [3]. Here, new standards for representing less clear and unclear cases have to be established (IPA and SAMPA etc. have been mainly used for description of well articulated clear speech utterances). Secondly, it should be pointed out that symbolic data receive a new kind of factuality if they are connected to real speech utterances within a large database (even if the physically recorded speech signals are not analysed themselves, but only used for the instantiation of the given facts). This condition allows us to say that Prolog is used to automatically analyse real phonetic facts by taking nothing but categorical representations from databases as input and so to compute new phonetic knowledge.

(III)
THE AREA OF EXTENSIONAL SPEECH THEORIES: FROM SYMBOL TO SIGNAL, FROM SIGNAL TO SYMBOL

The values of Prolog variables and predicates are not restricted to symbolically represented phonetic facts, but can also be extended to relate the possible categories and the analysable physical properties of the speech signal to each other, in both directions. There is of course no analytical relation between symbols that represent categorical facts and speech signals which *per se* are nothing but digitized time functions. In a speech database we empirically identify both sides according to Feigl's principle [15,6]: any complex category can be experimentally reproduced by repeating a given time function, and the time functions that we record when a category is repeatedly demonstrated (by different speakers in different situations) will be again instantiations of that category. Thus we can look for *aposteriorily necessary* connections (in the sense of Kripke's (1980) analysis) in order to answer the question as to what properties of the signal cause the perception of a category, and which properties are to be expected if the category is reproduced within a speech act. This will ultimately allow us to define the physical extensions of the phonetic categories of a spoken language.

In the PhonDat-project it has been demonstrated by Florian Schiel that it is a good step in this direction to compute the phonetic facts on a yet unspecified 3rd CRIL-level, by using the information of the 2nd canonic level as input to a speech verification procedure [19].

However, the final aim of this third area of C.P. is to be able to compute any speech signal that falls in the extension of a given category, and to determine the phonetic form that the words of a spoken language segmentally and prosodically take as soon as they are used by the speaker of this language in connected speech.

REFERENCES

[1] Carré, R. and M Mrayati (1990), *Articulatory-acoustic-phonetic relations and model-ling, regions and modes.*, pp. 211-240 in: W. Hardcast-le and A. Marchal (eds.), Speech pro-duction and speech modelling. Dordrecht.

[2] Draxler, C. (1995), *Intro-duction to the Verbmobil-PhonDat Database of Spoken German*, Prolog Applications Conference, Paris.

[3] Eisen, B., H. G. Tillmann, and C. Draxler(1992), *Consistency of judgements in manually labelling of phonetic segments: The distinction between clear and unclear cases,* ICSLP Banff pp. 871-875.

[4] Esling, J. (1990), *Computer-coding of the IPA: Supplemen-tary Report*, JIPA 20, pp.

[5] Fant, G. (196l), *Acoustic theory of speech production*, The Hague: Mouton & Co.

[6] Feigl, H. (1958), *The 'mental' and the 'physical'*, pp.370-497 in Feigl et al.(eds): Minnesota studies in the philosophy of science, Vol. II, Minneapolis.

[7] Kohler, K (1990), *Segmental reduction in connected speech in German: phonological facts and phonetic explanations*, pp. 69-92 in: W. Hardcast-le and A. Marchal (eds.), Speech production and speech modelling. Dordrecht.

[8] Kripke, S. A. (1980), *Naming and necessity*, Cambridge: Harvard University Press

[9] Lindblom, B. (1990), *Explaining phonetic variation: A scetch of the H and H theory* pp. 403-439 in: W. Hardcastle and A. Marchal (eds.), Speech production and speech modelling. Dordrecht.

[10] Meyer-Eppler, W., und G. Ungeheuer (1957), *Die Vokalartikulation als Eigenwertproblem,* Ztschr. f. Phonetik 10, pp. 245-257.

[11] Moore, R.: *Twenty things we still don't know about speech*, pp. 9-7 in H. Niemann, de Mori and Hanrieder (eds.): Progress and prospects of speech research and technology, Sankt Augustin 1994.

[12] Rabiner, L. R., and R. W. Schafer (1978), *Digital proces-sing of speech signals*, Englewood Cliffs: Prentice-Hall.

[13] Schiel, F. (1995), *An automatic segmentation program based on HMM*, internal report, to appear in FIPKM.

[14] Schroeder, M. (1967), *Determination of the geometry of the human vocal tract by acousic measurements*. Jasa 41, pp.1002-1010.

[15] Tillmann, H. G., (1993), *Why articulation matters in SLP*, FIPKM 31, pp. 11-28.

[16] Tillmann, H. G., (1995), *Kleine Phonetik und Grosse Phonetik*, to appear in Kohler-festschrift, Phonetica.

[17] Ungeheuer, G.(1957), *Untersuchungen zur Vokalartikula-tion*, Phil. Diss., Bonn.

[18] Ungeheuer, G. (1962), *Ele-mente einer akustischen Theo-rie der Vokalartikulation*, Ber-lin: Springer.

[19] Wesenick, B., and F. Schiel (1994), *Applying speech verification to a large database of German to optain a statistical survey about rules of pronunciation,* pp. 279-282, ICSLP, Yokohama.