

Title	聴知覚メカニズムに着目した音声の年齢知覚に関する研究
Author(s)	畠山, 達也
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15908
Rights	
Description	Supervisor: 鷗木 祐史, 先端科学技術研究科, 修士 (情報科学)

Study on the age perception focuses on
the auditory perception mechanism

1710163 Hatakeyama Tatsuya

Many people can correctly estimate the age of a speaker simply by listening to the speaker's voice. Age perception can be regarded as the perception of non-linguistic information, so it has been studied extensively of speaker identification and speaker individuality. If we can reveal the mechanism behind auditory perception in terms of how individuals perceive age from an observed voice, this finding will contribute to research on speaker identification, and also to the applications of various speech signal processing techniques such as age estimation. The main problem in studying speech science related to age perception is clarifying the relationship between the corresponding acoustic features on the basis of the speech production mechanism and age perception from a speaker's voice. In order to reveal the nature of age perception, it is necessary to investigate it from the viewpoint of speech production. In this paper investigate the relationship between the perceived age of a speaker's voice and the corresponding auditory factors (roughness, fluctuation strength, and sharpness) sound quality metrics.

I conducted experiments in which i labeled the perceptual ages to corresponding speech stimuli in three data based. The first speech corpus, CIAIR-VCV, includes the speech of men and women aged 6 to 12. The speech signals of all 14 speakers were used as stimuli in the child speech dataset. The second corpus, APP6BLA, includes thirty male and female speech of teens to 60s. The speech signals of 74 speakers were used as stimuli in the adult speech dataset. The third corpus, S-JNAS, includes male female voice over 60's. The speech signals of 28 speakers were used as stimuli in the adult speech dataset. Age perception experiments was carried separately for each dataset. Ten native japanese speakers with normal hearing (six males in their 20s, four female in her 20s) participated in the experiment. The both experiments were carried out in a soundproof room. Speech stimulus was randomly presented. The participants were required to answer the perceived age using values of one year. As the results, it was shown that the listener accurately estimates the age of the speaker to some extent. Whether it is a male speaker or a female speaker, the voice of the child is smaller or smaller than the adult. In the case of adult speech, the variation of the result becomes large, and the perceived age is estimated to be lower than the actual age. In the case of women, this tendency appeared more prominently.

Previous studies have shown that the spectral tilt is highly correlated with actual age in the case of male speech, and the fundamental frequency is highly

correlated with actual age in the case of female speech. So, i investigated the relationship between these two acoustical features of the speech stimuli that we used in the experiments and perceptual age. The spectral tilt in the dB per octave frequency was determined using the least mean squared method (LMS) from shortterm. Here, fourier transformation in which the window function is Hanning, frame length is 10 ms, and frame shift is 5 ms. The F0 was obtained using STRAIGHT TEMPO. F0 and spectral tilt were then averaged in the time domain to use them as the acoustical features.

From the result, looking at the ages as a whole, in the case of male voice, there is a tendency that roughness increases as the perceived age increases. On the other hand, looking at children, adults, and elderly people, roughness tends to decline in elderly counties. In the case of female voice, there was no tendency in roughness to perceived age when looking at the ages as a whole. However, when divided into small age groups, descending tendency was seen in child counties and elderly counties. Relationship between sensory age and fluctuation intensity of male speaker, relationship between perception age and fluctuation intensity of female speaker. Looking at the ages as a whole, in the case of male voice, there is a tendency that the perceived age increases and the fluctuation intensity decreases. On the other hand, when divided into children, adults, and elderly people, the tendency of roughness to decline was observed in children and adults, and the trend of fluctuation intensity tended to increase in elderly counties. In the case of female speech, there was no tendency in the fluctuation intensity with respect to perceived age as seen in the entire age. However, when divided into small age groups, a downward tendency was seen in child counties. Relationship between perceived age of male speaker and sharpness, from the result of the relationship between perception age and sharpness of female speaker In the case of male speech, there was no tendency in sharpness to perceived age. When dividing into children, adults, and elderly people, Sharpness was seen to rise in elderly people. There was no tendency in sharpness against sensory age in female speech. However, descent tendency was seen in child counties.

From the results, in the case of male voice, there is a tendency that the perception age increases and the roughness increases. On the other hand, when divided into children, adults, and elderly people, roughness tends to decline in elderly counties. In the case of female speech, there was no tendency to roughness to perceived age. However, descent tendency was seen in child counties and elderly counties. Relationship between perceived age and fluctuation intensity of male speaker, from the results of the relationship between perceived age and fluctuation intensity of female speaker. In the case of male voice, there is a tendency that the perceived age increases and the fluctuation intensity decreases. On the other hand, when divided into

children, adults, and elderly people, the tendency of roughness to decline was observed in children and adults, and the trend of fluctuating intensity was seen in the elderly counties. In the case of female speech, there was no tendency in the fluctuation intensity with respect to perceived age. However, descent tendency was seen in child counties. Relationship between perceived age of male speaker and sharpness, from the result of the relation between perceived age of female speaker and sharpness. In the case of male speech, there was no tendency in sharpness against perceived age. On the other hand, when divided into children, adults, and elderly people, sharpness was seen to rise in elderly people. In the case of female speech, there was no tendency in sharpness to perceived age. However, descent tendency was seen in child counties.

Comparing the case of seeing the relation between perceived age and acoustic feature and the case of seeing the relationship between perceived age and sound quality evaluation index, In men, F0 ($r^2 = 0.53$) was the most highly correlated with the overall perceived age of children, adults, and the elderly. The next highest was the spectral tilt ($r^2 = 0.4$). In the case of women, it was F0 ($r^2 = 0.38$). The next highest was fluctuating strenght ($r^2 = 0.35$). However, since the spectral tilt varies widely, other feature quantities may be involved.

Based on the perceptual age and spectral tilt, and the relationship between perceived age and F0, as the tendency was seen in perceived age and spectral tilt of male voice, it is possible to explain the result of this time with the acoustic feature which is conventionally known for perceived age. However, the value of correlation to perceived age is low. Therefore, regarding perceptual age, it is necessary to evaluate the impression of hearing. From the result of the relation between perceived age and sound quality metrics, i could not find an indicator that could explain the whole age. Therefore, further analysis was attempted by considering fluctuation of perceived age. From the results in Chapter 4, i extracted 10 data with small standard deviations and obtained relationships with each feature quantity. As a result, the perceptual age and the roughness could be related.