ABCI 3.0開発加速利用 （2025年度）成果概要（公開用）

課題名：	実施時期：2025/4～2026/3 所属機関名：産業技術総合研究所 代表者氏名：濱崎 雅弘
メディアインタラクション技術に関する研究開発	

成果概要：

人々の生活の豊かさの向上に資するメディアインタラクション技術の実現を目指して、メディアコンテンツの解析・理解・生成・活用を行うアルゴリズム・技術等の研究開発を行った。2025年度は、歌詞のテキストと背景音楽を含む歌声の音響信号との自動対応付け（歌詞アラインメント）の精度を向上させる新手法を開発した。

成果のポイント：

カラオケ字幕や歌詞アニメーション動画を生成するために、歌詞アラインメント技術は不可欠な基礎技術である。本研究では、単語レベルの正解時刻（アノテーション）しかない場合であっても、部分的に音素単位の正解時刻が分かる点に着目し、それを活用する深層学習モデルを提案してABCI3.0によって学習することで、従来手法を上回る精度を実現した。具体的には、確定時刻（フレーム）のみで交差エントロピー損失を計算する手法（masked cross entropy loss）により、不正確なラベルの影響を排除して部分的なアノテーションを効果的に活用した。査読付き国際会議 DAFx25にて発表。

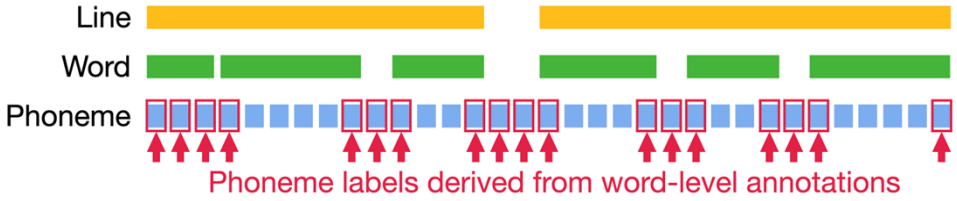

図1: 単語レベルのアノテーションから音素レベルの部分アノテーションを構築

Word	Phonemes	Phoneme labels	Onset	Offset
I	AY	6	0.087 seconds	0.184 seconds
feel	F IY L	14 18 21	0.281 seconds	0.377 seconds
like	L AY K	21 6 20	0.474 seconds	0.571 seconds

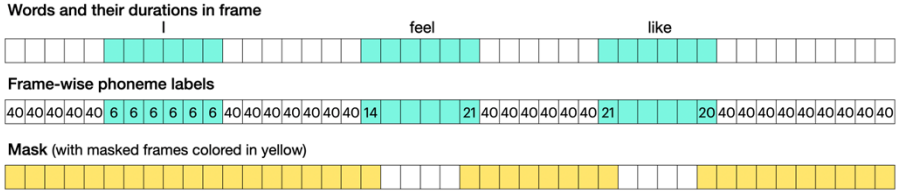

図2: フレーム単位の音素ラベルと不確定なフレームを区別するマスクを構築

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：

Tian Cheng, Tomoyasu Nakano and Masataka Goto, "Improving lyrics-to-audio alignment using frame-wise phoneme labels with masked cross entropy loss,“ In Proc. 28th International Conference on Digital Audio Effects (DAFx25), pp.342-349, 2025.
https://dafx25.dii.univpm.it/proceedings/