慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

| | |
|---|---|
| Title | I'd like to unravel the mechanisms of our speech and hearing : investigating what characterizes word recognition |
| Sub Title | |
| Author | 田井中, 麻都佳(Tainaka, Madoka) |
| Publisher | Faculty of Science and Technology, Keio University |
| Publication year | 2017 |
| Jtitle | New Kyurizukai No.26 (2017. 11) ,p.2- 3 |
| JaLC DOI | |
| Abstract | |
| Notes | The Research |
| Genre | Article |
| URL | https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO50001003-00000026-0002 |

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

The Research

Associate Professor Yukiko Sugiyama is featured in this issue,
whose field of research is phonetics focusing on the mechanism of speech and hearing.

# I’d like to unravel the mechanisms of our speech and hearing

## Investigating what characterizes word recognition

**We utter words when speaking, but it’s impossible to physically utter the same words twice even if the words are the same. If so, how do we recognize the words spoken to us, understand their meaning and communicate with one another? Or is it possible to distinguish words such as “*hashi* (bridge)” and “*hashi* (edge)” that are pronounced likewise but have different meanings in Japanese? Associate Professor Yukiko Sugiyama approaches the process and mechanism of speech production and perception from both aspects: words uttered by the speaker and their perception on the part of the listener.**

### What is “phonetics”?

Dr. Sugiyama’s specialty is a field of study known as “phonetics.” Phonetics is largely classified into three academic areas (Fig. 1): “acoustic phonetics” examines physical properties of spoken words; “articulatory phonetics” analyzes how speech sounds are produced in the oral cavity when humans speak; and “perceptual phonetics” investigates the process by which humans perceive speech.

“Phonetics is often regarded as a branch of linguistics. However, in order to examine speech, we need to identify its physical characteristics such as duration, frequency and intensity. Also, articulation deals with the workings of the oral cavity and vocal folds while speech perception concerns the human sensory mechanism. These factors require knowledge from a wide range of disciplines including physics, engineering, medicine and cognitive psychology, among others. It is indeed a multidisciplinary field involving both humanities and sciences,” explains Dr. Sugiyama.

With phonetics as the base, Dr. Sugiyama takes a two-way approach in proceeding with her research. One way is to analyze the physical characteristics of speech, and the other is to examine how a person perceives speech. By using this two-way approach she’d like to unravel characteristics of the Japanese language.

### Looking for characteristics that distinguish words

“The target I use for this purpose is Tokyo Japanese, or so-called the Standard Japanese. I collect and record samples of speech from Tokyo Japanese speakers. To begin with, I examine the physical characteristics of speech such as the frequencies and durations of speech segments. In the case of the Tokyo Japanese, for example, the word ‘*ame*’ (rain) is pronounced with a higher pitch for ‘*a*’ and a lower pitch for ‘*me*.’ On the other hand, the word ‘*ame*’ (candy) is pronounced with a lower pitch for ‘a’ and a higher pitch for ‘me.’ In other words, the pitch levels of high and low determine the meaning of words.”

But what if it comes to “*hashi*” (meaning “bridge” and “edge”) and “*tori*” (meaning “bird” and “last performer”)?

“Both words are pronounced with the same pitch pattern of low-high, making it difficult to distinguish them. However, when you say ‘*hashi o aruku*’ (the former meaning ‘walk over a bridge’ and the latter ‘walk along the edge’), the postpositional article ‘o’ that follows ‘*hashi*’ is pronounced with a low pitch for the former and with a high pitch for the latter. By putting words of interest in an environment where they minimally differ, we find out the characteristics that people use to identify words,” she continues.

As a matter of fact, if we analyze the frequency components in one’s speech and look at their spectrogram – the so-called “voiceprint” – we see rises and falls of the fundamental frequency (the rate at which the vocal folds vibrate per second, which we perceive as pitch) which serve to distinguish words (Fig. 2). Thus, in Japanese, we use pitch accent to distinguish one word from another.

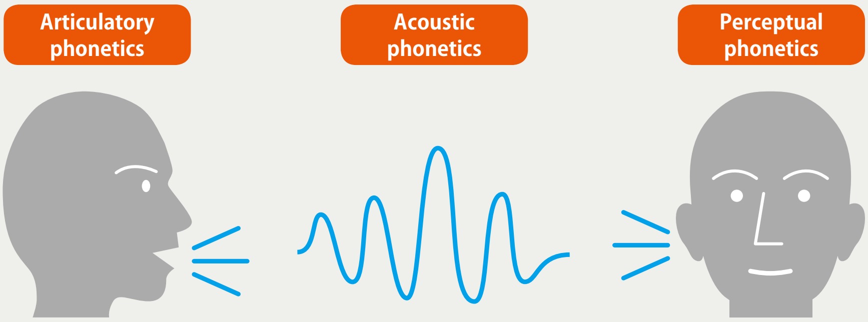
“In terms of distinguishing words by the movement of fundamental frequency, Japanese is similar to Mandarin Chinese, which is classified as a tone language. Meanwhile, the function of pitch in Japanese is similar to that of stress in English.”

Some propose that Japanese pitch accent is characterized not only by fundamental frequency but also by intensity and duration of segments as is typically observed in English and other stress accent languages.

“I don’t think that fundamental frequency alone is sufficient to distinguish between words in robust communication. In fact, English stress accent includes multiple elements such as intensity, duration and pitch. However, the meanings of Japanese words change if segment durations change. Then what elements can be used as acoustic correlates of pitch accent in Japanese?

**Fig.1 Phonetics**

Phonetics is largely classified into three areas as shown below:



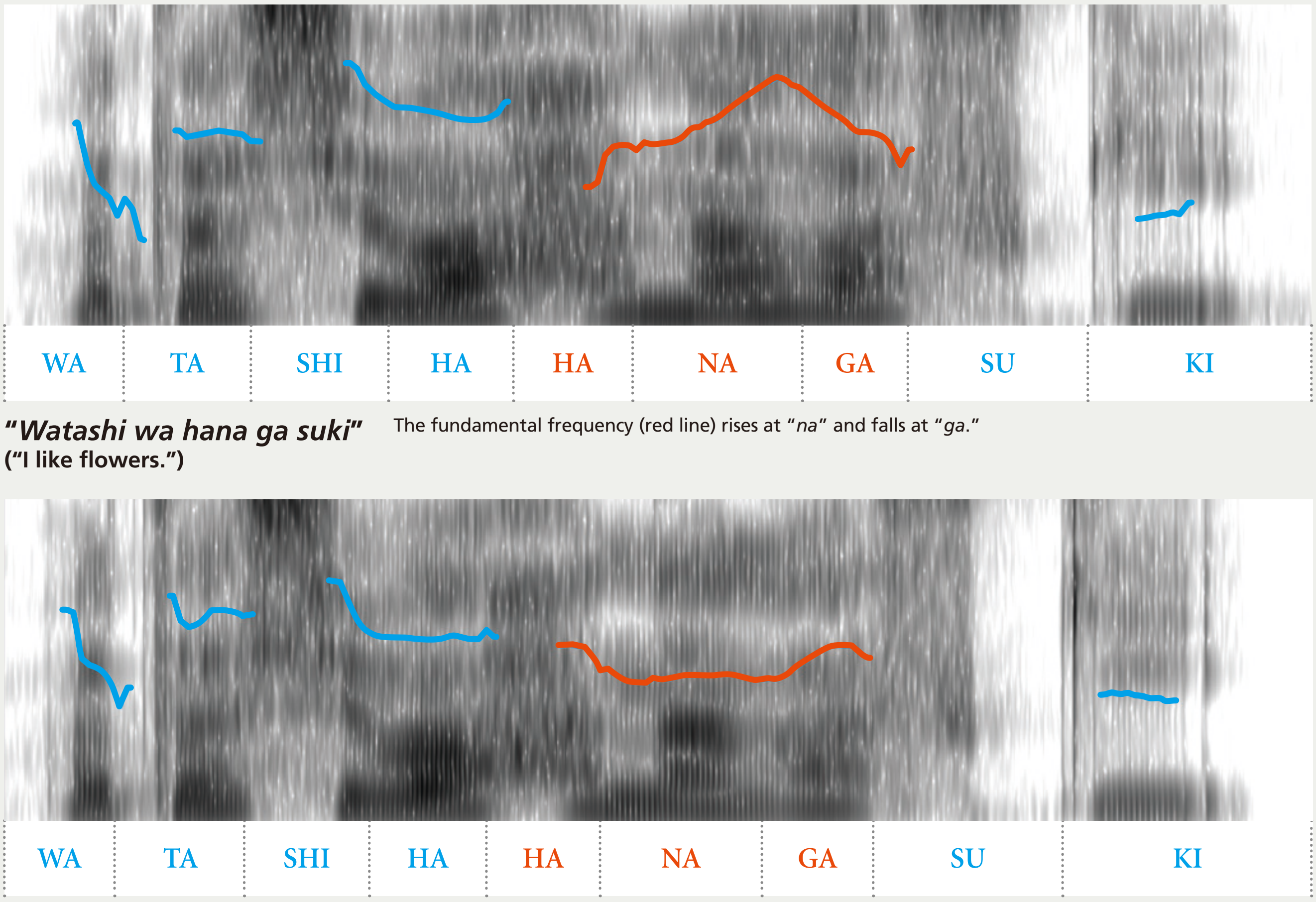
Analyzes how words are produced in the speech organ.

Physically analyzes speech as vibrations of air.

Analyzes how speech is perceived.

**Fig.2 Voice pitches and difference in meanings**

The dark areas with vertical striations are spectrograms (the so-called voiceprint). The vertical axes indicate frequencies (Hz). In the spectrograms, the darker an area, the greater the amount of energy. The blue and red lines lying on top of the spectrograms show pitch contours (Hz).



***"Watashi wa hana ga suki"***
**("I like flowers.")**

The fundamental frequency (red line) rises at "*na*" and falls at "*ga*."

***"Watashi wa hana ga suki"***
**("I like a nose.")**

The fundamental frequency (red line) remains relatively flat (strictly speaking, slightly rises at "*ga*") in the area from "*na*" to "*ga*".

This weighs on my mind," Dr. Sugiyama remarks.

To address this problem, Dr. Sugiyama conducts perception experiments using edited speech from which the fundamental frequency has been artificially removed. If listeners can successfully distinguish words such as "*hashi*" (bridge) and "*hashi*" (edge), even when there is no fundamental frequency, it would suggest that acoustic cues other than the fundamental frequency are present in the speech, enabling the listeners to use them to identify the words they heard.

Dr. Sugiyama says, "The results found that the listeners were over 95% correct in word identification when they heard natural speech. For the edited speech which contained no pitch information, the accuracy dropped to roughly 65%, but it was above chancel level. This leads to a conclusion that Japanese pitch accent is realized by certain other acoustic characteristics in addition to the fundamental frequency."

For future research, she would like to identify exactly what acoustic characteristics listeners use to identify words when there is no fundamental frequency.

## Would like to contribute to machine-based speech recognition and synthesis

In what way do these studies benefit us socially and academically?

"Academically, I think my research would contribute to a better understanding of the possible prosodic types that human language can have by revealing the acoustic details of Japanese pitch accent."

"I think it would also contribute to improving speech recognition systems and speech synthesis by indicating what acoustic correlates accompany pitch. In order to raise the accuracy of these systems, are there any other acoustic elements that need to be taken into consideration? If we can find an answer to this question, it will also help to synthesize more human-like speech," remarks Dr. Sugiyama.

While hearing aids and cochlear implants are very helpful to those who need them, their performance is still far from that of an actual human ear, causing difficulty in sensing pitch, having a narrower dynamic range, and introducing noise into what we actually want to hear. This is why hearing performance closer to the human ear is sought after.

"Also, the ability to recognize speech is known to vary largely from one person to another and much remains unsolved. For example, you can hear your name mentioned somewhere all of a sudden even when you are talking to someone in a noisy environment, a phenomenon known as the 'cocktail party effect.' Individual differences in perception mean that there is much more to be understood about the physiological details of pitch perception. To address these questions, it is necessary to collaborate with researchers from the engineering field, which will greatly help to formally characterize the acoustic details of speech."

In order to learn the methods used in signal processing, Dr. Sugiyama has sat in on an applied mathematics class together with second year students and gets help from a student whenever she has questions from the class. Dr. Sugiyama's challenge continues.

(Reporter & text writer : Madoka Tainaka)