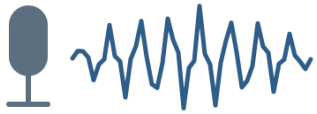
音声認識のしくみ（声の波から文字へ）

機械は、声をどのようにして文字にしているのだろう

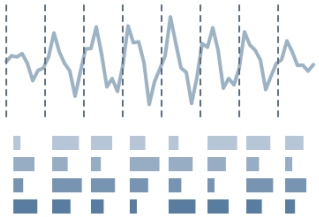
「あしたこうえんにいく」

1 声をひろう



マイクで、声の波を
とらえる

2 特徴を数にする



短く区切って、音の
特徴を数で表す

3 音を当てる

a　sh　i　t　a

k　o　u　…

特徴から、どの音かを
推定する

AI

4 ことばを選ぶ

同じ音でも候補が複数

あした公園に行く 0.81

あした講演に行く 0.11

あした公演に行く 0.08

ことばのつながりの
自然さで選ぶ

5 文字にする

**あした公園に
行く**

今は深層学習で、
いくつかの段をまとめて
一度に変換するしくみもある

聞き取りがずれやすいとき

周りの音が大きい

話し声や車の音などが
まざる

早口・小さい声

音の特徴がはっきり
しない

似た音のことば

こうえん（公園・講演・
公演）など

学んでいないことば

新しいことばや、
話し方のちがい

※ 模式図。例と確からしさの数は架空。しくみは、情報通信研究機構（NICT）の解説をもとにした。