

21

どんな音？物音を言葉で説明しよう

～系列変換モデルに基づく音響信号からの説明文生成～

どんな研究

これまでは、音声認識システムを用いても、人の話し声以外の音を的確にテキストに変換することはできませんでした。本展示では、**任意の音**に対して、それが**どんな音か**を説明するテキスト（**擬音語**や**説明文**）を生成する研究を紹介します。本研究により物音などを文字で伝えることが可能になりました。

どこが凄い

音を文字にすることで、**一見するだけでどのような音かを把握**できます。どの程度の**詳しくさ**で音を説明して欲しいかを数値で指定することで**目的に沿った文を生成**できます。擬音語や説明文が表現するような**主観的な音の近さ**に基づいて、イメージした音に最も近い音を**検索**することなどにも応用できます。

めざす未来

効果音や異常音などの**音の検索**がより便利になります。**動画中の音**を文字で表現することで、公共の場所や騒々しい場所などでの**動画視聴**も便利になることでしょう。また、**AIが人間に近い音の感覚を身につける**ことで、宅内ロボットなどでの日常のコミュニケーションが円滑になることも期待されます。

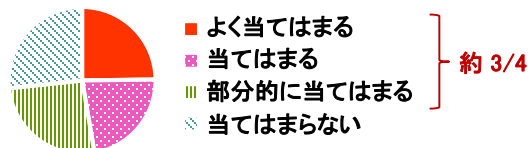
音の説明文生成とは

- 任意の音に対して **どんな音か**を言葉で説明する深層学習システム
- 「ベル」「車」といった単なる音源種類の認識（＝従来研究あり）だけでなく **変化の様子**なども含む詳しい記述も可能 **NEW**



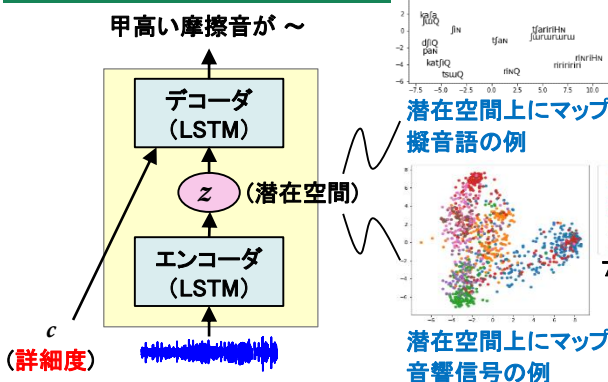
実験結果（説明文モードの例）

説明文が音に...

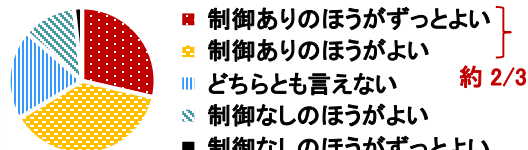


説明文の適切さ（詳細度制御なしの場合）

系列変換モデルによる実現



詳細度制御は...



説明文生成における詳細度制御の効果

生成された説明文の例（バスドラムの音）

c	生成された文
なし	低い音が一瞬だけ響く
20	低い音が一瞬だけ鳴る
50	量に打ち付けるような低い打撃音が一瞬だけ響く
80	つまらなそうにとでも低い音でドラムが一度だけ奏でられている
110	つまらなそうにとでも低い音で何かを叩くような軽くて少し低い音が小さな音で一瞬鳴ってすぐに消える

- 擬音語モード: 擬音語（音素(カナ)の系列）を出力
- 説明文モード: 説明文（単語の系列）を出力

【課題】文による説明のしかた(詳しくさ)に唯一の正解はない
 →【提案】**詳細度**により説明の詳しくさをコントロール
 詳細度: 単語単位で計算した“情報量”の総和
 文と詳細度の双方の誤差を小さくするように学習する

※本展示は 東京大学大学院情報理工学系研究科システム情報学専攻 中村宏教授との共同研究成果を含みます。

関連文献

- [1] Shota Ikawa, Kunio Kashino, "Generating sound words from audio signals of acoustic events with sequence-to-sequence model," In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)*, April 2018.
- [2] Shota Ikawa, Kunio Kashino, "Acoustic event search with an onomatopoeic query: measuring distance between onomatopoeic words and sounds," In *Proc. Detection and Classification of Acoustic Scenes and Events (DCASE 2018)*, November 2018.

連絡先

柏野 邦夫 (Kunio Kashino) メディア情報研究部
 Email: cs-liaison-ml at hco.ntt.co.jp



Innovative R&D by NTT
 オープンハウス 2019