Multivariate analyses of speech signals in singing and non-singing voices

Yoshitaka Nakajima[*1], Hiroshige Takeichi[#2], Saki Kidera[§3], and Kazuo Ueda[*4],

[*] *Department of Human Science and Center for Applied Perceptual Research, Kyushu University, Japan*
[#] *RIKEN Nishina Center, Japan*
[§] *School of Design, Kyushu University, Japan*
[1]nakajima@design.kyushu-u.ac.jp, [2]takeichi@riken.jp, [4]ueda@design.kyushu-u.ac.jp

ABSTRACT

Background

Language and music are two main domains of auditory communication, and, in order to clarify the mechanism of either, it is necessary to understand in which aspects they are similar to and different from each other (Patel, 2008; Deutsch et al., 2011). It must be of crucial importance to examine the acoustic properties of speech signals in musical and non-musical contexts. We analyzed spoken sentences in eight languages/dialects (e.g., Ueda et al., 2010), and calculated power fluctuations extracted by critical-band filters, which were supposed to simulate the function of the peripheral parts of the auditory system. It was revealed that three factors that were related to four frequency bands, typically divided at 600, 1800, and 3200 Hz, appeared constantly in all languages/dialects. These factors were considered to be important to convey linguistic information. One of these factors was related to a frequency range around 1000 Hz, typically 600-1800 Hz, and it seemed that this factor could be associated with vowels, especially with open vowels. Of particular interest was whether these factors would appear in the same way in singing voices. Because something other than linguistic information must be playing an important role in the perception of human voice in a musical context, acoustic difference between singing and non-singing voices might appear. The present analysis could be a step to reveal acoustic cues specific to music.

Aims

We were interested in whether we could recognize stable factors in singing voices that could be related to frequency bands similar to those obtained for non-singing voices in our previous research. This simultaneously means that we were also examining whether we could find any systematic differences between speech-generating conditions in terms of power comodulations between critical bands.

Method

Two male and two female amateur singers were employed. They (1) sang two simple tunes [“Umi (The Sea)” and “Yūyake Koyake (The Evening Glow)” from the songs for primary schools collected and edited by the former Ministry of Education, Japan], modified for the present purpose, in Japanese, (2) sang variations of these tunes in which tone duration (as notated) or pitch was fixed, and (3) read the lyrics aloud at three different tempi. These speech signals were recorded, and analyzed utilizing a critical-band-filter bank covering a frequency range 50-6400 Hz. The power fluctuation of the band-pass waveform from each critical-band filter was smoothed utilizing a moving Gaussian window of σ = 5 ms. Factor analyses were performed to find common factors of these power fluctuations extracted from the 20 critical-band filters. A correlation-coefficient matrix between such power fluctuations was calculated for each speech sample as a first step of the factor analyses. The correlation matrices for all samples obtained from all participants were also compared directly with each other.

Results

In all speech-generating conditions, the same three factors could be identified as in our previous research (e.g., Ueda et al., 2010), in which only spoken sentences had been analyzed. One of the factors corresponded to a frequency range of several critical bands around 1000 Hz, and this means that its factor scores should have been high for vowels. This factor seemed important for the perception of pitch and rhythm.

The Euclidean distances between the correlation-coefficient matrices indicated a clear distinction between the speech-generating conditions: reading aloud, singing with a fixed pitch, and singing with the original pitch pattern. It has been revealed that there was indeed clear acoustic difference between singing and non-singing voices, but we are still examining the acoustic properties corresponding to the different speech-generating conditions.

Conclusions

Singing and non-singing voices were basically similar in terms of power comodulations between critical bands. However, the correlation-coefficient matrices of power fluctuations were different between the speech-generating conditions. The present approach is likely to lead to the discovery of acoustic properties characterizing musical voices.

Keywords

singing, speech, multivariate analysis, spectral changes, critical bands

REFERENCES

Deutsch, D., Henthorn, T, and Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America, 26*, 2245-2252.

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford University Press, Oxford.

Ueda, K., Nakajima, Y., and Satsukawa, Y. (2010). Effects of frequency-band elimination on syllable identification of Japanese noise-vocoded speech: Analyses of confusion matrices. *Proceedings of the 26th Annual Meeting of the International Society for Psychophysics*, 39-44.

Supported by Grants-in-Aid for Scientific Research Nos. 14101001, 19103003, 20330152, and 23653227 from the Japan Society for the Promotion of Science.