WAVE MECHANICS OF THE SPEECH SIGNAL

VICTOR SOROKIN

Institute for Information Transmission Problems
Academy of Sciences, USSR
Moscow 101447

ABSTRACT

Processes of control, muscular contraction, articulators deformations, and acoustical oscillations take place in continuous media. Description of these processes can be done by means of the same mathematical technique. Some important properties of the speech production processes are described.

Speech production processes proceed in different physical media : articulation control processes - on a set of α and γ motor neurons and muscle fibers, mechanical oscillation processes of the vocal folds, tongue and lips - in viscous - elastic tissues, acoustical processes - in air cavities of the vocal tract. There are certain special features in each medium which determine characteristics of the processes, but there is also a very important similarity that enables to use actually the same mathematical technique of wave mechanics. This similarity comes from the fact that on each level - acoustical, mechanical and control, the system of speech production is a system with distributed parameters.

Motor units in the articulation control system are discrete elements but their number is great, and their parameters overlap rather a wide range. Thus it is possible to consider that processes of muscle contraction control take place in a certain continious medium. In that medium, for instance, there is a possibility to control dynamic characteristics of the muscle by the working point shift of the muscular receptors and also by displacement of the area of the sarcomers contraction along the muscular fibers from one muscle end to another.

Amplitudes of eigenfunction of muscle elastic deformations change due to these displacements. Although distributed control systems have been studied in engineering, special features of the articulation control system are little known and deep investigations in the area are of great necessity. Some results concerning articulation control system properties are described in [6].

Geometrical parameters of the vocal folds are rather small. Thus folds elastic vibrations accompany waves in all three dimensions. It can be well seen in high speed motion pictures that elastic waves propagate along the focal folds and also in transverse and vertical directions. Besides, surface waves are observed after folds collapse [2]. Characteristics of the waves are determined by mechanical properties of the vocal folds tissues. For example, surface waves dissapear when cancer tumour evolves.

Mathematical models of vocal folds elastic vibrations were investigated in [4,5,6,7,9,10]. Computer modeling shows, that for the description of folds elastic oscillations in a vertical direction only one eigenfunction is sufficient, in a transversal direction two eigenfunctions are required, and oscillations along the vocal folds require three eigenfunctions. Vertical movements create a new, unknown before, excitation source - a piston source, which is active during close vocal slit interval also. A speech synthsizer excited by the described vocal source produces a speech signal with high naturalness.

Geometrical sizes of the tongue and lips are comparatively large. Thus elastic waves do not propagate in the articulators but their movements are "wave-like", as it is seen in the cinemaradiographic motion pictures. Elastic deformations of the tongue and lips are described by the same mathematical technique as elastic deformations of the vocal folds.

It is sufficient to have only one eigenfunction for description of the shape and movements of the lip. This eigenfunction for the lip is just half-wave of sine. Five eigenfunctions must be used for the tongue shape description. In the case the approximation error in the uniform metrics is about 6 - 7 %. Change of the tongue shape is achieved by means of modes control of elastic oscillations. [6]. The velum is an elastic body with distributed parameters too, thus to calculate its deformations the same mathematical technique as for the vocal folds, tongue and lips must be used.

For the frequencies above 200 - 300 Hz the vocal tract is a system with distributed parameters where waves of acoustical oscillations propagate. Fast change of speech parameters and nonstationary processes domination are properties of acoustics in the vocal tract. There is a set of pecularities of speech production acoustical processes which play crucial role both for speech synthesis and speech recognition.

First of all one must reject an idea that the vocal slit is a starting point of the vocal tract. During production of voiceless fricatives and stops the area of the vocal slit is comparable to the minimal area of the vocal tract. Even during phonation the resistance of the vocal slit turns out to be not so high as it was supposed just recently. As a result, processes in the lungs and the trachea have influence upon acoustical parameters of speech signal and, therefore, the vocal slit is located almost in the middle of the vocal tract. During phonation formant frequencies are subjected to noticeable alterations, sometimes up to 20 - 30 % and those alternations are synchronous with vocal folds oscillations [6] . The formant frequencies variations are pretty fast, therefore both in speech synthesis and speech analysis the vocal tract must be considered as a parametric system with fast alternation of parameters.

In addition to those variations the formant frequencies sometimes undergo fast alternations during articulators movements. For example, the rate of the first formant freqiency variation can be 10 times as much as the velocity of articulator movements, if the minimal area of the vocal tract cross-section is sufficiently small. As a result an abrupt change of formant frequencies is observed before and after a closure.

Vocal tract walls yielding is a cause of radial oscillations, which dominate during closure. The first formant frequency in this interval is equal to the radial resonance frequency (150 - 350 Hz) instead of zero as it were in an acoustical system with absolutely rigid walls. There is radiation of a speech signal through the yielding vocal tract walls and, as it was shown in [1] , the radiation occurs mainly in two areas - around the lips and the pharynx. Yielding of the vocal tract walls leads to the "shut" effect, when in areas with a small cross-section the propagation of low-frequency oscillations stops due to walls oscillations in the antiphase.

As a result of all above mentioned facts the only right method of description of speech production acoustics is a method nonsteady-state wave processes. Therefore the formant method of speech synthesis is unadequate in consequence of the hypothesis on steady stateness of processes and slow variations of the vocal tract parameters. This is confirmed by the low quality of formant speech synthesizers. More over, there is a limit (not very high) for improvement of such synthesis naturalness and intelligibility.

As a basis for description of acoustical processes during speech production a method of travelling waves like Kelly - Lochbaum scheme [3] can be used. Originally Kelly - Lochbaum scheme has a set of serios shortcomings. Particularly, it generates specific noise during alternation of the vocal tract area function. However it is possible to solve the noise problem by means of dynamic matching of boundary conditions between cylindrical sections which approximate the vocal tract shape [7,8] . Further improvement of that scheme should concern section and vocal tract length alternation.

Characteristics of the turbulent source of excitation are little dependent on the place of articulation. Difference in acoustical characteristics of fricatives and bursts of stops are results of various positions of the turbulent source in the vocal tract accompanied with change in values of zeroes of the vocal tract transfer function. This effect is also a consequence of the fact that the vocal tract is a system with distributed parameters.

Thus, physics of speech production processes is much more complicated then it was supposed some time ago, but cognition of those processes inspires a hope of development of high quality synthesizers and reliable speech recognition systems.

1. Fant G., Nord L., Branderud P. A note onthe vocal tract wall impedance. STL QPSR, 1976, N 4, p. 13-20.
2. Hirano M. Data from high speed motion pictures. Vocal Fold Physiology Conf. Kurume, 1980, p. 4a1-4a6.
3. Kelly J.L., Lochbaum C.C. Speech synthesis. Proc. 2 Int. Congr. Ac., 1962, paper G42, p. 1-4.
4. Sorokin V.N. Some questions of the general model of speech production generation. Proc. ARSO-8, Lvov, 1974, p. 97-100.
5. Sorokin V.N. The voice source as a system with distributed parameters. Acoust. J., 1981, v.27, N 3, p. 434-440.
6. Sorokin V.N. The speech production theory, 1985.
7. Sorokin V.N. Travelling waves in the vocal tract. Acoust. J.,1986, N 4, p. 506-510.
8. Strube H.W. Time-varying wave digital filters for modeling analog system. IEEE Trans. v.ASSP-30,1982, N6, p.864-868.
9. Titze J.R., Strong W.J. Normal modes in vocal cords tissues. JASA, 1975, v. 57, N 3, p. 736-744.
10. Titze J.R. On the mechanisms of the vocal-folds vibration. JASA, 1976, N 6, p. 1366-1380.