G.A. Kulikov, N.G. Andreeva

DETEMINATION OF BOUNDARIES OF PERCEPTIVE VOWELS PHONEME CONSTANCY DURING FREQUENCY TRANSPOSITION

St. Petersburg State University, Russia
199034, St. Petersburg, University Emb. 7/9
Phone : (812) 328-9706; Fax: (812) 328-2454; E-mail: kulikovga@mail.ru

The present study measured the frequency ranges of vowel phonetic constancy during the frequency transposition. There were several natural spoken and sung vowels [a], [o], [u] whose spectral component frequencies were changed from the original (by means of frequency transposition scheme) and whose amplitude ratios were kept constant. Received in such a way model signals were then presented to the listeners. It was shown, that effect of frequency shift results in a shift of perceived vowel quality. Thus, while sequence of phonetic changes were identical in all model series (from [u] to [o] and to [a](in upwards shifting), but the frequency values at which the transition in vowels label occur were different and depend on spectral component amplitude ratios of the signal.

The search for acoustic correlates of vowel quality remain one of the most challenging and fundamental problems in speech science. There is a long-standing debate concerning the efficacy of formant-based versus whole spectrum models of vowel perception [1-4]. However each of these approaches faces with a number of difficulties by consideration of concrete questions. So, there has, however, been some controversy suggesting that neither formant frequencies nor whole spectral shape hypotheses can be regarded as providing exclusive cues for vowel perception. However, it does not seem to be appropriate to view these models as opposed to one another because formant frequency is also involved in the "whole-spectrum" characteristics. These two, apparently, polar approaches (designated often as discreteness-continuity problem) are considered by a number of researchers as complementary [5].

In our previous studies, it was found that the perceived phonetic quality of vowels, produced with a high fundamental frequency (independent on speech realization mode) might be achieved by information about spectral maxima frequencies and their amplitude relations [6-9].

It was shown that due to differences in relative amplitude ratios different vowels occupy separate areas in the corresponding coordinate space. The parameters used and its interrelation makes it possible to discriminate the sounds [a], [o], [u] under the similar frequency characteristics of spectral components [10, 11]. But the relationship is not quite simple. The representation of the same vowels shift in a complicated manner as far as the F0 value rose, and different vowels occupy separate but other areas in the corresponding coordinate spaces.

It should be note that the shift of areas vowels occupied suggest that spectral envelope of a given vowel differs at different F0 ranges, and on the contrary, sounds with the similar spectral shape(spectral component amplitude ratios) will be perceived as different vowel depending on frequency range. But extent to which a given relative amplitude pattern can represent phonetically different vowels should be investigated.

The present study measured the recognition of the frequency transposed vowels whose spectral component frequencies were changed from the original and whose amplitude ratios were kept constant. The stimuli for the investigation were 39 spoken and sung vowels [a], [o], and [u] produced by adults and children (fundamental frequency values varies from 172 to 990 Hz, duration – 230–450 ms)

To construct the stimulus continuum, frequency transposition scheme was used (program Cool Edit Pro, Constant Stretch function). This manipulation applied to each naturally produced vowel resulted in a 39 model continua consisting of a number of signal which first harmonics' frequencies (h1 or "fundamental frequency") varied in stepwise fashion from 110-120 Hz to 800-900 Hz in one semitone steps. After generation, maximum amplitudes of the signals were normalized in order to equalize sound intensities within a test. It's necessary to note that preliminary suppression of singers' formant region at 25-35 dB in sung vowels was performed. The model stimuli were analyzed acoustically for spectral component frequency and intensity values. It was revealed that their spectral component amplitude ratios were kept constant and did not differ from the original.

The resulting vowels model continua were then presented to the trained listeners (5 subjects). On the basis of preliminary auditory analysis some signals (with extremely high and/or extremely low h1 frequencies) were excluded because of distortions in sounding. The remained signals were perceived as vowels [u], [o], [a] and some intermediate phones (u/o, o/a). A test set of model signals consisting of a randomized list of sounds differing in h1 frequencies was prepared of each modeling continuum.

In a test series a sets of model signals were presented to naïve normal-hearing listeners. Subjects were psychology students, all were native speakers of Russian. None of them had participated in the previous experiment. The subjects listened to the sounds using calibrated AKG K-141 headphone, in a sound-proof room, with the sound volume adjusted to a comfortable level. Taking into account the results of the preliminary auditory analysis vowel recognition was measured in 3-alternative identification paradigm: listeners were instructed to identify the vowel sound perceived as [a], [o] or [u] (forced-choice identification task). The stimuli were presented on-line by blocks of random orderings 25-30 tokens (with the insertion of a three-second inter-block interval and a one-second inter-stimulus interval). Each block was consisted of three replication of the same stimulus (one-second inter-stimulus interval) and was anticipated by block number. Listeners were given 3 s to respond after which another stimulus number was announced before next stimulus triad was presented. For each of model stimulus continuum, responses of at least ten subjects (as a rule, 16-25 subjects) were received.

Response data were collected to calculated level of identification; the significance of identification was defined by binomial criterion *m*. To characterize the shift in vowel identification the values of u_{max} and a_{min} were used. They were defined as h1 frequency maximum and minimum of authentic ($p<0.01$) /u/- and /a/-labeling responses respectively.

It was found that an upward or downward frequency transposition of natural vowels systematically altered vowel category assignments. In every of model stimulus continuum upwards frequency shift results in the changes of perceived vowel quality ranging from [u] to [o] and to [a] (Fig. 1).

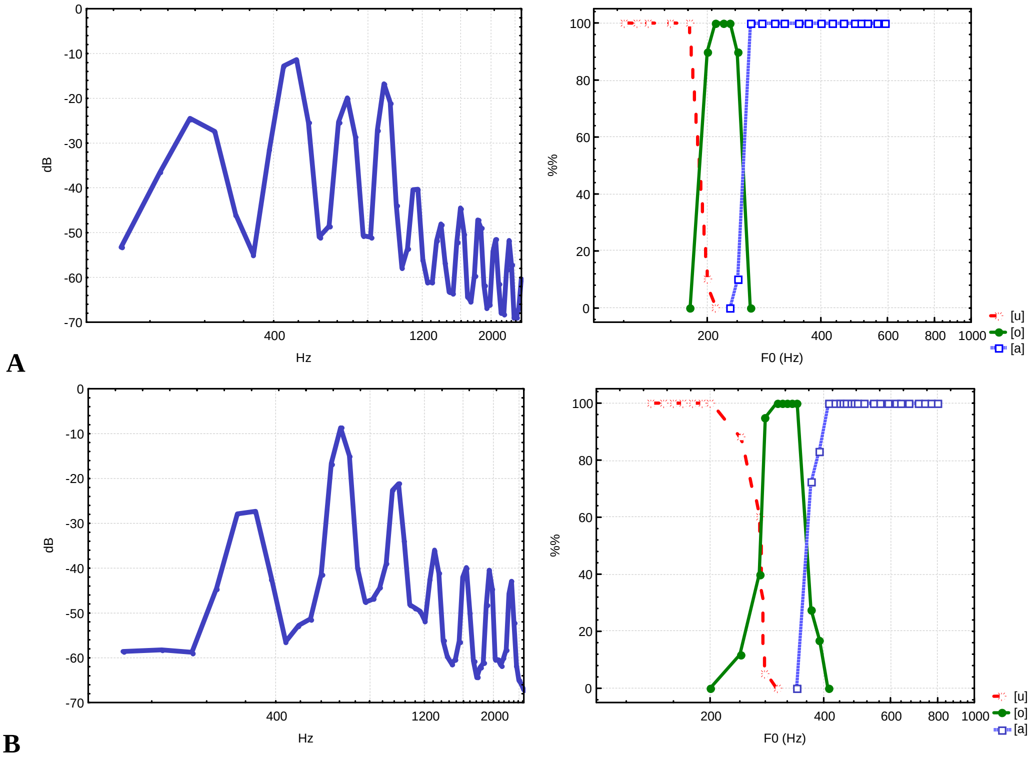

Fig. 1. Four different vowels' spectra (left) and identification score of their frequency transposed sounds (right).

Abscissa – frequency (Hz), ordinate – amplitude (dB) – left, and identification score (%%) – right; [a] – □, [o] – ●, [u] – ∗.

Frequency ranges of a-, o-, and u-labeling responses ($p<0.05$) as well as ranges of ambiguous identification (intermediate values of identification score) were observed corresponding to signal spectral shape. For those with a clear formant pattern (Fig. 1, A) a border of frequency values where phonetic shift occur depends on formant (spectral maxima) frequency values, and sounds labeled as [a], [o], [u] occupy different areas on two-formant plain similar to the natural vowels.

For signals displaying formant pattern ambiguity or signals with a lack of formant structure (Fig. 1, B) areas of perceived phonetic constancy as well as extent of ambiguity areas seems to be varying in accordance with "amplitude cue". To determine whether the relative amplitude ratios could account for the patterns of identification we implemented Spearman rank correlation test. The test showed statistically significant correlation ($p<0.001$) between amplitude ratios of first-second (A h1h2), second-third (A h2h3) harmonics and u_{max} and a_{min} values. There was no significant interaction between A h3h4 and u_{max} as well as between A h3h4 and a_{min} values. For these signals (unlike formant-pattern ones) there is highly significant linear correlation ($r= 0.9489$, $p<<0.0001$) revealed between u_{max} and a_{min}, values which characterized beginning of perceptual shift from [u] to [o] and lower boundary of [a] consequently (Fig. 2).

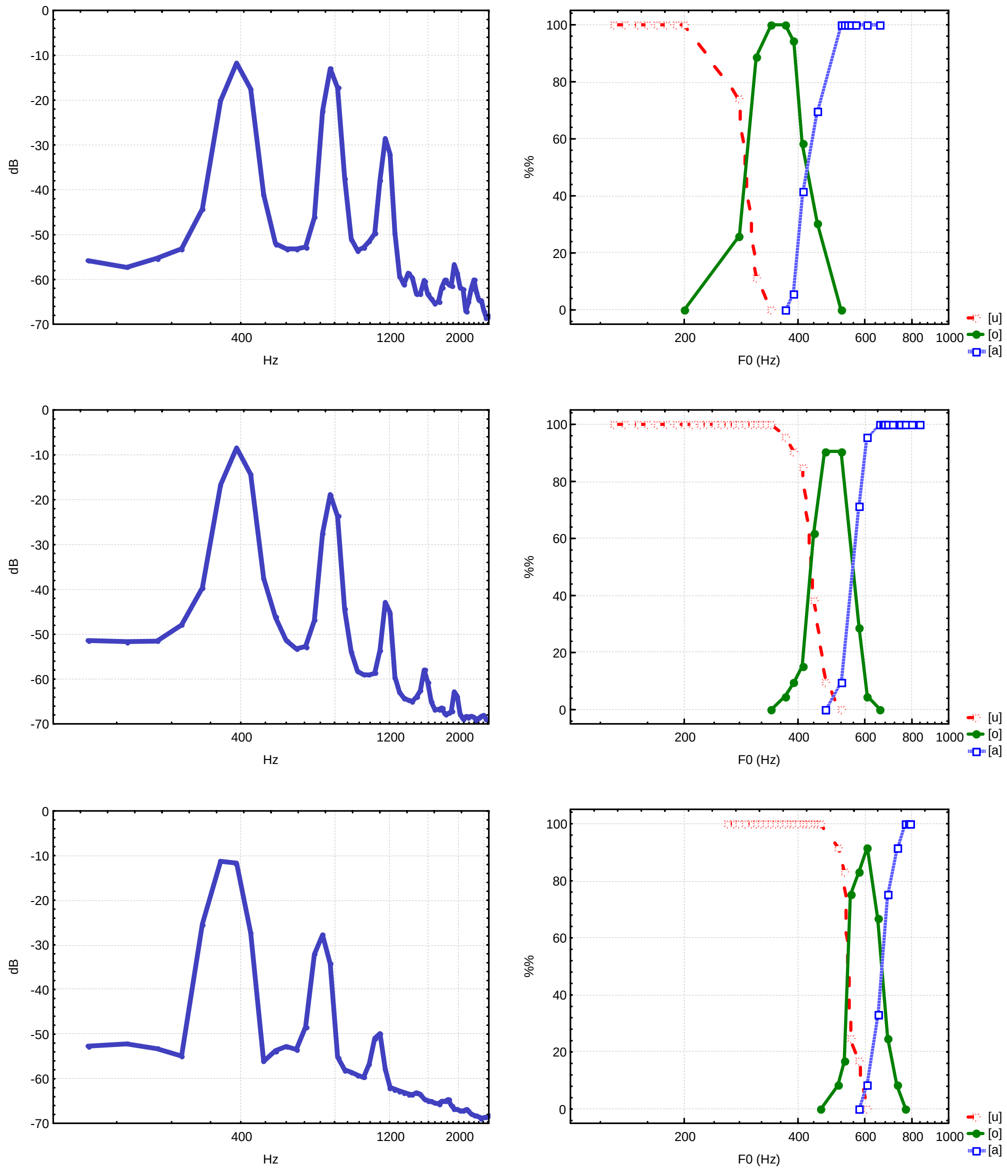

Fig.2. Identification score (right) of the model sounds received by frequency transposition of three vowels differs in spectral component amplitude ratios (left).
Legends: see Fig. 1.

Thus, at displacement of a vowel sounds [u], [o], [a] along a frequency axis at preservation of amplitude relations of first four spectral components, phonetic interpretation of vowels regularly varies. Boundaries of phonetic transition for vowels of the same category dependent are on initial peculiarities of the spectrum envelope.

Recently interest to investigation of influence of frequency transposition on speech perception [12-14] has renewed. One of the reasons of it is attempt to use frequency transposing of a speech signal in the area of residual hearing in hearing impaired persons [15, 16]. Results of the present study allow concluding, that for preservation of phonetic constancy at frequency transposing it is necessary to take into consideration amplitude relations of spectral components (the form of a spectrum envelope) of initial signal.

The study is supported by grant RFH № 07-06-00546a "Perceptive significance of amplitude ratios of vowels spectral components and the problem of vowels frequency transposition for children with residual hearing".

REFERENCES

1. Peterson G.E., Barney H.L. Control methods used in a study of the vowels // Journ. Acoust. Soc. Am. 1952. V.24. P. 175-184.
2. Bladdon R. A. W., Lindblom B. Modeling the judgment of vowel quality differences // Journ. Acoust. Soc. Am. 1981. V. 69. № 5. P. 1414–1422
3. Hillenbrand J., Getty L.A., Clark M.J., Wheeler K. Acoustic characteristics of American English vowels // Journ. Acoust. Soc. Am. 1993. V. 97. № 5. P. 3099–3111.
4. Zahorian S.A., Jagharghi A.J. Spectral-shape features versus formants as acoustic correlates for vowels // Journ. Acoust. Soc. Am. 1993. Vol. 94. № 4. P. 1966–1982
5. Molis M. R. Evaluating models of vowel perception// Journ. Acoust. Soc. Am. 2005. V. 118. №2. P. 1062–1071.
6. Kulikov G. A., Andreeva N. G., Pavlikova M. I., Samokishchuk A. P., Characteristics of vowel-like sounds in infants under six month of age // Doklady Biological Sci. 1999. V. 368. P. 471-473
7. Andreeva N. G., Kulikov G. A., Samokishchuk A. P., Common features of the amplitude – frequency characteristics of vowels in different forms of speech. // Acoustical Physics. 2002 V. 48. P. 620-622.
8. Andreeva N. G., Kulikov G. A., Sung vowel's characteristics under different fundamental frequency. // Sensory Systems, 2004. V. 18. P. 172-179. (in Russian)
9. Kulikov G. A., Andreeva N. G., Samokishchuk A. P., Aleksandrov A. Yu. Spectral maxima amplitude significance in vowel recognition. // Proc. XV Session of Russian Acoustical Soc., M., 2004. V.3. P. 420-423.
10. Andreeva N. G., Aleksandrov A.Y., Kulikov G.A. Acoustical characteristics of vowel sounds under various fundamental values. // Proc. XVI Session of the Russian Acoustical Soc., M., 2005. V.3. P. 611-613.
11. Kulikov G.A., Andreeva N. G., Goryainova G.Y. Perceptual significance of frequency and amplitude characteristics of women vowels at different fundamental frequency values // Proc. XIX Session of Russian Acoustical Soc., M., 2007. V.3. P. 597-600
12. Fu Q-J., Shannon R.V. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing // Journ. Acoust. Soc. Am.1999. V. 105. № 3. P.1889-1900
13. Assmann P.F., Nearey T.M., and Scott J.M. Modeling the perception of frequency-shifted vowels // Proc. of the 7th International Conference on Spoken Language Processing. 2002. P. 425- 428
14. Glidden C.M., Assmann P. F., and Nearey T. M. Effects of frequency shifts on vowel category judgments // Journ. Acoust. Soc. Am. 2002. V. 112. №1. P. 112, 249 (A).
15. Mc Dermott H.J., Dorkos V.P., Dean M.R., Ching T. Y.C. Improvements in speech perception with use of the AVR TranSonic frequency-transposing hearing aid // Journ. Speech, Lang. Hear. Res. 1999. V.42. № 6. P. 1323-1335.
16. McDermott H. J.; Dean M.R. Speech perception with steeply sloping hearing loss: Effects of frequency transposition // Brit. Journ. Audiol. 2000. V. 34. № 6. P. 353-361.