Study on redundancy contained in temporal amplitude envelope of Japanese speech

Yasutaka Nishino (1110047)

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

February 12, 2014

Keywords: Japanese speech, temporal amplitude envelope, redundancy, modulation frequency components.

Speech is robust perceptually. Important features related to linguistic information are redundantly contained in speech signals. Therefore, we can easily recognize linguistic information even if some of the features are deleted or modified. A well known phenomenon showing the redundancy of linguistic information is robust perception of noise-vocoded speech. Noise-vocoded speech contains only the temporal amplitude envelope of speech and is generated by amplitude modulation with a noise carrier. Information related to the intelligibility of voice has a lot of temporal amplitude envelope of speech. It was reported by previous studies of noise-vocoded speech. However, those have not been investigated for redundancy contained in temporal amplitude envelope. Accordingly, We elucidate the redundancy contained in temporal amplitude envelope of Japanese speech.

Speech redundancy is the characteristic that recognition of linguistic information is possible even if some features are lost. Because, it is contained distributive and redundant in speech signal that features related to speech perception. Technique utilizing this property is speech coding technique and audio compression. The feature which does not affect listening comprehension is deleted and transformed from the original speech signal. Therefore, we must focus on features which do not affect recognition of language information even when deformation or remove features

of temporal amplitude envelope. We focus on the modulation frequency components. The following two points have been reported to modulation frequency components. Modulation frequency components are important to speech intelligibility from 4 Hz to 16 Hz. And, modulation frequency components have a peak peculiar from 2 Hz to 8 Hz. However, those previous works are not discussing that recognition of linguistic information by controlling this feature. Therefore, it is controlled systematically that modulation frequency components are contained in temporal amplitude envelope, and recognition of linguistic information investigate.

In order to systematically control the modulation frewuency component contained in temporal amplitude envelope, we focused on time structure of the Mora speech and time structure of consonants. The prosodic structure of Japanese is based on mora. Morae are repeated in a steady rhythm. This means that the cycle of morae is important for Japanese. The total time length of the speech signals of four morae used in the experiment was about 1000 ms; thus, the duration of one mora was about 250 ms. If this duration is one cycle of morae repetition, the outline of the primary shape of the amplitude modulation based on the moraic syllable structure should be represented by 4 Hz modulation. This suggest that the shape of the amplitude envelope of the moraic syllable structure is an important factor in recognizing linguistic information. We predicted upper limit frequency of the modulation component is high, with listening experiments focusing on the time structure of the consonant. The upper limit frequency of the modulation components must be high because consonant of linguistic information should be recognized. However, the upper limit of frequency modulation component was low. This indicates that recognition of linguistic information can be hearing without a part of temporal amplitude envelope of consonant. Because that is a clue to the attack of temporal amplitude envelope of the consonant, recognition of language information of consonant was possible.

Modulation frequency components necessary for recognition of language information, only to be saved modulation frequency components of only can reproduce time structure of Mora. By the difference in attack of temporal amplitude envelope, recognition of linguistic information is possible in a part of consonant. From these results, redundancy contained in temporal

amplitude envelope of Japanese speech is modulation frequency components are removed by low-pass filter.