V.Galunov

THE ACTUAL PROBLEMS OF THE SPEECH ACOUSTICS

Saint-Petersburg State University
quay Universitetskaya 9/11
AudiTech Ltd.
Russia, 195220 Grazhdansky 22
Tel./fax: 7 (812) 535-95-86
e-mail: auditech@online.ru

Three possible approaches to the problem of speech acoustics are under consideration:

1. *Starting from concrete practical tasks of automatic speech recognition and synthesis;*
2. *Guided by the classical theory of speech production;*
3. *Considering speech as a system providing it's principal distinctive features such as production and noise immunity.*

There are three possible approaches arising in connection with the speech research.

The first one, which is the most popular recently, proceeds from the priority of solving applied tasks. Popularity of such an approach is boosted by the active development of so called speech technologies, i.e. the systems of automatic speech recognition, synthesis and compression. But for all that the natural-science knowledge on the object under research is substituted for utilitarian purposes. The example of contraposition of a plane and a bird constitutes a standard justification of such an approach: a plane doesn't flap with its wings, but it solves its tasks better than a bird. Of course, among the tasks of speech technologies there are some narrow problems being solved without any accurate knowledge on the speech structure. And the corresponding utilitarian approach has the right to existence but, of course, is not related to what is called science.

Besides we should bear in mind that the most perspective fields for the speech-technology applications are connected with interaction with man and demand at bottom of fact repetition of the methods of speech information processing used by man in the technology system. It's possible to give some examples, rather effective but taking science away form true knowledge. As for the field of automatic recognition, the use of latent Markov models exemplify this. It is obvious that speech communication is organized by some rather complicated rules on which we don't know everything or, better to say, know very few. The use of probability model (HMM) it's an attempt not to study these rules but to substitute them with a random search. Imperfection of the method is obvious: absence of noise-immunity, need of a great statistics. Within the last fifteen years compilation method (e.g. SPOLA) became the most frequently used in the speech synthesis. Creating of the systems like this they practically fully refrain from the reveal of the speech signal acoustic structure and scientific investigations are focused only on the linguo-phonetic levels. At present speech compression is practically focused on the methods of head-on approximation of the speech signal without taking its acoustic-informational structure into consideration. Such methods allow to reach the signal rate of 2400-1200 baud, of course, with loss of noise-immunity and some other speech signal characteristics. At the same time it should be mentioned that it was already in the middle sixties in Russia when the attempts to apply in practice the special-application compression systems for 600-800 baud. With all this rather primitive models of the speech signal information structure were taken into account (but still not the direct signal approximation), without taking the speech signal specificity into consideration.

The second possible approach corresponds to what is now accepted to call the acoustic theory of speech production. It was Helmholtz who laid the foundation of this theory [1].

The main idea of this approach is the following assumption: if we manage to deal with the acoustic mechanisms of speech production, we'll be able to understand the speech signal structure. In compliance with Helmholtz' ideology speech is being formed by means of speech apparatus resonance excitation by the sound sources (voice, noise or impact) which are independent from corresponding resonance characteristics. This theory took more or less modern shape in the middle of the last century [2, 3, 4]. There are some alternative versions in corresponding models but the following statements are considered more or less common:

- speech signal consists of sound, linear sequence of the speech segments, corresponding to phonetic structure of the speech message;
- the main parameter distinguishing linguistic units are the formants which now are regarded as the peaks in the speech signal dynamic spectrum (in the initial model – resonance frequencies of articulation filter) and the sound source characteristics;
- the work of the sound source and that of articulation filter are independent.

The ideology of the acoustic theory of speech production in the form above-mentioned is implicitly accepted as canonical by almost all speech investigators though the facts discordant with this theory are known even to the most insignificant of them.
One can mention the following:

1. the idea of rough linear phonetic structure was introduced to the theory of acoustic speech from the phonetically-oriented written language customary in many countries. Indeed a man can convert an acoustic speech signal into a linear phoneme sequence and vice versa. But we should keep in mind that the linguistic categories (of the phonemes) are abstract and not required to have physical correlates.
2. The vowels, which acoustic correlates well with customary canons, contain next to nothing useful data on the speech message though they have much more energy than consonants which are exactly the main information carriers.
3. There is a range of speech transmission systems in which the formants in no way could be discovered: band vocoder, telephone line (especially with carbon microphone, not in use today), etc.
4. A number of experiments points out to the fact that transitions between phonemes are the most important in the speech transmission.

This list of vagueness can be followed. Retaining the basic ideas of the acoustic theory of speech production it still has several running problems.

1. Independence of the sound signal and articulation filter has no experimental proof.

Even for such a well-simulated sound source [5] as a voice source it turns out that the linguistic quality of the speech has been considerably formed already at the larynx output [6]. If it is true, the statement on the leading role of the articulation filter in forming parametrical form of phonetic elements becomes rather doubtful.

2. Besides the leading role of the formants, the alternative statement on determining role of correlation of the power levels in certain spectrum zones was put in the forefront [7]. In this case the formants are not the main speech parameter but only a way of realization of absolutely different kind of speech characteristics, and the mechanisms of speech perception become leading in the formation of the speech acoustic structure rather than the process of speech production.

3. consideration of the acoustic processes in speech production fundamentally distinct from the model by Helmholtz – Fant, for example the modulation model [8].

The third approach to the speech examination can be started with A. Einstein's statement: "Does it make sense to describe Beethoven's symphony in terms of air-pressure waves?" If we are able to understand the order of the process of speech communication, only then we'll be able to raise clearly specific acoustic problems. One may suppose the three components participating in the process of speech communication: the source of data, communication channel, the receiver. One possible model of organization is represented in the papers [9, 10]. There are two considerable factors to be pointed out which influence the speech signal organization. The first one is a need of noise-immunity ensuring. N. Vinner: "Speech is a mutual game of the speaker and the listener against the forces causing disorder". Speech message should be redundant only by direct multiplication of distinctive signs and provide the speech communication functioning under different kind of external and internal noises. It follows from this that speech messages should simultaneously posses discriminate parameters defining and providing transmission of the same language data. On the other hand, obvious redundancy of acoustic and parametric provision of the speech message makes the complete parametric provision of the speech message mandatory in each specific speech communication act.

At present quite a few is known on the ways of speech noise immunity provision [7, 9, 11]. Thus in each specific situation of speech contact it's almost clear what can be given up in the structure of the speech message.

In conclusion we should once again pay attention to the fact that the speech in a real acoustic communication and written speech are fundamentally different processes. At the same time at present the most vague question of the speech communication is the two-fold question concerning the way of temporal organization of the speech signal:

- does the linear format realize substantively in the speech signal (much as it happens in written speech for phonetically oriented kinds of written languages);
- where the considerable behaviourally useful information is concentrated in the speech acoustic signal.

R E F E R E N C E S.

1. Heimholtz H., Die Lehre von der Tonempfindungen als psychologische Grundlage fuer die Theorie der Musik. Brannschweig, 1870.
2. Chiba T., Kajiama M. The Vowel, its nature and structure, Tokio, 1941.
3. Fant G. Acoustic Theory of Speech Production. M., Nauka, 1964.
4. Ungeheuer G., Elemente einer akustischen Theorie der Vokalartikulation, Springer-Verlag, Berlin, 1962.
5. Galunov V., Tampel I., Mechanical Works of Speech Source. Acoustic journal, XXVII, # 3, 1991.
6. Galunov V., Krylova B., Stankevich S., Kantemirov R. The Research of Aerodynamic Processes in Larynx. The III Otolaryngologic congress of the Russian Federation, 1972
7. Galunov V., Garbaruk V. «Acoustic Theory of Speech Production and The System of Phonetic Signs». 100 Years of Experimental Phonetics in Russia. Materials of the international conference. St. Petersburg, 2001, 58-60.
8. Galunov V., Uvarov V., Once More on the Mechanism of Voice Production. Materials of the XI session of the Russian acoustic society, v. 3, 2001
9. Galunov V., Noise Immunity as the System-organizing Factor of Speech. The Problems and the Methods of Experimantal and Phonetical Research. St.Petersburg, 2002
10. Galunov V., Speech as a System. The present collected articles.
11. Galunov V., Koroleva I. «Provision of noise-immunity when processing information in auditory system». Sensory Systems, v. 2 № 2, 1988