CONCEPT APPROACH OF THE HUMAN-COMPUTER IN VOICE PROCESSING FIELD

Anne-Marie Claudia DUMBRAVĂ
dumnravaannemarie@yahoo.com

Abstract. This paper aims to study the way to produce speech through mathematical analysis and graphical representation of the fundamental tone period and the voice tract parameters corresponding to certain typical situations regarding ways of pronunciation and accent. It is a model of human speech generation in terms of change the signal transfer function that produces speech. The fonetic mechanism which produces the sounds is exemplified by covering operation of fricative and plosive sounds.

Keywords:

1. HUMAN SPEECH GENERATION MODEL - FEATURES, DESIGN AND RESULTS

Speech production process can be modelled to achieve technically artificial speech.

Noise excitation signal en(n) with period T0 and noise en(n) are chosen by switch vocalize/non-vocalize KV/N.

With a level adjustment provided by the gain G, the level excitation is applied to a system that changes over time and transfer function H(z).

Resulting the signal spoken s(n) = h(n)* e(n) the spectral variation: S(z)=H(z)E(z)

The transfer function H(z). It follows the signal spoken s(n)=h(n)*e(n) with the spectral variation: S(z)=H(z)E(z)

Fig.1. The speech generation
The transfer function $H(z)$ can have poles and zeros in the complex plane whose positions change over time. It can be achieved by the band pass filters in cascade (synthesis formantic), by systems with only poly (synthesis by linear prediction), from cascades formed by electric models of the acoustic cylindrical tubes of different sections.

![Cylindrical acoustic tube modeling for different sections](image)

**Fig. 2.** Cylindrical acoustic tube modeling for different sections

The human auditory model for perception of sound – In the perception of sound problem the basic questions which arise are related to how sounds are perceived by humans, what a person hear and how different components of the sound interfere one with another other.

If a sound can be heard or not, this depends on the strength and range of sound (perceptibility is discussed in terms of hearing thresholds) (Zost, 1994). What a person hears at given sound response is a more complicated problem and at this level the discussion is limited of how structure of sound relates to the perceived sound frequency.

Once the speech the signal is generated and propagated to the listener begins the perception process (recognition).

The spoken signal is received by the external ear, acoustical transformed in the middle ear and processed in the inner ear along the basilar membrane, resulting a spectral analysis of it.

The cochlea is the first place where the time-frequency analysis in the auditory system.

The sound entering the ear determines the appearance of vibration, that propagate along the basilar membrane (MB) and reach a maximum at a certain
point, which depends on the frequency of the excitatory tone.

The mechanical vibrations are felt by the nervous cells, leading to activation of the nerve fibers.

Each of these nervous cells is connected to about 10 nerve fibers of different diameters and at different synaptic connections.

In a manner still not understood by humans, neural activity along the auditory nerve is converted into a code language for the highest centers of processing in the human brain where is performed understanding of the message.

The cochlear processes of the vibration MB and of the neural activity are quite nonlinear, and as a result, the perception of sound energy at a frequency is dependent on the distribution of the sound energy on another frequency.

Auditory threshold - The ear is able to hear sounds in a wide range of frequencies from 16Hz to 18kHz. Frequency sounds below 1kHz or over 5kHz demand a higher energy to be heard than those in (1-5)kHz.

The intensity of the sound is measured in terms of sound pressure level (SPL) in decibels. The reference level SPL is an intensity of 1026 watt/cm2 at a frequency of 1 kHz, which corresponds to dB. At the extremes of the audible frequencies, sounds can be perceived only in a narrow field of amplitude, so at the frequencies where the ear is more sensitive from 1kHz to 5 kHz, sounds are detectable in a dynamic range exceeding 100 dB. The minimum intensity at which sounds can be perceived is known as the hearing threshold or auditory threshold, which is growing rapidly outside the above. The auditory threshold remains almost constant on a large range of frequencies. The most relevant threshold for frequencies below 700Hz, which is characteristic of the first formant region and of the fundamental tone and of the most powerful harmonies.

On average, the perception threshold of the first formant is 10dB higher than that of the other two. For the vowels, at physiological level, all harmonics up to the fourth formant can be heard. However, when the speech amplitude is reduced, as if speaking in a whisper, it is likely that the fundamental and the first harmonic should not be perceived. These frequencies are not the crucial for intelligibility. An example is talking on the phone, where the signal band begins after the 300Hz signal band and although the loss is irrelevant in terms of intelligibility, the naturalness signal is affected.

2. THE FONETIC MECHANISM – INTERFACE FOR ORAL MAN-MACHINE

Fonetic mechanism - Speech is a product of the human vocal tract, controlled by the central nervous system which continuously receives information by the acoustic reaction (auditory reception of the signal). The human vocal tract is no special apparatus for this purpose, but is achieved by adapting to the function of some organs belonging mainly to the masticator and respiratory equipment. Therefore the production of speech (phonation) is a voluntary act and coordinated of the breathing and masticator apparatus.
Respiratory apparatus provides the energy required for speaking, when air is exhaled through the trachea. At the top of the trachea is the larynx, where the air pressure is modulated before being applied to the canal (tract) tag that extends from larynx at lips (fig. 3).

![Diagram of respiratory and vocal tract]

**Fig. 3.** Scheme of the respiratory and vocal tract

Larynx is a set of muscles and cartilage surrounding a mobile located at the upper cavity of the trachea. Larynx is a set of muscles and mobile cartilage surrounding a located at the upper cavity of the trachea.

In the center of the larynx are the vocal cords, placed across the larynx. Vocal cords are ligaments formations symmetrical in number of two pairs, one lower and one higher. Between the two pairs of string vowels, lower ones have a much higher elasticity and are essential in producing of speech. They can close completely the larynx, and by removing they form a triangular opening, called throat.

The brain system control through nerves the movement the vocal cord, putting them into vibration through contractions.

In this way it can increase or decrease the size of the glottis, achieving a modular flow of it.

Free air passes during breathing and whispered voice, and also during training the muffled sounds or the non-sound ones (speech occurs during expiration air). Audio sounds are produced, on the contrary, through the regular vocal cords, which makes the vocal tract and be applied to periodic pressure pulses.

Vocal tract (vocal channel) is a set of cavities located between the throat and lips.

- **Pharyngeal cavity**;
- **The oral cavity**;
- **Nasal cavity** (in derivation with the the oral cavity).

Vocal tract can be considered as a series of tubes or acoustic cavity lengths.
and different sections. The radiation sound (sound emission) can occur through the oral opening (mouth) or simultaneously through the opening oral and nasal orifice (nostrils).

Derivation of airflow to the nasal cavity is controlled by the soft palate (the moving part of the palate).

Audio sounds are produced therefore by the vocal tract excitation by periodic pressure pulses generated by the vibration of vocal cord: sudden opening of the glottis upstream the release accumulated pressure, glottis closes after a while longer.

At the top of the larynx (above the upper opening thereof) is the epiglottis.

This is a triangular fibro cartilage cell below the language, serving to covert the top of the throat when swallowing, thus separating the food from the respiratory route during swallowing.

Oral vowels (i, e, u, ...) are issued without the intervention of the nasal cavity, which is isolated by closing the soft palate. Nasal vowels and consonants are produced with the involvement of the nasal cavity.

The coupling between glottis and vocal tract is a weak coupling in that tract deformation affects less pressure wave generated at the glottis opening.

The sound intensity is closely linked and driven by air pressure upstream of the larynx. The sound height is determined by the vibration frequency of vocal cord, called the fundamental frequency (F0).

The inverse of this frequency, denoted by T0, is called the fundamental period. Since F0 height determines the sound, it is also called pitch (the English term, adopted in French and Romanian).

Fundamental frequency may vary between the limits:
- 80-200Hz a male voice;
- 150-450Hz a female voice;
- 200-600Hz for a child's voice.

Two sounds with the same intensity and the same height are distinguished by timbre, which is determined by the relative amplitudes of the fundamental harmonics.

A sound audio signal corresponds to a quasiperiodic signal. It is observed spectral lines on the chart (radius) corresponding to F0 fundamental frequency harmonics.

This harmonic structure of the spectrum is called pitch structure.

The tire of these rays have maximum named Format and corresponding to their own frequencies Fi (i = 1,2,3, ...) of the vocal tract. The first three formats are essential in order to characterize the voice signal spectrum, higher order formants having a limited influence.
Fricative sounds are produced as a result of air passing through a narrow portion (constriction) of the vocal tract, situated in a sense of vocal tract, particularly the lips and teeth. Fricative sounds are classified into two categories.

- Sonorous (v,s);
- Non sonorous (f,s).

The plosive (occlusive) are produced by closing (occlusion) of the vocal tract at some point, followed by sudden relaxation.

Vocal chords remain idle, and air flow passes freely through the throat to the place in which the occlusion. Here air compression occurs during occlusion after the sudden opening of the tract causes the release of retained air, thus generating the sound respective with explosive effect.

Sounds sisters may be of two types:

- Sonorous (b, d);
- Non sonorous (p, t).

3. ARTIFICIAL INTELLIGENCE IN ROMANIA

The failure of Artificial Intelligence shows also the need not to disregard the biological sources of inspiration and to closely integrate research in AI with the other cognitive science and with neuroscience and cognitive psychology in particular.
In Romania it is relatively little research in Artificial Intelligence, and what is already exists is generally confined within traditional paradigms and in a formal approach, not interdisciplinary.

For example, the cognitive sciences are not mentioned at all, as area, into the government department responsible for research programs.

That while France, for example, encourages the development of the research in cognitive sciences through incentive actions and in a recent report by the National Science Foundation (U.S.) the cognitive science are identified as national priority in research together with the bio technologies, the nanotechnologies, informatics, mainly in their future perspective.

On the other hand, in Romania there is the potential of performing a competitive research in AI, because there are competences in computer science, and for the research in this area are not required expensive equipment in laboratory, as in other areas with high potential, the biotechnologies, the nanotechnologies.

Tendencies and future projections – Development of the Romanian technology (IT), which is a booming sector and recognized as a priority by the government, may have applications of research results in Artificial Intelligence, given that now is based on outsourcing 90%, much less profitable than software products. Implementation of a development strategy in Romania of research in artificial intelligence and cognitive science may have beneficial effects for long term storage of the international competitiveness of Romanian IT industry.

4. BIBLIOGRAPHY

