A? WA? What Japanese phonetic sounds sound the same?

IZUMI Kyotaro OMIYA Nobuto OKADA Yuji
TOKUDA Riko RYU Fumihiro

Abstract

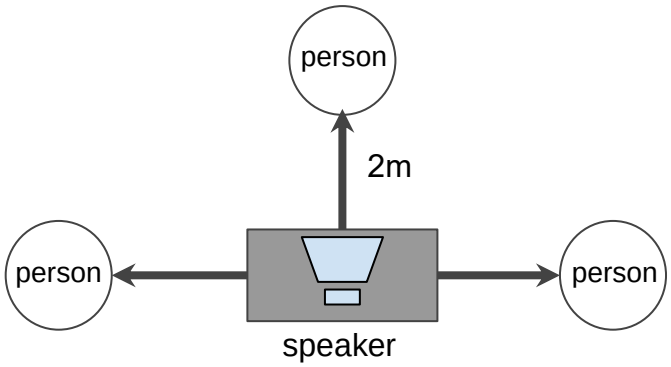
Our daily lives are directly impacted by the language we use everyday. But because of the COVID-19 pandemic, we were obliged to wear masks, which presented problems in daily communication. Even now, though COVID-19 is not considered a pandemic anymore, many Japanese people still choose to wear masks. Masks have great health benefits, but there are a number of drawbacks. One big problem is that many people cannot hear other people's voices as clearly as before. Therefore, we conducted various experiments on phonetic listening comprehension. Using Fourier transform, we discovered which Japanese syllables were difficult to hear, and analyzed syllables used daily in Japanese in order to find similarities among the syllables.

The COVID-19 pandemic has forced people around the world to wear masks. Under these unique conditions, there are few opportunities to see other people's mouths so it is difficult to read lips and thus distinguish what words people say. Likewise, masks muffle people's speech which reduces the ability to distinguish words in daily conversation. This has become a problem in our daily lives. We are taking the first steps to help improve the quality of daily communication in the age of the mask. By elucidating the causes of mishearing spoken words and the difficulty in hearing, we would like our research to aid future phonetic research such as voice recognition. If the sound frequency components are similar, then it is difficult to distinguish between the similar sounds. We want to reveal which syllables are difficult to recognize among the 50 Japanese monosyllables and why some syllables are more difficult to recognize than others.

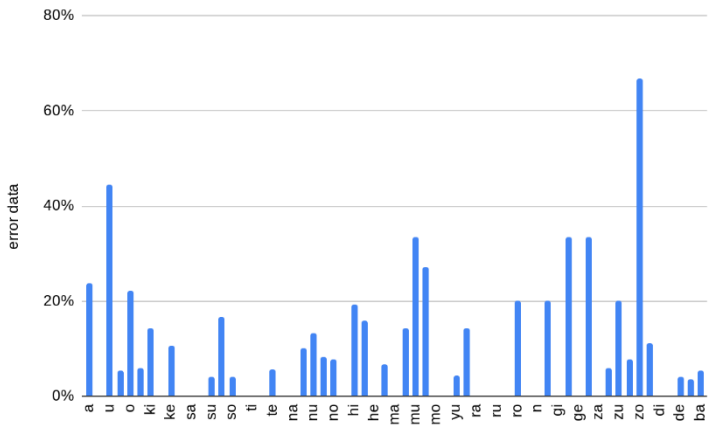
In the first experiment, we made a program which chose five random monosyllables from Python b.

1. By using this program, we played a series of five tones from a speaker and had the subject listen to them.
2. We had our test subjects listen to those five monosyllables. (Based on the references, place the distance interval for conversation at 2 m and the speaker volume at 60 decibels.)
3. Subjects were asked to write down on a white board what sound they heard.
4. Analyze data using a spreadsheet by conducting this experiment.

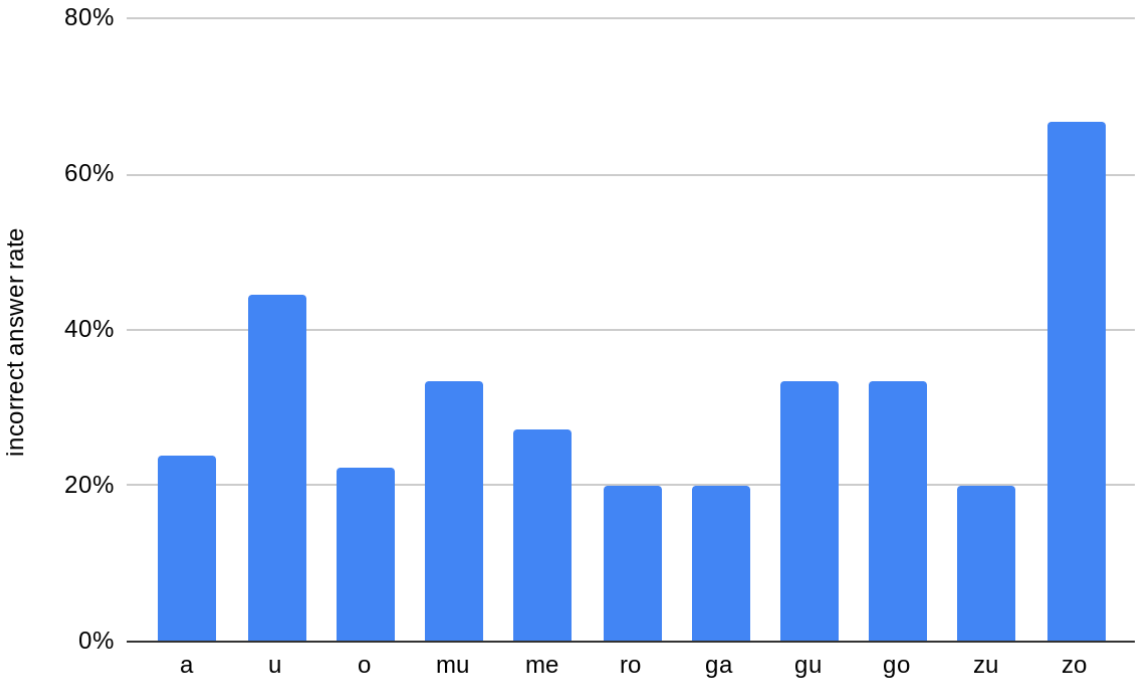
By conducting this experiment,we found there were eleven monosyllables that were often misheard by people. Those syllables included “A,” “U,” “O,” “MU,” “ME,” “RO,” “GA,” “GU,” “GO,” “ZU,” “ZO.”



(Figure1)state of experiment



(table1) The error percentage of all syllables



(table 2) Syllables where the error percentage exceeded 20%

Hypothisis
Syllables with the same vowel type, such as "a" and "ka," tend to be easily misheard. In addition, muddled and semi-muddled sounds are easily misheard.

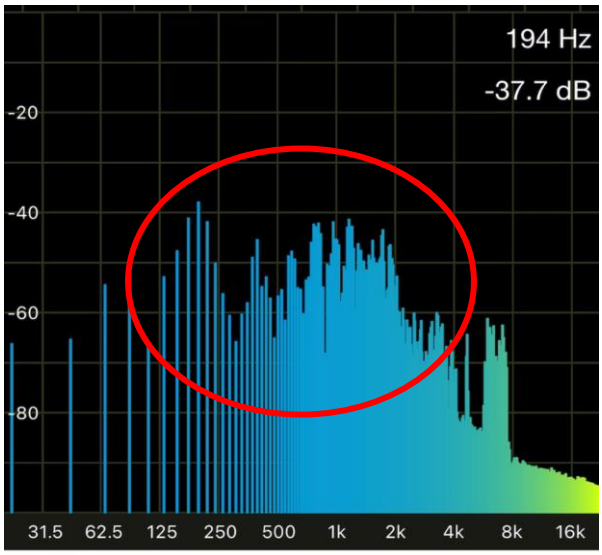
Results
Table 1 shows the results for all syllables, and Table 2 shows syllables where the error rate exceeded 20% or more.The vertical axis represents the error rate and the horizontal axis represents the different syllables.

Discussion
From Table 2, voiced sounds and vowels are considered to be the most difficult syllables to recognize in the Japanese language. A significant difference was observed between the syllables that were frequently misheard and those that were often heard correctly.

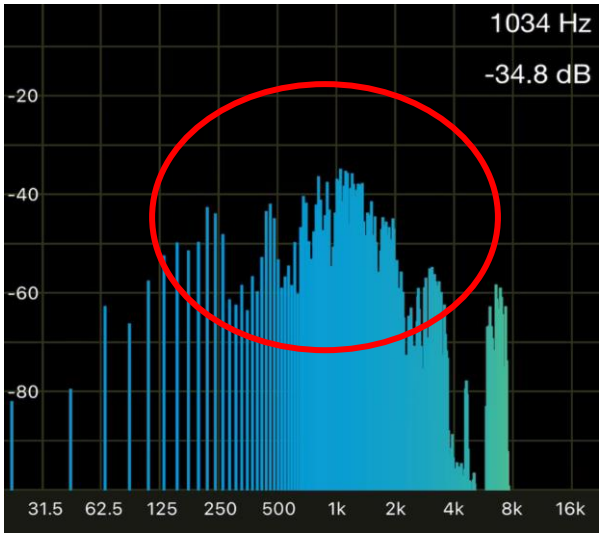
Second experiment
We researched why monosyllabic words make a difference. First, we measured the sound waves of all 50 monosyllables with an application, and converted these sound waves into audio spectrums. The audio spectrum is a graph that breaks down a syllable wave into its frequency components by Fourier transform and arranges the magnitude of each component. We compared the shape of each audio spectrum, especially 400 Hz to 2kHz, because humans can easily hear the frequency of sound waves in these Hertz.

Hypothesis
It is known that if the sound waves are similar, the tones are similar. Therefore, we think if the spectra extracted from the sounds are similar, then the tones will be similar. For this reason, we assumed that the audio spectra of the misheard syllables and that of correct syllables would be similar.

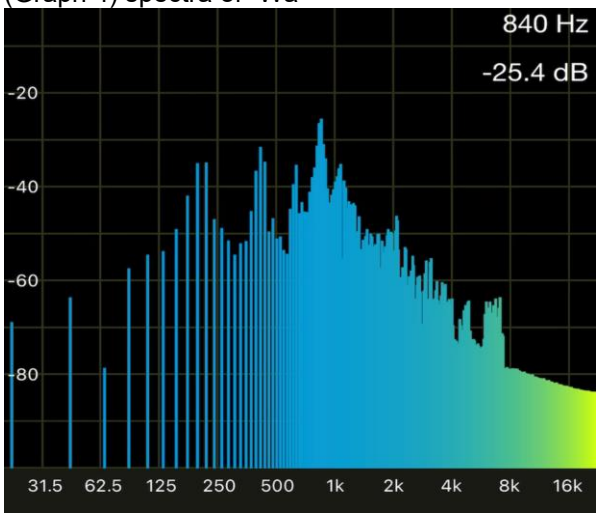
Results and Discussion
1.Comparison of “A” , “Wa” and “Za”



(Graph 3) spectra of “A”



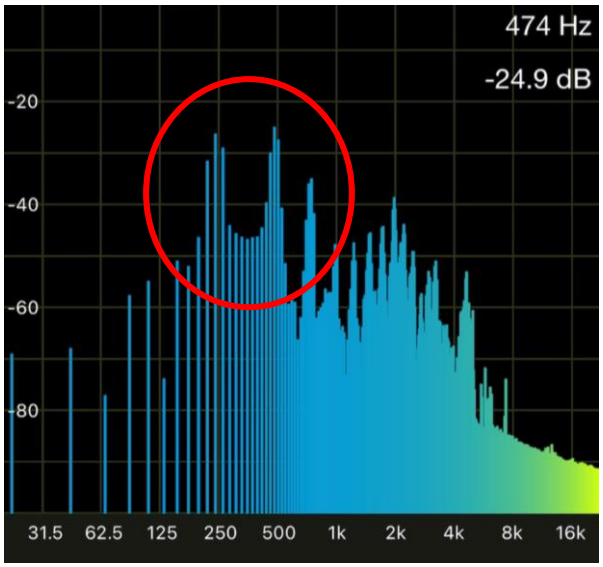
(Graph 4) spectra of “Wa”



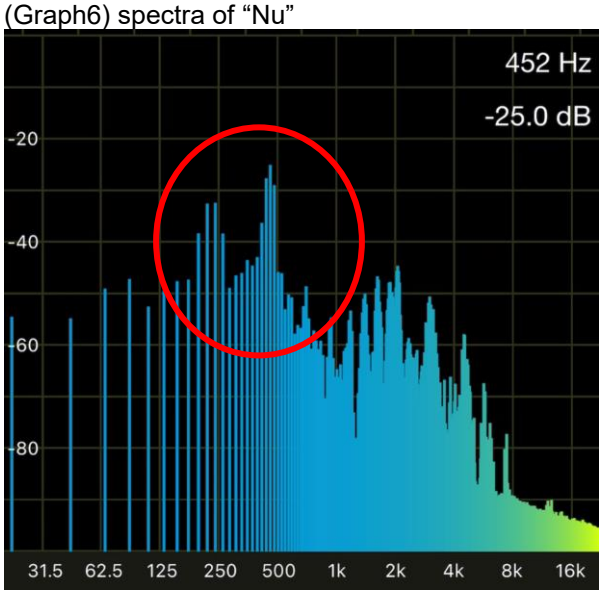
(Graph 5) spectra of “Za”

We determined that A’s spectrum is similar to Wa’s spectrum from the red circles. However people didn’t mishear “A” for “Za”, the graph 3 and 5 showed A’s spectrum isn’t similar to Wa’s spectrum. Therefore, we thought people often misheard “A” for “Wa” because their spectrum is similar.

2.Comparison of “Nu” and “Mu”

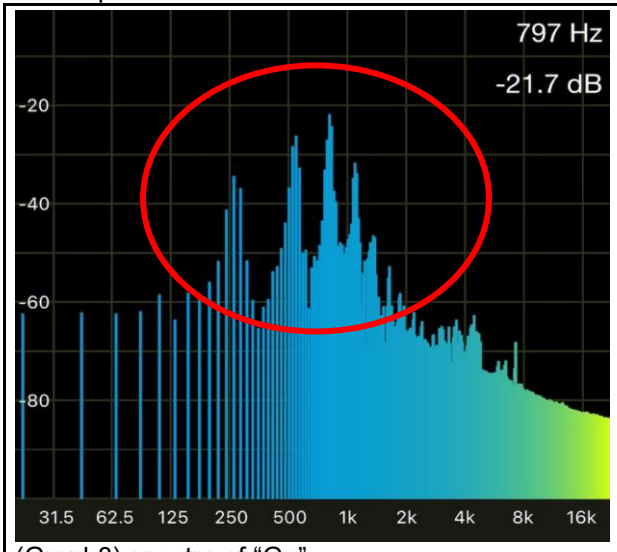


(Graph6) spectra of “Nu”

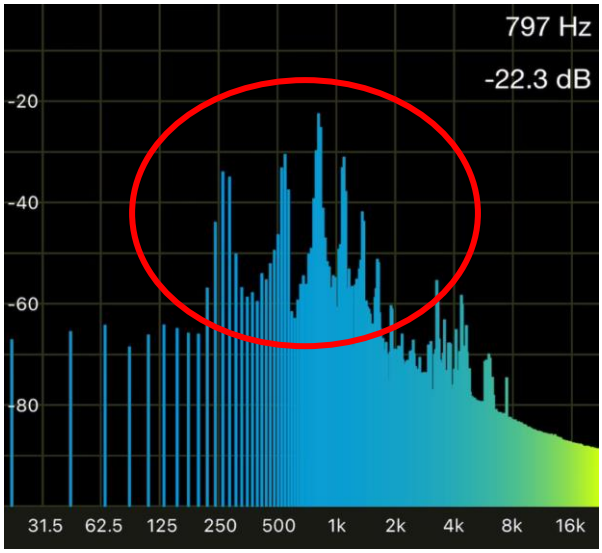


(Graph7) spectra of “Mu”

3.Comparison of “Go” and “Bo”

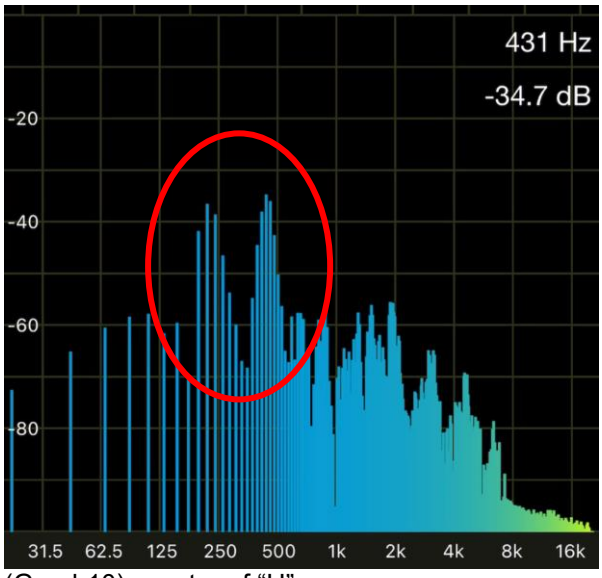


(Graph8) spectra of “Go”

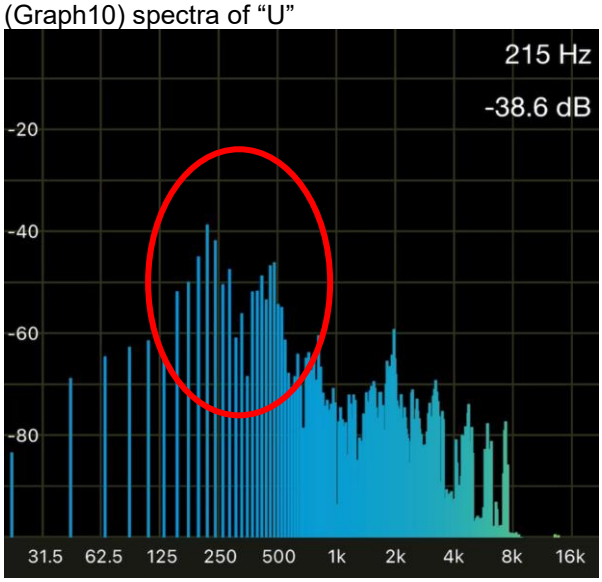


(Graph9) spectra of “Bo”

4.Comparison of “U” and “Yu”



(Graph10) spectra of “U”

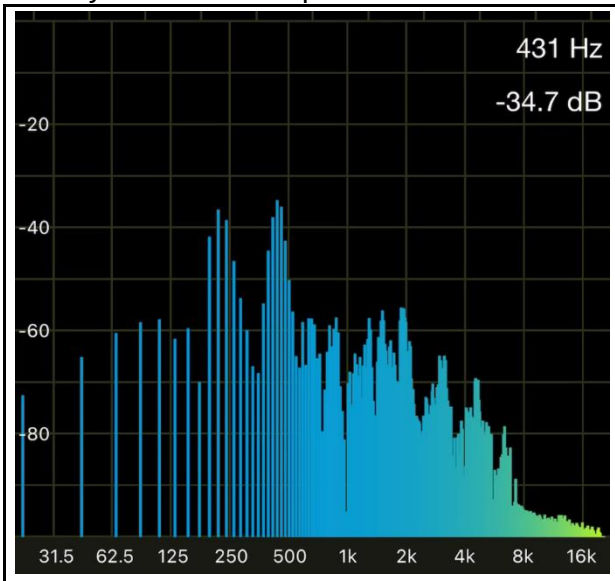


(Graph11) spectra of “Yu”

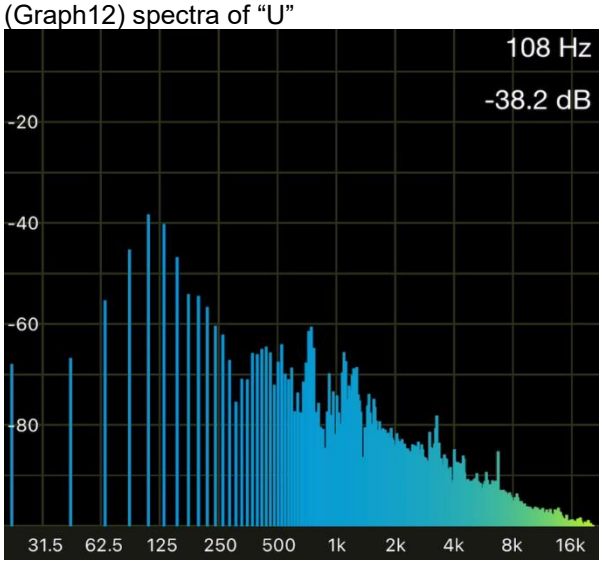
In the same way, we compared “Nu” with “Mu”, “Go” with “Bo”, and “U” with “Yu”.

5. Future Prospects

1.Syllables whose spectra are not similar but misheard

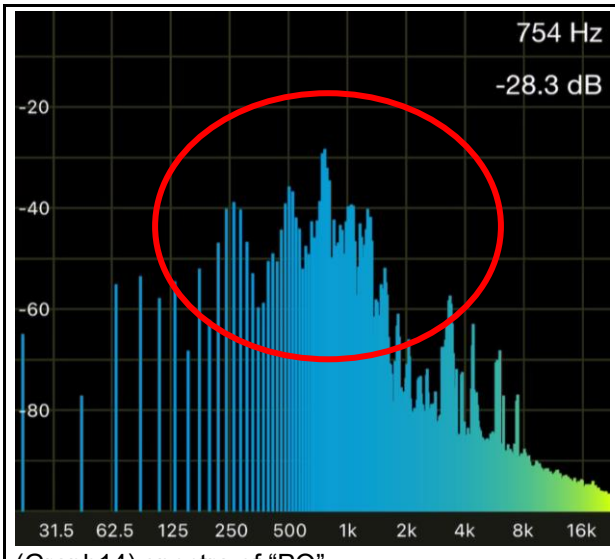


(Graph12) spectra of “U”

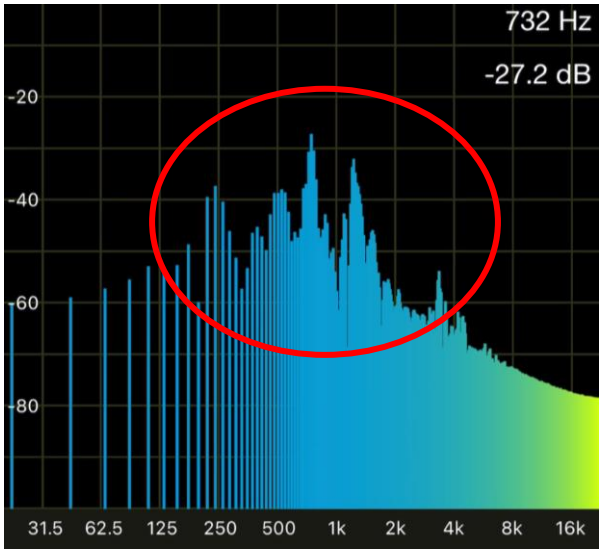


(Graph13) spectra of “O”

2.Syllables with similar spectra but not misheard



(Graph14) spectra of “PO”



(Graph15) spectra of “KO”

Conclusion

From the first experiment of searching which syllables are difficult to hear,we found that syllables that have the same vowel are difficult to recognize.By second experiment comparing the syllable spectra,we discover syllables which misheard has similar spectra shape.also it is difficult to distinguish between sounds with similar spectra around 450 hz. There were some exceptions: syllables that were misheard even though the spectra were not similar ("u" and "o") and others that were not misheard even though the spectra were similar ("po" and "ko"). We would like to understand more clearly the reasons for this.

To quantify this, we would like to compare not only visual "similarity" judgments when looking at speech waveforms, but also other factors. We would also like to examine the influence of syllable-to-syllable connections on the ease of mishearing.