

|                  |   |
|------------------|---|
| Title            | 人間が話したり、聞いたりするしくみを解き明かしたい：<br>何が単語の認知を特徴づけているのかを探る  |
| Sub Title        |   |
| Author           | 田井中, 麻都佳(Tainaka, Madoka)   |
| Publisher        | 慶應義塾大学工学部   |
| Publication year | 2017  |
| Jtitle           | 新版 窮理図解 No.26 (2017. 11) ,p.2- 3  |
| JaLC DOI         |   |
| Abstract         |   |
| Notes            | 研究紹介  |
| Genre            | Article   |
| URL              | <a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO50001002-00000026-0002">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO50001002-00000026-0002</a> |

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.

# 人間が話したり、聞いたりする しくみを解き明かしたい

何が単語の認知を特徴づけているのかを探る

人は話すとき、同じことばであっても、物理的にまったく同じ音声を発することはできない。では、なぜ私たちはそれを聞き取り、意味を理解し、意思の疎通を図ることができるのだろうか。あるいは、「橋」と「端」のように、同じ発音の単語の意味を汲み取ることができるのだろうか。杉山由希子准教授は、話し手の音声、聞き手の聴覚の両面を調べることで、音声を話したり、聞いたりするプロセスやそのしくみの解明に取り組んでいる。

## 「音声学」とは？

杉山さんの専門は、「音声学」と呼ばれる学問分野だ。音声学は大きく分けて、人がしゃべった言葉の音波を調べる「音響音声学」、人間が言葉をしゃべる際に口の中で音を整える調音について研究する「調音音声学」、そして聞く側について調べる「聴覚音声学」の3つの分野からなる(図1)。

「音声学は、言語学の一分野と捉えられることもあります。音波を調べるためには音声の周波数や時間の長さなど、物理的な特徴を調べなければなりません。また、調音であれば口腔や声帯のはたらき、聴覚であれば人間の知覚のしくみも関係してくることから、工学や医学、認知心理学など、幅広い知識が必要になります。文理両方の要素を併せ持つ学問分野なのです」と杉山さんは説明する。

その音声学をベースに、杉山さんは2つのアプローチで研究を進めている。1つは、音声が物理的にどのような特徴を持っているかを調べることで、もう1つはそれを人がどのように聞いているかを、実際に聞き取りの実験を行って調べるという方法だ。こうした双方向のアプローチにより、日本語が持つ特徴を明らかにしたいという。

## 単語を区別している特徴を探る

「対象にしているのは、東京方言、いわゆる標準語です。複数の東京方言話者から音声を録音して集め、まずは音声の周波数や長さなど、物理的な特徴を調べます。たとえば東京方言の場合、『雨』は『あ』の音が高く、『め』が下がりますが、『飴』では『め』を高く発音します。このように、音声の周波数成分に由来す

る音の高低(ピッチ)が単語を意味づけています」

しかし、「橋」と「端」ではどうだろうか。あるいは「鳥」と「トリ(最後)」では？

「どちらも同じ低高のパターンで発音するため、単語からだけでは区別が付きません。ところが、『橋を歩く』と『端を歩く』では、『はし』の後ろにつく助詞の『を』のピッチが前者は低く、後者は高いまま発音されます。このように、人が何を手がかりにして単語を理解しているのか、その特徴を探るのです」。

実際に、音声に含まれる周波数成分を分析して可視化したスペクトログラム、いわゆる「声紋」を見ると、音の高低を聞き分ける際の手がかりになる基本周波数(基底となる周波数成分)の上がり下がりが見てとれる。日本語では、このように音のピッチで単語を区別する、ピッチアクセントを頼りに意味を汲み取っていると考えられる(図2)。

「音声の高い低いで単語を区別するという点では、日本語はトーン言語に分類される北京語に似ています。一方で、日本語のピッチの持つ役割は、英語のストレスアクセント(強勢)に近いものでもありません」。

そうしたことから、日本語のピッチアクセントには、音の高低だけでなく、英語に代表されるストレス言語と同様に、音の強さや長さも関係しているのではないかとされることもある。

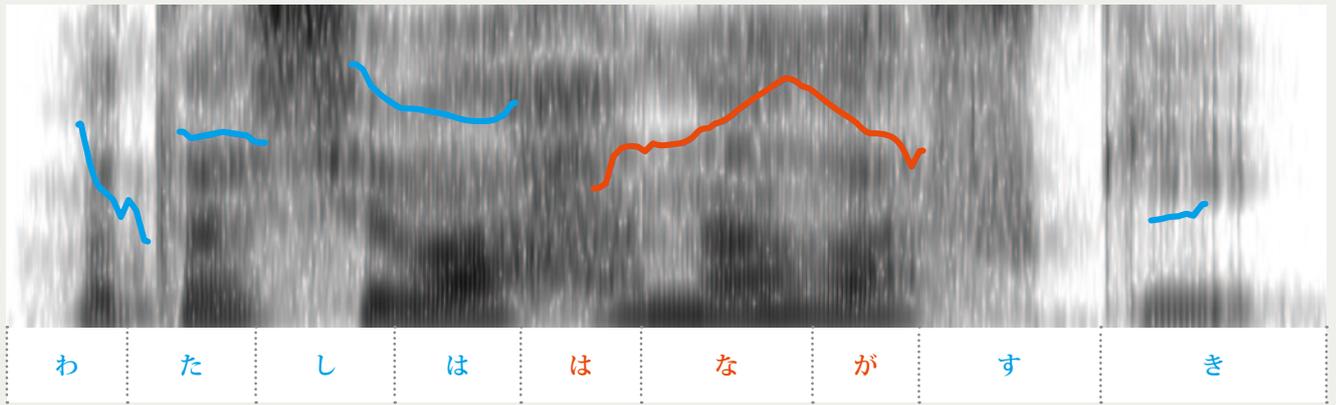
「単語の区別をピッチアクセントだけに頼っていたのでは、さまざまな状況下でのコミュニケーションにおけるロバストネス(頑健性)を担保しきれないように思います。実際に、英語のストレスアクセントには、強さ、長さ、ピッチと

## 図1 音声学

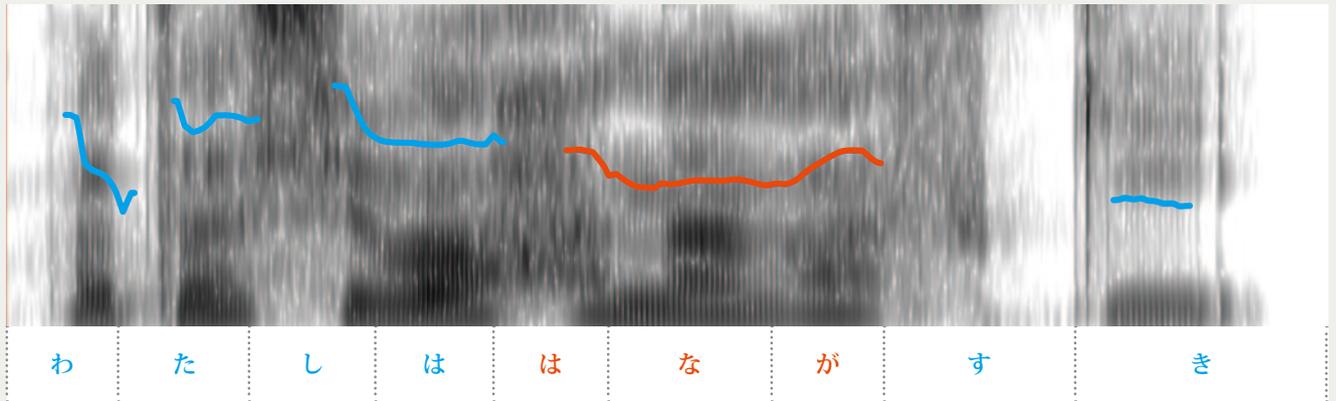
音声学は、大きく分けて図のように3つの分野に分けられる。



**図2 声の高さと意味の違い** 図の黒い部分はスペクトログラム（いわゆる声紋）で、縦軸は周波数（Hz）である。色の濃い部分は、エネルギー量が多いことを示しており、青い線と赤い線は声の高さ（Hz）を示している。



「私は花が好き」 声の高さ（赤線）を見ると、「な」で上がり、「が」で下がっているのが分かる。



「私は鼻が好き」 声の高さ（赤線）を見ると、「な」から「が」にかけて比較的平坦（厳密に言うと「が」で少し上がっている）なのが分かる。

いった複数の要素が含まれます。しかし、日本語の場合は長さが変わると意味も変わってしまう。では、何を使っているのか、とても気になるところです」。

そこで、杉山さんは、声帯の震えを伴わないさやき声や、音声の周波数成分の中から基本周波数などを人工的に取り除いた、ピッチが聞き取れないはずの加工された音声を使って、それでも単語を聞き取ることができるかどうかの実験をしている。

「実験の結果、自然な音声では正答率が95%を超えますが、基本周波数とその倍音を除去した合成音声でも正答率は65%近くにのぼり、偶然に当たるレベルを超えています。このことから、日本語のピッチアクセントにも基本周波数以外に、単語を区別する何らかの音響特徴があると考えられます」。

今後は、ではなぜ基本周波数が存在しなくても声の高低が聞き取れるのか、その手がかりとなっている音響特徴を同定

してゆきたいという。

### 機械による音声認識や 音声合成に貢献したい

ところで、こうした研究は、学術的、あるいは社会的にどう役立つのだろうか。

「1つは、日本語が他言語と比較してどのような特徴を持っているのか、音声学の側面から類型化できればと考えています」。

もう1つ、現在、機械による音声の認識・合成の活用が始まっていますが、この精度を上げるために、どのような要素に着目すれば、単語の認識率がより高くなるのか、あるいは人間らしい声を再現できるのか、といったことにも役立てられるでしょう」と杉山さん。

たとえば、人工内耳や補聴器による「聞こえ」は、人間の耳の性能に比べるとまだまだ劣る。ピッチが感じ取りにくかったり、ダイナミックレンジが狭かっ

たり、耳障りな音が聞こえたりして、より人間の耳に近い聞こえが求められているのだ。その際に、単語を理解する上で役立つ特徴を強調できれば、よりよい聞こえに貢献できるだろう。

「ただ、音の認知は個人差が大きく、まだまだ解明されていないことがたくさんあります。カクテルパーティ効果といって、うるさい場所でも自分の名前だとパッと聞き取れたり、同じ音を提示しても、人によって音の高さが違って聞こえたり、解明されていない謎はたくさんあります」。

そうした謎に迫るため、工学系の研究者と組むことによって、音声の物理的特徴をより詳細に探る必要があると思っています」。

そのために、杉山さんは信号処理を学ぼうと、学生と肩を並べて授業を聴講し、ときには学生からも教わるという。真理の追究のため、杉山さんの挑戦は続く。

（取材・構成 田井中麻都佳）