

課題名： 少量データで性能を発揮するAI開発のための音声音響基盤モデルの構築と利活用	実施時期：2025年4月1日-2026年3月31日 所属機関名：国立研究開発法人 産業技術総合研究所 代表者氏名：深山 覚
---	---

成果概要：
 自己教師あり学習モデルはラベルなし音声データから汎用的な特徴表現を学習し、音声認識などの下流タスクに高い性能をもたらす。一方で大量のデータ、大規模なストレージ、長いデータ処理時間が必要である。さらに自己教師あり学習モデルの音声言語の種類が下流タスクに影響することが知られている。そこで大規模計算資源であるABCI3.0を活用し、低コストな自己教師あり学習モデルの構築と、構築したモデルにおける学習データの音声言語の種類が下流タスクに与える影響の検証を行った。

成果のポイント：

① **ニューラルオーディオコーデックを用いた低コストな自己教師あり学習モデルの構築**
 ニューラルオーディオコーデック (NAC) を用いた自己教師あり学習モデルNAC HuBERTは、事前学習においてストレージおよび計算コストを削減し、音声認識・音声感情認識においてNACを使わない場合と同程度の性能を発揮した。一方、NACによる音声圧縮に起因して性能がわずかに低下することが確認された。

表1 NAC HuBERT 事前学習における計算時間と GPU 使用量 (括弧内は削減率)

SSL	言語	データ形式	サイズ [GiB]	学習時間 [ms]	GPU 使用量 [MiB]
HuBERT	英語	FLAC	57.98	529.94	94713.8
	日本語	WAV	517.39	502.79	98662.5
NAC HuBERT	英語	離散トークン	5.8 (-90%)	430.32 (-18.8%)	82407.7 (-13%)
	日本語	離散トークン	29.8 (-94.2%)	424.73 (-15.5%)	83426.9 (-15.4%)

② **自己教師あり学習モデルにおける学習データの音声言語の種類の下流タスクへの影響を検証**
 下流タスクの言語と同じ言語で事前学習したNAC HuBERTを用いた場合に、下流タスクの性能が最も高くなることがわかった。一方でNACの学習に用いる言語の違いによる性能差は限定的であった。NAC HuBERTの構築においては下流タスクと同じ言語の事前学習が有効である一方、NACについては特定言語にあわせて再学習の必要は低いことがわかった。

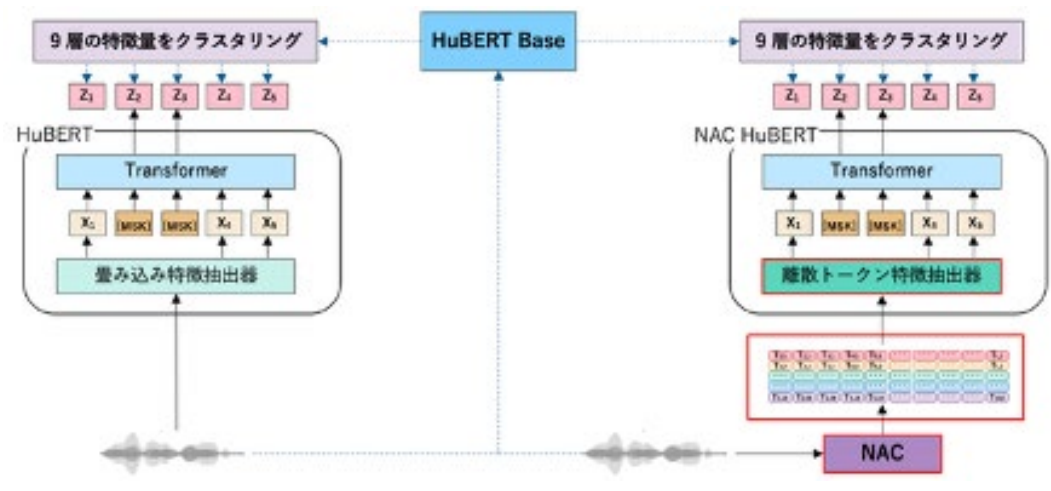


図1. ニューラルオーディオコーデックを用いた自己教師あり学習モデルNAC HuBERTの構築法

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：

- Neural Audio Codecを用いた低コストな自己教師あり学習モデルの検討, 瀧澤大吾, 中村友彦, Samuele Cornell, William Chen, 深山覚, 渡部晋治, 日本音響学会講演論文集 (2026年春季), 2026.
- Neural Audio Codecを用いた自己教師あり学習モデルにおける事前学習データの言語が下流タスクに与える影響の検証, 瀧澤大吾, 中村友彦, Samuele Cornell, William Chen, 深山覚, 渡部晋治, 日本音響学会講演論文集 (2026年春季), 2026.