Study on Speaker Individuality Contained in Temporal Envelope of Speech

Zhu Zhi (1310029)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 12, 2015

Keywords: speaker individuality perception, modulation spectra, temporal amplitude envelope, noise-vocoded speech.

1 Introduction

The human voice not only includes linguistic information but also non-linguistic information such as emotional, gender, age, and speaker individuality. Of these, the characteristic of speaker individuality is very important, and humans can easily determine speakers by using it. In addition, speaker individuality also plays an important role in the cocktail party effect that we can focus on only one person's voice from the background of many people's voice. Thus, speaker individuality has a very important meaning in human speech communication. However, what features determine speaker individuality are still unclear in the mechanism responsible for human auditory perception.

If the factors that are contributing to the speaker individual perception can be clarified, those findings can lead to the elucidation of the mechanism of speaker individual perception. For engineering, in addition the speaker recognition system, it is possible to apply for the speech synthesis methods that can generate speech with speaker individuality, the speech recognition systems that adapt to the speaker and many other voice processing technology.

The basically reason of the produce of speaker individuality is the variation of vocal organs those produce voice (vocal cord and vocal tract) depending on the speaker. Moreover, vocal organs' congenital difference of shape and the acquired difference of motion are all contribution to the identification of speaker. From this viewpoint, the conventional study about speaker individuality were focusing on the product of interacting acoustic cues: source characteristics of the vocal cords (F0 and its dynamics) and filter characteristics of the vocal tract (spectral envelope and formants). Furthermore, each feature has been studied separately regarding static and dynamic components.

Kazama et al. (2009), on the other hand, focused on individuality information contained in the narrow band temporal envelope of speech, instead of the viewpoint of speech production. It was reported to be possible to express individuality information from the difference in narrow-band temporal envelope correlation matrices (ECMs). However, the ECMs only represented the correlation between different frequency bands. These features that contributed to speaker individuality in the temporal envelope and whether speaker individuality would be used by people to identify speakers still remained unknown.

From the viewpoint of auditory perception, temporal amplitude envelope is known as a very important factor for speech perception. It is also well-known that the low modulation frequencies, particularly below 16 Hz, play a crucial role for speech intelligibility (e.g., Drullman et al., 1994). Moreover, the presentation of modulation information of only a few acoustic bands such as noise-vocoded speech is sufficient for speech recognition (Shannon et al., 1995).

The ultimate goal of this study is to clarify whether there were physical features related to speaker individualities in the temporal envelope of speech waveforms from the viewpoint of auditory perception. In this study, it is hypothesized that "the physical quantities that are highly different due to the speaker are contributing to speaker individuality perception". At first whether individual differences in temporal amplitude envelopes obtained from the output of an auditory filterbank could be observed is investigated. Then, the relationship between these individual differences and speaker individuality perception is investigated in a psychoacoustic

experiment.

2 Analysis of Individual Differences on Modulation Spectrum of Speech

The individual differences in temporal amplitude envelopes are investigated by analyzing variances in the modulation spectra of different speakers. The modulation spectrum corresponds to the frequency spectrum of the temporal envelope. To compute the modulation spectrum, the speech signal is filtered by a 33-channel filterbank firstly. The filter bandwidths are characterized by the ERB_N which is a scale that is very similar to a human's frequency decomposition function Then, the modulation spectrum is obtained by taking discrete Fourier transform on the envelope of each band.

As a result of the variances in the modulation spectra of different speakers, it is found that variances in modulation spectrum over 20 ERB_N–number were clearly larger than those of the others. Variances at modulation frequencies below 15 Hz for differences were the largest in frequency bands from 20 to 29 ERB_N–number. The largest variances at whole modulation frequencies were observed in higher frequency component ranges over more than 30 ERB_N–number. These results suggested that these larger variances could be interpreted as speaker differences in the modulation spectrum.

3 Speaker individuality perception with band-limited modulation spectrum

In the experiment, in order to focus on the modulation frequencies, the simulation strategy was implemented by using Noise-vocoder speech. The speech signal was split up into a series of frequency bands (width of 1 or 2 ERB_N–numbers) and the envelope for each band was low-pass filtered to investigated the impact of the changes in the modulation components on speaker individuality perception.

The results showed that the performance of speaker identification was as a function of the cutoff frequency of low-pass filter. For the situation

of bandwidth of 1 ERB_N–number, the average recognition rate reduced when the cutoff frequency was reduced. However, because of the large variance, this result was not significantly by using ANOVA. In the situation of bandwidth of 2 ERB_N–numbers, the recognition rate reduced when the cutoff frequency was reduced from 16 Hz significantly and it was reach to 60 % when the upper limit of modulation spectra was 1 Hz.

4 Conclusions

The aim of study was to clarify the speaker individuality information included in temporal envelope of speech. The individual differences was investigated by analyzing variances in the modulation spectra of different speakers. A psychoacoustic experiment was carried out to clarify the relationship between these individual differences and speaker individuality perception.

In the results, the speaker individuality information was remarkable in the low modulation frequency components of the temporal amplitude envelope. These results also suggested that modulation filtering affects the perception of speaker individuality and there were dominant modulation regions (bellow 16 Hz) for speaker recognition.