Creation of Sounds from Linguistic Information (Continuous Sounds)

Marie OHNO*, Taichi SATO**, Kihachiro TANAKA***

*Intelligent Mechanical Engineering, Graduate School of Science and Engineering, Tokyo Denki University
Ishizaka, Hatoyama, Hiki-gun, Saitama 350-0394, JAPAN
e-mail: mre@satolab.n.dendai.ac.jp

**Department of Intelligent Mechanical Engineering, School of Science and Engineering, Tokyo Denki University

***Department of Mechanical Engineering, Faculty of Engineering, Saitama University

Abstract

Many people use the same onomatopoeic word to express a given sound. To explore the connection between continuous sounds and their corresponding onomatopoeia, we created sounds based on the linguistic information of onomatopoeic words. We then carried out an auditory perception experiment using these sounds, before finally analyzing the actual and pronounced sounds with a 1/3 octave analyzer. We found that the frequencies of a sound perceived to be an onomatopoeic word are similar to the frequencies comprising the corresponding actual sound, and also close to the frequencies of the onomatopoeic word when actually uttered.

1. Introduction

Recently, there have been strong demands to investigate how people perceive sounds generated from various devices [1], [2] and [3].

People have the ability to express sounds they have heard as onomatopoeia. They can also do the reverse – i.e. infer what sounds onomatopoeic words represent. We believe this means there is some kind of close relationship between onomatopoeia (linguistic information) and the sounds themselves.

Many onomatopoeic words are used to represent car noises, and on this topic in particular, much has been reported on the links between the onomatopoeia and the reason for different car noises being produced [4]. There has not, however, been any investigation into why a particular onomatopoeic word is used to express a given sound. This paper shall thus investigate the relationship between onomatopoeic words (linguistic information) and sounds, by creating sounds from onomatopoeia.

2. Selection of onomatopoeic words

We decided to begin by selecting onomatopoeic words to use in this study. We decided that suitable onomatopoeia were those which are used by many people and from which sounds can easily be inferred. Next, we conducted a survey asking which onomatopoeic words corresponded to various continuous sounds.

When compiling our results, we focused on whether many people used the same onomatopoeic phrases. When the continuous sounds were ranked in descending order of frequency of the same word being used, the highest-ranking four (4) were:

1. the sound of a toilet flushing: jaa (98/143)
2. the sound of a detuned television: zaa (94/143)
3. the sound of an aerosol spray: shuu (84/143)
4. the sound of a shower running: shaa (75/143)

The bracketed numbers show how many of the 143 respondents express a given sound with the same onomatopoeic phrase. This means, for example, just under 70 percent of respondents use the onomatopoeic phrase “jaa” to express the sound of a toilet flushing. The sounds with the highest values are listed above. We therefore chose to study this group of continuous sounds.

3. Creation and output of sounds corresponding to onomatopoeic phrases

From the information gained from onomatopoeic words, it is predicted that the corresponding sounds are white noise type waves composed of a mixture of different frequencies. We passed white noise through a band-pass filter, as shown in Fig.1, and generated a white noise type waveform. For a band-pass filter, we prepared 10 types of waveform by dividing the 50 Hz-12.5 kHz range of frequencies into 10 (see Table 1). We set the length of the continuous sounds at 1 second, ample time for them to be recognized as continuous sounds. We fed the white noise output of the function generator through a band-pass filter before feeding it into a personal computer. We then processed the resultant waveform to generate a one-second continuous waveform as shown in Fig.1. We output the generated sound from the personal computer, and carried out an auditory perception experiment using headphones.
4. Auditory perception experiment method

We performed the auditory perception experiment described below on 3 testees (referred to as A, B, and C; all male university students with normal hearing). To ensure the auditory perception experiment would not be affected, we did not provide the testees with any information at the time.

(1) In addition to the 10 sounds mentioned earlier, we selected 5 sounds to be references. The standard sounds were made by passing white noise through frequency filters of 160-250 Hz, 400-630 Hz, 1,000-1,600 Hz, 2,500-4,000 Hz, and 6,300-10,000 Hz, respectively.

(2) We had the testee listen to two sounds – one of the reference sounds and one from the group of 10 sounds. In the case of “jaa”, for example, we had the testee decide which of these sounds sounded more like “jaa”. Here, we will explain the procedure for the 160-250 Hz reference sound. We went through all possible combinations of this standard sound and the group of 10 sounds, i.e. the 160-250 Hz sound and the 50-200 Hz sound; the 160-250 Hz sound and the 200-315 Hz sound; ... the 160-250 Hz sound and the 8,000-12,500 Hz sound.

(3) The inspector would draw a circle in the corresponding box on the auditory perception scorecard (Table 1) if he thinks the sound from the group of 10 sounds sounds more like “jaa”, and a cross if he does not think so.

(4) Steps (2), and (3) are repeated for the other references. This is done for all references. That is, each testee performs 50 pair comparisons of sounds for each onomatopoeic word.

(5) Steps (2), (3), and (4) are repeated 7 times for each testee and each onomatopoeic word.

5. Auditory perception experiment results

In the experiment described above, the more like “jaa” a noise sounds, the more circles are marked in the scorecard in Table 1. As there are 5 references, a maximum of 5 items will be circled. This means that 5 is the maximum number of points possible in Table 1. We used the scorecard to find which frequencies range sound like “jaa”. We showed our results with grayscale as shown in Fig. 2. This figure shows the mean of the auditory perception experiment performed 7 times. The vertical axis represents frequency, and the results for the various onomatopoeic sounds – “jaa”, “zaa”, “shuu”, and “shaa” – are shown from left to right. In this diagram, the shading becomes lighter as rankings approach 5, indicating the frequency region which sounds closest to the onomatopoeic phrase in question.

Testee A’s results, for example, can be summarized as follows:

(1) “jaa”: frequency of 200 Hz-800 Hz
(2) “zaa”: frequency of 315 Hz-1.25 kHz
(3) “shuu”: frequency of 1.25 kHz-5 kHz
(4) “shaa”: frequency of 800 Hz-3.15 kHz

The above represents the frequencies range that are interpreted for these onomatopoeic sounds. This shows us that the frequencies range which can be interpreted as being a given onomatopoeic sound varies from sound to sound. While the results from the 3 testees show individual variation, they display the same trend, in that sounds in a low frequency range tend to sound like “jaa” and “zaa”, while those in a high frequency range tend to sound like “shuu” and “shaa”.

6. Relationship between actual characteristics of continuous sounds and auditory perception experiment results

As shown in Diagram 2, the frequencies range which can be interpreted as representing a given onomatopoeic word differs from word to word. This is one important finding of the auditory perception experiment. To consider why this is the case, we investigated the characteristics of the actual continuous sounds. In practice, this involved recording the continuous sound corresponding to, for example, “jaa” – the sound of a toilet flushing, and investigating its characteristics. We did the same with the other onomatopoeic words.

Our method of recording continuous sounds is summarized simply below.

(1) toilet flushing: we flushed a Japanese-style toilet
(2) detuned television: we unplugged the antenna from a TV and turned the TV on
(3) aerosol can: we discharged a spray can in a semi-anechoic chamber
(4) shower: we turned on a shower nozzle in a bathroom

<table>
<thead>
<tr>
<th>Table 1 Auditory perception scorecard</th>
</tr>
</thead>
<tbody>
<tr>
<td>f Hz</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Mark of scorecard

Fig.2 Auditory perception experiment results (Testee A)
Fig. 3 shows the results of a 1/3 octave analysis of these continuous sounds. The graphs are, from top to bottom, the sound of a toilet flushing, the sound of a detuned television, the sound of an aerosol can, and the sound of a shower. One can see that the peak frequency differs for different sounds.

We examined what type of relationship exists between the distribution in frequencies and the results of the auditory perception experiment. The experimental results showed that a relatively low-frequency sound range was perceived as “jaa” (Fig. 2). The 1/3 octave analysis of the actual sound (that of a toilet flushing) (Fig. 3(a)) also peaks at a relatively low frequency. With the onomatopoeic word “shuu”, on the other hand, sounds in a relatively high frequency range were perceived as sounding like this word. The corresponding 1/3 octave analysis of the actual aerosol can sound (Fig. 3(c)) also showed a relatively high frequency range. We therefore considered that there is some kind of relationship between the two groups, and summed this up below.

The relationship between the results of the auditory perception experiment and the actual 1/3 octave analysis of sound is summarized in Fig. 4 in terms of frequency distribution. In Fig. 4, the horizontal axis shows the frequency found from the results of the auditory perception experiment, and the vertical axis shows the frequency found from the 1/3 octave analysis of actual sounds. A, B, and C on the graph represent the three testees. The squares on the graph were produced as described below, based on the two sets of results.

(1) The range of frequencies which are perceived as the onomatopoeic word in question are found from the results of the auditory perception experiment. That is, we find the white area in Fig. 2. To find the frequencies more rigorously, however, we use the scorecard described in Chapter 4. Of all the frequency ranges, the 3 frequency ranges that scored highest on the 5-point scale were selected to define the frequency ranges obtained from the experiment. These frequency ranges were made the horizontal axis of the square in Fig. 4. The same was repeated for each subject (A, B, and C on the graph).

(2) The region from the highest peak frequency to the 3rd-highest peak frequency for the 1/3 octave analysis results of actual sounds was made the vertical axis of the square in Fig. 4.

(3) We plotted the square as per (1) and (2), and placed it on the coordinate axes shown in Fig. 4.

We can see from this graph that the plotted square is
man, there is a need to investigate the relationship between the results of the auditory perception experiment and human utterances. We had our testees utter the various onomatopoeic words (“shuu”, for example), and performed a 1/3 octave analysis of these sounds. The results of testee C pronouncing “shuu” are shown in Fig.5. In this figure, (a) is the waveform immediately after the beginning of the utterance, and (b) after a short time has passed, representing the waveform at 0.03 second after (a). While (a), immediately after the utterance has begun, shows waves of low amplitude and high frequency, in (b), waves of high amplitude and low frequency have begun to appear on the right hand side. This is because /u/, the vowel of “shuu”, has begun to appear. We can see that there is a difference in peak frequency between results of the 1/3 octave analysis before and after the /u/ sound appears.

The 1/3 octave analysis results from before the occurrence of /u/ displays a very similar trend to that observed in the 1/3 octave analysis results for the actual sound, shown in Fig.3 (c). This shows us that when considering and investigating utterances, there is a need to investigate waveforms only for the moment one begins to expel air from the mouth, or a short time after the utterance has begun.

**8. Conclusion**

Noticing that many people express a given sound with the same onomatopoeic word, we created sounds based on linguistic information acquired from onomatopoeic words. We conducted an auditory perception experiment to determine which created sounds were perceived to sound like the onomatopoeic words studied. We investigated the relationship between our experiment results with actual sounds, and with human utterances. As a result, we showed that the frequencies perceived to sound like the onomatopoeic words were close to the frequencies of the actual sounds, as well as the frequencies of the utterances. We believe that a detailed examination needs to be conducted on the influence of vowels and consonants on utterances in future.

**References**