

# 環境音の包絡パワー分布が音声了解度に及ぼす影響

真鍋 優奈<sup>†</sup> 玉川 勝也<sup>†</sup> 佐藤 敬子<sup>††</sup>  
<sup>†</sup> 香川大学大学院工学研究科 <sup>††</sup> 香川大学工学部

## 1. はじめに

音声了解度推定モデルに関する研究では、マスキングとして主に定常雑音を用いて検討されてきた。しかし、サウンドマスキングにおいて、定常雑音をマスキングとして用いると、不快に感じる可能性があるため、現実的には、川の流、小鳥の鳴き声等の環境音をマスキングとして用いるほうが望ましい。しかし、音声了解度推定モデルを用いて、環境音による音声のマスキングを実験的に検討した研究は少なく、既存のモデルを環境音に適用する上での問題点が明らかにされていない。これを明らかにすることができれば、環境音下の音声了解度推定モデルの開発に貢献することができる。そこで本研究では、既存の音声了解度推定モデルが環境音下の音声了解度をどの程度推定できるのかを検討した。

## 2. 実験

実験には、正常な聴力を有する学生 10 名(男性:10 名, 平均年齢:23 歳)が参加した。マスキングには、親密度別単語理解試験用音声データセット[1]の中から 750 語を選択した。マスキングには、流水音等の代表的な環境音 28 種と、定常雑音であるホワイトノイズとピンクノイズを合わせた 30 種を選択した。マスキングとマスキングのサンプリング周波数は 20000 Hz, 帯域幅は 20~10000 Hz, 継続時間は 1 秒間であった。また、マスキングの提示レベルは、等価騒音レベルで 50 dB, マスキングは、45, 50, 55, 60, 65 dB であった。実験は簡易防音室にて行い、防音室の暗騒音は等価騒音レベルで 24.7 dB であった。参加者には、1つの刺激音の提示が終了するごとに、聴取した音声をタイプ入力してもらった。

## 3. 結果と考察

実験に提示した単語と、参加者がタイプしたひらがなが完全に一致した場合を正解とし、各マスキングレベル、各マスキング 30 種に関して、参加者 10 名が回答した 5 単語の正答率の平均を単語了解度として算出した。実験の結果、どのマスキングも、提示レベルが上がるにつれて単語了解度が低くなった。また、同じ提示レベルであってもマスキングによって単語了解度に差が見られ、55 dB と 60 dB で顕著に見られた。

次に、既存の音声了解度モデル(mr-sEPSM モデル[2], STOI モデル[3], sEPSM<sup>corr</sup>モデル[4])の 3 種を用いて、音声了解度を推定し、実験結果と比較した。モデルの推定精度を比較するため、30 種のマスキングごとの

表1. 3 種の単語了解度モデルによる推定精度

mr-sEPSM		STOI		sEPSM <sup>corr</sup>	
$\rho$	MAE(%)	$\rho$	MAE(%)	$\rho$	MAE(%)
0.90	21.39	0.95	6.06	0.96	6.44

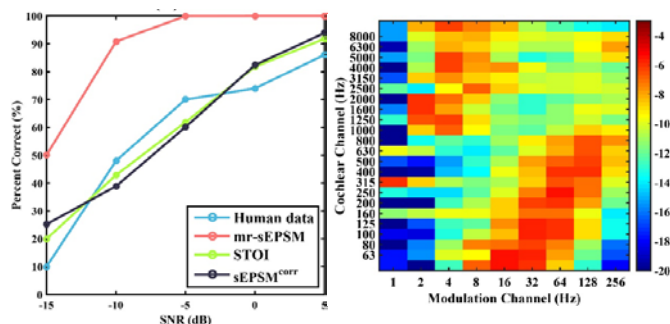


図1. 音声了解度推定結果(左)と包絡パワー分布(右)

モデルの推定値と実験結果の相関係数 ( $\rho$ )および平均絶対誤差 (MAE)を算出したものを表 1 に示す。また、図 1 左側に、環境音として、小鳥の群れをマスキングとして用いたときの音声了解度の推定結果を示す。

表 1 及び図 1 左より、STOI モデルと sEPSM<sup>corr</sup> モデルで精度良く単語了解度を推定することができた。両者のモデルは相関に基づくものであり、環境音下の音声了解度は包絡の相関関係が重要であることを示唆している。さらに、小鳥の群れは、別の環境音(小さい流れ)と長時間スペクトルが類似するにも関わらず、マスキング効果に差が見られた。これは、図 1 の右側に示すように、小鳥の群れが小さい流れに比べて、低い変調周波数帯域 (Modulation Channel < 16 Hz)に大きいパワーが存在することに起因すると考えられる。

## 参考文献

- [1] 天野成昭, 他, “親密度別単語理解試験用音声データセット (FW03),” NII 音声資源コンソーシアム, 2006.
- [2] S. Jørgensen *et al.*, “A multi-resolution envelope-power based model for speech intelligibility,” *J. Acoust. Soc. Am.*, vol. 134, no. 1, pp. 436–446, 2013.
- [3] C. H. Taal, *et al.*, “An algorithm for intelligibility prediction of time-frequency weighted noisy speech,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.
- [4] H. Relaño-Iborra *et al.*, “Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain,” *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2670–2679, 2016.