Humans have the ability to express certain sounds as onomatopoeia, and to imagine the real sounds that onomatopoeic words correspond to. 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 uttered 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. Onomatopoeia is language which people utter. Therefore it is necessary to clarify the relationship between the frequencies perceived to sound like the onomatopoeic word and uttered sounds. In this report, we paid attention to both the vowel and the consonant of uttered sound of onomatopoeia. As a result we found that the utterance time of a consonant changes by the mechanism of utterance. And the ratio of energy between the vowel and the consonant influences frequency of the sound that an onomatopoeia expresses.

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 Sounds corresponding to onomatopoeic phrases

2.1 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 six (6) were:

1. The sound of an untuned television: \( \text{za} \) (89/115)
2. The sound of an aerosol spray: \( \text{shu} \) (85/115)
3. The sound of a flushing WC: \( \text{ja} \) (72/115)
4. The sound of a Bunsen burner: \( \text{bo} \) (62/115)
5. The sound of a shower: \( \text{sha} \) (58/115)
6. The sound of air flowing in a duct: \( \text{go} \) (54/115)

The above list was obtained as the result of a questionnaire study on onomatopoeia. The figures in brackets indicate how many out of 115 questionnaire respondents used this onomatopoeia to represent the corresponding sound. For example, almost 80% of the respondents used the onomatopoeia \( \text{za} \) to represent the sound of an untuned television. The onomatopoeia listed above are those for which this number was large.

2.2 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 from 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. We also performed listening tests to ascertain which frequencies sound like the corresponding onomatopoeia.

2.3 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 “jā”, for example, we had the testee decide which of these sounds sounded more like “jā”. 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 “jā”, 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.

2.4 Auditory perception experiment results

In the experiment described above, the more like “jā” 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 “jā”.

For example, Fig. 2 shows the results for testee A. The vertical axis represents the frequency, and the results for each type of onomatopoea are arranged horizontally. This figure shows the mean of the auditory perception experiment performed 7 times. In this figure, paler regions indicate that the sound more closely resembles the onomatopoeia in question. From the results for testee A, the frequency ranges that correspond to the different onomatopoeia are as follows:

(1) bo̊: 50 Hz – 500 Hz
(2) go̊: 50 Hz – 500 Hz

<table>
<thead>
<tr>
<th>f Hz</th>
<th>Standard sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>160-250</td>
<td>400-630</td>
</tr>
<tr>
<td>① 50-200</td>
<td>② 200-315</td>
</tr>
</tbody>
</table>

Figure 1: An example prepared wave of sound (in the case of white noise passed through band-pass filter of 50 Hz-200Hz)
The results in Fig. 2 are arranged in order of the corresponding onomatopoeia that the sounds resembled at low frequencies in the hearing evaluation, regardless of the magnitude of the numbers in the abovementioned questionnaire test results. We have showed that the frequencies perceived to sound like the onomatopoeic words were close to the frequencies of the actual sounds [5]. Next, the relationship between these frequencies that sound like onomatopoeia and the sounds that are spoken when uttering these onomatopoeia is investigated below.

3 Acoustic characteristics of uttered sounds

Onomatopoeia are words uttered by humans. Accordingly, it is important to clarify the relationship between the listening test results and the sounds uttered by people.

When these onomatopoeia are spoken, they start with a consonant and then change into a vowel sound (Fig. 3). Since the frequencies that sound like onomatopoeia are thought to be affected by the frequency and acoustic pressure of the uttered consonants and vowels, we first concentrated on the duration for which the consonants are uttered. The consonant durations were read from the time axis waveforms as shown in Fig. 3.

The results, which are shown in Table 2, can be divided into onomatopoeia with short consonant durations (jā, zā, bō and gō) and onomatopoeia with long consonant durations (shā and shū). Although the onomatopoeia jā and shā are very similar, the lack of a voiced consonant in the latter was found to result in a longer consonant sound.

Based on these findings, we looked into the possibility of classifying between voiced and unvoiced sounds based on the consonant length. In Japanese, voiced sounds correspond to syllables beginning with consonants like g, z, d and b that are the voiced counterparts of unvoiced syllables beginning with k, s, t and h. The Japanese language also contains plosive syllables beginning with p, which are also regarded as unvoiced.

Here, in order to study other forms of onomatopoeia, we studied the following five groups of onomatopoeia comprising voiced and unvoiced sound pairs:

**Table 2 Time of consonant**

<table>
<thead>
<tr>
<th>Onomatopoeia</th>
<th>Time ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>jā</td>
<td>46</td>
</tr>
<tr>
<td>zā</td>
<td>52</td>
</tr>
<tr>
<td>bō</td>
<td>59</td>
</tr>
<tr>
<td>gō</td>
<td>60</td>
</tr>
<tr>
<td>shā</td>
<td>142</td>
</tr>
<tr>
<td>shū</td>
<td>165</td>
</tr>
</tbody>
</table>

**Table 3 Time of consonant**

<table>
<thead>
<tr>
<th>Onomatopoeia</th>
<th>Time ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unvoiced</td>
<td></td>
</tr>
<tr>
<td>Voiced</td>
<td></td>
</tr>
<tr>
<td>ki</td>
<td>112</td>
</tr>
<tr>
<td>shi</td>
<td>177</td>
</tr>
<tr>
<td>fi</td>
<td>163</td>
</tr>
<tr>
<td>kya</td>
<td>65</td>
</tr>
<tr>
<td>byu</td>
<td>142</td>
</tr>
</tbody>
</table>
an open vowel whereas according to the references, grouped together as vowels that are not narrow. As narrow vowels, and the vowels classifications of vowels (open, narrow) and durations into four groups by combining these figure 4 shows the result of classifying the consonant position between narrow vowels and open vowel. In articulatory phonetics they can be classified according to the following three criteria [6]:

Vowels are generally voiced, and in terms of their articulatory phonetics they can be classified according to the following three criteria [6]:

(a) Rounding of the lips
(b) The front-to-back position of the top of the tongue
(c) The vertical position of the top of the tongue

With regard to (c), vowel sounds produced when the top of the tongue is in a low position are called “open vowels”, and vowel sounds produced with the top of the tongue in a high position are called “narrow vowels”. In Japanese, the vowels i and u are classified as narrow vowels, and the vowels a, e and o are grouped together as vowels that are not narrow. According to the references, a is clearly positioned as an open vowel whereas e and o occupy an intermediate position between narrow vowels and open vowel.

Figure 4 shows the result of classifying the consonant durations into four groups by combining these classifications of vowels (open, narrow) and consonants (voiced, unvoiced). Here, since e and o are not narrow vowels, they are grouped together with the open vowel a and are referred to as open vowels in this study. In this figure, larger circles represent longer consonants (i.e. syllables with a longer consonant duration). It is clear that the longest consonant durations occur in the bottom left region of this figure, where unvoiced consonants are paired with narrow vowels. Conversely, the shortest consonant durations tended to occur with voiced consonants and open vowels. Since open vowels have greater energy than narrow vowels, it is thought that the sounds uttered for onomatopoeia that use open vowels are more likely to be affected by the vowels.

4 Effects of the sound pressure of uttered vowels and consonants

In the previous section we found that open vowels in onomatopoeia result in greater vowel energy and shorter consonant durations. This causes the consonant energy to become relatively small, and it is thus plausible that when onomatopoeia is uttered, the magnitude of the vowel and consonant parts may affect the frequency distribution of the uttered sound.

We therefore analyzed the frequency distribution of vowels and consonants in the uttered sounds corresponding to various onomatopoeia. The results are shown in Fig. 5. From top to bottom, this figure shows the frequency spectra of acoustic pressure for the onomatopoeic words jā, zā, shā, shū, bō, and go. The area graphs show the vowel frequency distributions, and the bar graphs show the consonant frequency distribution.

4.1 Onomatopoeia with the same vowel “a”

The onomatopoeic words jā, zā and shā have the same vowel sound a but different consonants. According to the results of listening tests, the frequencies used in these onomatopoeia are considered to increase in the order jā, zā, shā (Fig. 2). From Fig. 5(a), (b) and (c), it can be seen that the vowel frequency distributions (area graph) of these three onomatopoeia have the same characteristic in that they possess a fixed peak close to 1 kHz. On the other hand, the consonant frequency distributions (bar graph) have peaks at frequencies increasing in the order jā, zā, shā. This matches the trend obtained from the results of listening tests, and shows that the frequencies used in these onomatopoeia are considered to vary according to the consonant. Also, for the onomatopoeia zā, it can be said that the magnitude of the consonant sound was smaller than that of the vowel and the effect of the
vowel is strongly expressed. Consequently, as in the results of the listening tests, it seems that the formant component of the vowel a is strongly expressed.

### 4.2 Onomatopoeia with the same consonant

Next, we investigate the onomatopoeic words shā and shū (Fig. 5(c) and (d)). In the listening test results, the onomatopoeia shū was considered to correspond to a higher frequency range than shā.

Figures 5(c) and (d) show that the consonant frequency distributions (bar graphs) all have high frequency peaks. In particular, shū has a large peak over a wide range of frequencies above 3 kHz. This is thought to be the reason why, with regard to the onomatopoeia shā and shū, the latter was reported to sound like higher frequencies in the listening tests.

The frequency distribution of the vowels (area graphs) shows that the vowel u in the onomatopoeia shū has a peak at low frequency that becomes gradually smaller with increasing frequency. The vowel a in the
onomatopoeia $sh\check{a}$ has a relatively high acoustic pressure at intermediate frequencies. In the onomatopoeia $sh\tilde{a}$, the consonant frequency component is masked by the vowel frequency component and the consonant has little effect, while in the onomatopoeia $sh\tilde{u}$, it appears that the high-frequency peak of the consonant is not masked by the frequency component of the vowel and the effects of the consonant are strongly expressed. As a result, although the vowel $a$ has components at higher frequencies than the vowel $u$, it seems that the overall effect of the onomatopoeia $sh\tilde{a}$ was considered to have a higher frequency than that of the onomatopoeia $sh\check{a}$ in the listening test results.

4.3 Onomatopoeia with the same vowel “$o$”

The onomatopoeia $b\tilde{o}$ and $g\tilde{o}$ consist of different consonants with the same vowel $o$. In the listening test results, these onomatopoeia were both regarded as having low frequencies.

From Fig. 5(e) and (f), it can be seen that both of these onomatopoeia have similar low-frequency peaks in the vowel frequency distributions (area graphs), but differ in terms of their consonant frequency distributions (bar graphs). However, the frequency distributions of the two consonants are smaller than those of the vowels, which means that the effects of the consonants are not readily expressed. This is thought to be the reason why there was not much difference between the two in the listening test results.

5 Summary

We have conducted listening tests to determine what sounds correspond to various onomatopoeic Japanese words. Since onomatopoeic words are spoken by humans, we studied them from a phonological viewpoint. We conducted a physical investigation of the relationship between onomatopoeia and the sounds they express, and we examined the relationship between the two. As a result, we have shown that sounds resembling the onomatopoeia we studied are governed by both the duration and the frequency characteristics of the consonant and vowel parts of the uttered sound of the onomatopoeia.

References


