ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research (ASSEHR), volume 122
45th International Philological Conference (IPC 2016)

Исследование работы голосового источника и артикуляторного аппарата для гласных звуков на разных стадиях речепроизводства

Евдокимова Вера Вячеславовна[1] **Скрелин Павел Анатольевич**[2]
Евграфова Карина Владимировна[3]
[1]Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург, Россия
[2]Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург, Россия
[3]Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург, Россия

Investigating the source-filter interaction of the vocal tract on different stages of speech generation

Evdokimova Vera[1] **Skrelin Pavel**[2] **Evgrafova Karina**[3]
[1]Saint Petersburg State University (SPSU), St. Petersburg, Russia
[2]Saint Petersburg State University (SPSU), St.Petersburg, Russia
[3]Saint Petersburg State University (SPSU), St.Petersburg, Russia

Аннотация

Для получения экспериментального материала были выполнены синхронные записи речевого сигнала в разных местах речевого тракта человека: на выходе и в области голосовых связок диктора. Была проведена синхронная запись речевых посылок текста длительностью 15 минут для 6 дикторов: 3 мужчин и 3 женщин. Были записаны изолированные гласные на разной частоте основного тона, а также список слов. Проводился слуховой, акустический анализ и перцептивный эксперимент. Была построена передаточная функция артикуляторного аппарата, по которой производился синтез новых гласных звуков.

Abstract

The paper presents an original method of synchronous recording the glottal wave and an output speech signal. The method was employed to obtain the experimental material. The recordings of 3 male and 3 female subjects were performed. The recording of each subject was 15 minutes long and included isolated vowels and words. The auditory analysis, perceptual experiments and acoustic analysis were carried out. The transfer functions of the articulation for the Russian vowels were obtained. The transfer functions and voice source signals of different vowels were used to generate new signals. The obtained signals were analyzed. The constructed model of the filter part of the vocal tract completely corresponds to the basic phonetic laws. It adds the accuracy to the existing models of the speech production and can be used for solving specific problems of speech technologies.

Ключевые слова: фонетика, сигнал голосового источника, форманты, передаточная функция речевого тракта.

Keywords: phonetics, voice source, source-filterinteraction, formants, transfer function of vocal tract.

Introduction

The traditional approach to phonetic research of the vocal tract assumes that there are several successive stages of speech production which are initialization, phonation, articulation and radiance of speech signal. The fact of the presence of the interaction between the two parts of the vocal tract does not make the classic linear source-filter theory completely consistent. To obtain the voice source signal which is not influenced by the articulation system and analyze its nature is an important up-to-date problem for different fields of speech science and speech technology.

Our research was aimed at analyzing the signal of the voice source and the output speech signal to consider the non-linearity of the vocal tract system. The coprocessing of these signals allowed constructing the formant structure of the vowels, the frequency constituents of different kinds of vowels and their variations. The transfer functions of some vowels and the voice source signals of others were used to synthesize the new vowel sounds. The obtained speech signals were tested in the perception experiment.

Equipment and subjects

The recordings were made in the recording studio. Multichannel recording system Motu Traveler and WaveLab program were used. The recordings had a sample rate of 44100 Hz and a bitrate of 16 bits. Two microphones were used during the recording. The capacitor microphone AKG HSC200 was placed in the output of the speakers mouth (ME). The miniature microphone QueAudio (d=2.3 mm, waterproof) was located in the proximity of the speaker`s vocal folds (MI) with the use of special medical equipment. This procedure was performed by a phoniatrician. The subjects of the experiment were 3 male and 3 female speakers. Each speaker pronounced each of the 6 Russian vowels: /a/, /e/, /i/, / ɨ/, /o/ and /u/ in different pitch modes: comfort, high, low, rising and falling. Apart from the isolated vowels, the speakers were asked to read a set of words [1], [4].

Perceptual experiment

The aim of the experiment was to find out if a voice source signal could be identified as a speech sound and which Russian vowel it could be associated with. A group of informants (23 individuals) were involved into perceptual tests. The samples were organized on a random basis. The informants were asked to assign each stimulus to one of the six Russian vowel phonemes. The questionnaire had also no decision option. The tests results showed that the vowel [a] stayed most intelligible and were identified correctly in most cases. The vowels [e], [o] and [u] were second intelligible. However, there were strong confusions of [i] and [u], [i] and [ɨ] and [u] and [ɨ]. Besides, some informants reported that all vowel types were perceived as labialized [4].

Acoustic analysis

The analysis of the vowel spectra shows that the signal from MI contains the frequency constituents of the vowel formants (resonance frequencies of the set of pharynx, nasal and oral cavities) However, the frequency constituents are weakened. It can be assumed that it is caused by the refection of the acoustic energy from the articulation system upstream [2]. As well as this, the signals can be very different for the two microphones.

Thus, the reflected energy in the nonlinear acoustic system of the vocal tract has an influence on the operation of the voice source and the glottal wave characteristics. This energy is reflected again to the articulation system and its frequency constituents are changed.

The next step was the discrimination and modelling of the transfer functions of the articulation. The transfer functions and the formant positions were estimated using the algorithm described in the research by Evdokimova [3].

Vowel synthesis

The obtained transfer functions of the vowels were used to generate the new signals. The voice source signals of different vowels were the input for these transfer functions. Our aim was to find out whch of the following would influence the resulted signal more: the characteristics of the voice source and the feedback section of one vowel or the transfer function of the articulation system of the other one.

The resulted signals were presented to the 10 lay participants. The aim was to identify the resulted signals as one of the Russian vowel phonemes. The experiment showed that all of the signals can be recognized as a vowel speech signal. However, not all the vowels were identified as a vowel phoneme the transfer function of which was used for the generation. Sometimes the participants recognized the resulting signal as the input /a/ or /ɨ/ signals. The analysis of the results of the perceptual experiment showed that, if the formant characteristics of the feedback section in the MI input signal were closer to the vowel formants in the transfer function, the vowel was identified as a source signal. For example, the signal generated from the source MI signal /ɨ/ with the transfer function characteristics of /i/ was perceived as /ɨ/. The generated vowel from the source MI signal /a/ with the transfer function characteristics of /e/ was perceived as /a/.

As well as this, if the formant characteristics of the feedback section in the MI input signal differed significantly from the vowel formants in the transfer function of articulation, the vowel was identified as a vowel the transfer function of which was used. For example, the signal generated from the source MI signal /ɨ/ with the transfer function characteristics of the vowel /u/ was perceived as /u/. The signal generated from the source MI signal /a/ with the transfer function characteristics of /i/ was perceived as /i/.

Discussion and conclusions

The acoustic analysis and perception experiments allowed us to specify and improve the source-filter model. The results confirmed the fact that the acoustic energy from the filter component reflects backwards. This justified the introduction of the feedback component into the speech production model. The used approach allowed reliable automatic discriminating of the vowel formant structure by processing the real speech data.

The filtering the supraglottal signal by the transfer functions of other vowels showed that it was not always possible to synthesize a new vowel. The reflected signal of the feedback section was sometimes stronger and influenced the resulting signal more. In the perspective experiments we plan to investigate the influence of the fundamental frequency on the resulting signals. Besides, as the microphones used in the recording were different, there is a problem of normalization of the MI and ME signals.

The constructed model of the filter part of the vocal tract completely corresponds to the basic phonetic laws. It adds the accuracy to the existing models of the speech production and can be used for solving specific problems of speech technologies.

Acknowledgements

The authors would like to thank Saint-Petersburg State University. This work has been carried out in the framework of SPbSU project n. 31.37.353.2015 “The Phonetic Aspects of Speech Signal Synthesis with a High Degree of Naturalness”.

Литература

Евдокимова В. В., Скрелин П. А., Евграфова К. В., Чукаева Т. В., Швалев Н. В. // Теоретическая и прикладная лингвистика. — 2015. — вып. 1. — № 3. — С. 37–49.

Evdokimova V., Evgrafova K., Skrelin P. INVESTIGATING SOURCE - FILTER INTERACTION TO SPECIFY CLASSIC SPEECH PRODUCTION THEORY. In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 04.621.1–5 retrieved from http://www.icphs2015.info/pdfs/Papers /ICPHS0462.pdf

Evdokimova V.V. The use of vocal tract model for constructing the vocal structure of the vowels SPECOM’2006, Saint-Petersburg, 25-29 June 2006, p. 210-214, 2006.

Evgrafova K. V., Evdokimova V. V., Skrelin P. A.,.Chukaeva T. V, Shvalev N. V. A New Technique to Record a Voice Source Signal // Models and analysis of vocal emissions for biomedical applications: 8 th international workshop: December 16–18, 2013, 20

References