ISS-P-128

聴覚末梢系モデルの高雑音下における音声特徴量抽出精度の改善

安永 貴仁† 畔津 忠博†† 末竹 規哲††† 内野 英治†††,††††
† 山口大学理学部 †† 山口県立大学国際文化学部
††† 山口大学大学院創成科学研究科 †††† 一般財団法人ファジィシステム研究所

1. はじめに

現在の音声認識は統計モデルに基づいているが，高雑音下では認識精度が極端に低下する．また，ヒトは街中の様々な音の中から必要な音声を選び取り，話したい相手と会話ができる．本研究では，ヒトの聴覚末梢系モデルを構築し，新たな処理を加えることで，高雑音下における音声特徴量の抽出精度の改善を図る．

2. 聴覚末梢系モデルの構成

ヒトの聴覚末梢系において，音声特徴量の抽出は，主に内耳の基底膜，有毛細胞，聴神経が影響している．本研究では，基底膜モデル，内有毛細胞（順応特性・位相固定性消失）モデル，聴神経モデルから聴覚末梢系モデルを構築し，聴神経モデルに新たに簡単な処理を加える．基底膜モデルは，Gammatone filter[1]で構成されるフィルタバンクを使用し，内有毛細胞（順応特性）モデルは，Meddis inner hair cellモデル[2]を使用する．また，内有毛細胞（位相固定性消失）モデルと聴神経モデルは牧らのモデル[3]を使用する．

3. 特徴量抽出

聴覚末梢系モデルの出力である時系列パルス信号から，PSTH[4]を用いて特徴量抽出を行う．PSTHとは，時間軸上を一定幅の小区間に分割し，単位時間あたりのパルス数を各区間の値としてヒストグラムを作成したものである．

4. 提案モデル

牧らの聴神経モデルは，パルス生成の前段階として，内有毛細胞モデルより得られた離散信号から，膜電位を生成する．得られる膜電位の数は，基底膜モデルを構成するフィルタバンク，有毛細胞，聴神経の数と同数である．本提案モデルではそれぞれ331個とした．ここで，入力音に雑音が加わると，膜電位の極大値の数が増加する．本研究では，聴神経モデルから得られる膜電位$V_n(t)$の代わりに，以下の式から得られる$V_n'(t)$を新たな膜電位として置き換えたモデルを提案する．

$$V_n'(t) = V_n(t) \Big/ \sqrt[4]{V_n(t)\text{の極大値数}}. \qquad (1)$$

式（1）により膜電位を置き換えたとき，雑音の影響が大きい周波数帯域における聴神経の発火が抑制される．また，その周波数帯域で入力音の音圧が高ければ，無雑音時に見られる入力音の特徴は残る．

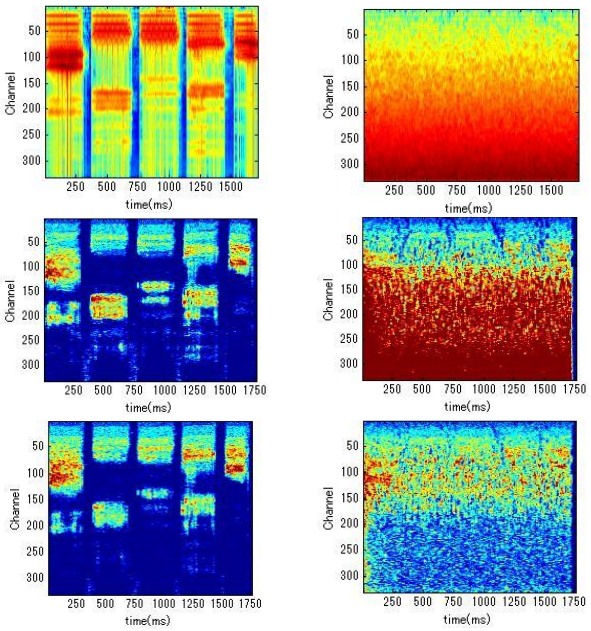

図1．無雑音時（左）と雑音下（右）における出力結果
（上からEP，PSTH（従来型），PSTH（式（1）を適用））

5. PSTHとExcitation Patternの比較

B.C.Mooreらによって考案された聴覚末梢系モデルの特徴量として，Excitation pattern(EP)[5]がある．高雑音下におけるEPと本モデルの出力（従来型，提案モデルの式（1）を適用)を図1に示す.ただし，入力音の音圧は45dBとし，雑音は55dBのホワイトノイズである．図1より，雑音の影響が大きいとき，EPよりPSTH（式（1）を適用）の方が無雑音時の特徴を読み取り易いといえる．さらに，式(1)により，雑音による発火を抑制できていることがわかる．正規化平均二乗誤差による比較では，高雑音下において，PSTH（式（1）を適用）が優れていた．

6. まとめ

本研究で新たに追加した処理は，雑音の影響が大きい周波数帯域における聴神経の発火を抑制する．今後は，さらにモデルの耐雑音性を向上させる．

参考文献

[1] R. D. Patterson and J. Holdsworth, Advances in Speech, Hearing and Language Processing, vol.3, pp.547-563, 1996.
[2] R. Meddis, J. Acoust. Soc. Am., vol.79, pp.702-711, 1986.
[3] 牧勝弘，赤木正人，廣田薫，日本音響学会誌65巻5号，pp.239-250, 2009.
[4] 銅谷賢治，伊藤浩之，藤井宏，塚田稔，脳の情報表現，朝倉書店，2002.
[5] B. C. J. Moore, B. R. Glasberg and T. Baer, J. Audio Eng. Soc., vol.45, no.4, pp.224-239, 1997.