

## Why current speech technology is false phonetics

"The test of a model is its performance in its intended applications" (G. Fant)

Sven Öhman

Dept. of Linguistics, Uppsala university  
Sven.Öhman@ling.uu.se

### Abstract

*Some of the best known physical-acoustic models of speech production that have been proposed by Speech Technologists such as G. Fant (Mouton 1960) at KTH in Stockholm, Sweden and by K. N. Stevens (MIT Press 2000) at MIT in the US are discussed in this paper in light of the principle which has been chosen as its motto, a principle I attribute to Fant, since his very concrete and practical guidance as my thesis advisor once made me aware of its fundamental importance in modern Science and Technology.*

*I argue that the models in question are **not** adequate as intellectual tools for classical phonetics. As such, they are in fact inferior to the tools already available to phoneticians, but which they have been taken by many to replace. I try to show that this last interpretation of the new models is due to misunderstandings.*

### 1 Background

During the Second World War the American linguist Martin Joos made his military service on a submarine in the US Navy where he was trained to use special technical equipment that had been developed for the purpose of detecting enemy submarines on the basis of the under-water sound they emitted. Much owing to Joos' efforts this equipment was later made commercially available to linguistics and phonetics researchers by the Kay Electric Corporation as the so-called SonaGraph.

In 1948 Joos published the results of his phonetic war time experiments as a supplement to the international Journal *Language* under the title *Acoustic Phonetics*. The most impressive of these results was Joos' demonstration, inspired by the writings of the German physicist and physiologist H. v. Helmholtz' (1821 – 1894), that the *Vowel Triangle* of Classical Phonetics could be approximately reproduced by a two-dimensional plot of the vowel formants F1 against F2. (This "triangle" is of course best known today as the vowel "Quadrangle" of the International Phonetics Association, IPA, for vowel identification in field research and elsewhere.) Many linguists – including Joos himself – saw, in this correlation, a promise for phonetics to actually become the *exact science* that it had striven to be in the final two decades of the 19<sup>th</sup> century.

A few years later Potter, Kopp and Green at Bell Labs published a book containing a collection of Sonagraphized speech samples with the title *Visible Speech* (1947 N. Y. USA) which also happened to be the name of a speech transcription system that had been worked out in the 1880's by the famous Scottish-American phonetician and inventor (of the telephone) Alexander Graham Bell whose son had also given his name to the Bell Telephone Company (later part of A. T & T.). The impetus that this gave to the then brand new engineering research field of Acoustic Speech Studies thus had an orientation toward Phonetics from the very beginning. For instance, the French phonetician Pierre

Delattre was able to synthesize vowels and even the stop consonants p, t and k by means of a – by modern standards – rather primitive speech synthesis device at the Haskins Laboratories, then in New York. This orientation toward phonetics, especially Joos' "acoustic" variety of it, was challenged in the Monumental *Acoustic Theory of Speech Production* published by G. Fant in 1960. Here Fant, using mathematical-acoustic theories that had been developed at MIT when Fant was a student there, argued that the "formant-cavity affiliations" assumed by Joos, which had of course been the key to his and several other people's claim that the lowest-frequency formant (F1) reflected the "big cavity" in vowel production, and the higher-frequency one the "small cavity", were not justified when a model derived by a more detailed physical analysis was used. In fact by this model the so-called "vowel spectrum" consisted of an *infinity* of formant frequencies none of which was more "affiliated" with any one cavity than with any of the other cavities!

In the eyes of many linguists, this meant a death blow to Acoustic Phonetics as a method of interest to the Science of Linguistics (and Phonetics). To them it became a pure engineering concern, and it occurred to few critics that Fant's results need not be interpreted in this way. A more reasonable conclusion is that Fant's results left Phonetics unaffected by physical acoustics, and that Joos' somewhat amateurish choice of model was Fant's primary target. Yet many people – among them Roman Jakobson and Morris Halle, and later, also K. N. Stevens – concluded that further progress in phonetic research would be nearly impossible without building on the most recent engineering approach.

The submissive attitude on the part of linguists and phoneticians to the new approaches should be seen in the light of the ideological climate at that time.

The anticommunist currents in the Western World including Western Europe (with Sweden) reached its climax in the Cold War. The US, as the leading military and technological power of the world, started to take over the role as "World Police" which had been the prerogative of Great Britain before the war. The US defined itself as the power whose duty it was to make the world "safe for democracy". Antidemocratic movements anywhere around the world were considered to threaten American national security. Therefore the US started to invest huge sums of money into armaments, as well as science and technology, not only at home, but also abroad, e. g. in Sweden.

It is my impression that the linguists of the world were so overwhelmed by this quite new situation of Wealth for Physics and Technology and relative poverty for the Humanities and Social Sciences that they were struck by what might be called "technoplegia", a condition in which a prostrated attitude becomes natural! This condition actually still prevails in many places. One of its symptoms was the attempt by Chomsky and his followers to rewrite linguistics as an earlier unheard-of "Natural Science" working with quasi-mathematical methods which included a "generative phonology", and which was *also* to "replace" classical phonetics. Classical Phonetics was beginning to be overtaken by *two* "scientific" competitors – Technological Speech Research and Generative Phonology! The most recent sign of this movement is K. N. Stevens' *Acoustic Phonetics*, at MIT Press 2000 in which the influence from Generative Phonology is evident on almost every page.

#### 1.1 The intended applications of classical phonetics models

Referring back to the motto concerning model evaluations – "The test of a model is its performance in its intended applications" – we must consider the applications intended for the classical vowel quadrangle and other ideas of articulatory and auditory, i. e., Classical Phonetics. They were meant as tools that individual language students should be able to use "in the field" i.e. in study situations in which one encountered events of sound

production in the course of speech in some language. The student was not supposed to have to rely on aid by any particular tools or technical gadget of any kind. The language student's only tools for "speech synthesis" and "analysis" were supposed to be her or his own mouth and ears, respectively. To write down the speech sounds that were heard in the scientific phonetic notation which had been mastered by means of the student's acquiring control over his or her articulatory abilities and hearing sense. What was needed was some training in the rendering of the positionings of one's own articulators according to the "traditional articulatory maps" worked out, and still being worked on, in the science of Phonetics. The "fidelity criterion" was that another language student with sufficient familiarity with phonetics was herself able orally to reproduce the sounds noted down in a way that could satisfy native speakers of the language in question. This was essentially the idea of it all.

In so far as it is not an anachronism to talk about "methods" in Classical Phonetics, it would perhaps not be unfair to describe them as being thoroughly *subjective*, and hence basically incompatible with those of Modern Science and Technology. In particular the fundamental *concepts* of articulatory-auditory phonetics should be understood as certain human abilities to find one's way about in the articulatory-auditory space of speech sound production, both so as to identify the sounds and to reproduce them. One acquires these concepts through relatively extensive experimentation with one's own sound production abilities, keeping an eye on the phonetic relationships between the various sounds one encounters. Since we are not aware of hearing (or feeling) formants, or anything else that speech physicists describe in the mathematical language of Speech Technology, and since there is no method available to bring them within the grasp of the naked Human senses, we have no way of incorporating them into the linguistically relevant conceptual order of Classical Phonetics. Thus, the basic concepts of articulatory-auditory phonetics can not be *translated* into physical-technical concepts, i. e. into explicit physical-acoustic conceptual constructions, without serious distortion. The translations become *different concepts*, which cannot be used to *replace* their Phonetic originals with a hope of achieving the *same goals*. For, they are not within their reach. That this is so is well illustrated by the example of the "visible speech" of the SonaGraph. To this day no one has been able to read it fluently! If "translation" is attempted then, *nolens volens*, a *new* conceptual order is created whose relationship to its source of inspiration is necessarily obscure. So obscure in fact that it is highly doubtful whether the translation should at all be called *phonetics*. Stevens' new book "Acoustic Phonetics" ought rather to have been called *Phonetic Acoustics* since it is much more of a book in acoustics than in phonetics!

## 2 Future for phonetics?

I do not mean to say, of course, that Classical Phonetics is beyond change or that it cannot be improved upon. Nor do I mean to suggest that developments in modern Science and Technology must necessarily be without value to language students and phoneticians. Quite on the contrary I believe that the impressive progress in electronics, sound engineering and computer technology that we have witnessed in the last few decades provide linguists with several excellent ways of simplifying language study and making it more efficient. However, language study is done by human beings, not by computers. And in order for the computer itself *not* to become a problem over and above those posed by the language studies, the technology must be adapted to the demands posed by the student's linguistic tasks rather than the other way around; with the student having to adapt to the technical and conceptual demands of the equipment. This is however just what current "Acoustic Phonetics" and Generative Phonology require by forcing the

linguist and phonetician to try to master an utterly non-intuitive translation of relatively straight-forward concepts into the quasi-scientific language of Generative Phonology, and the scientific language of Physical Acoustics.

The Acoustic Phonetics of Martin Joos was in fact much more adequate to the purposes of phonetic language study than the version now proposed by K. N. Stevens, because it could be understood in terms of the well established ideas of a long phonetic tradition whose performance in the intended applications of language study had been thoroughly tested.

## 3 Work in progress

I am in the process of making a thorough investigation of the larger part of the conceptual framework that has been in the focus of this brief paper. I hope to be able to report on this work (in English) in not too distant a future.

## References

- M. Joos. 1948. *Acoustic Phonetics* Language Nr 24, Supplement
- Hermann von Helmholtz. 1954. *On the Sensations of Tone*, Dover N. Y.
- G. Fant. 1960. *Acoustic Theory of Speech Production*, Mouton.
- K. N. Stevens. 2000. *Acoustic Phonetics*, MIT Press.
- S. E. G. Öhman <http://www.ling.uu.se/Sven> Publications found under *Curriculum Vitae*