

# 難聴の個人差に対応する補聴方式のための調音様式識別

森山 剛<sup>1,a)</sup> 高木 美奈<sup>1</sup> 高橋 伸弥<sup>2</sup> 小野博<sup>3</sup> 水町光徳<sup>4</sup> 坂田俊文<sup>2</sup> 白石君男<sup>5</sup>

概要：現在の補聴器は、ノイズ低減と高周波数帯域の増幅を行うものが主流である。これは、主に加齢による感音性難聴が、統計的に高周波域の可聴値が上昇して聴き取りが悪くなる傾向があるためだが、実は個人によっては摩擦音や破裂音、鼻音といった調音様式に応じて難聴の傾向が異なることが知られている。そのため、全てのユーザに対して特定の高域強調を行うだけの補聴器をつけても聴き取りが改善しないことが起こる。また、聴力を補うために音声認識によるテキスト書き起こしができるほど、話者が明瞭な発声を行うとも限らない。従って、発話内容が未知の音声を、感音性難聴の個性に合わせて強調することによって、聴き取りを改善する技術が求められる。そこで本研究では、異聴が無声子音及び通鼻音において多く出現するという既知の傾向に基づき、日本語音素を母音、破裂音、摩擦音、鼻音という共通する調音様式にクラス分けし（これを音韻クラスと呼ぶ）、音韻クラスごとの音声強調を実現するために、分析フレームで切り出された音声波形をいずれかの音韻クラスに識別するアルゴリズムを提案する。提案する音韻クラス識別アルゴリズムは、ベイズ理論に基づく最大事後確率推定法である。本報告では、音響特徴量としてゼロ交差数、エネルギー及びスペクトル平坦度を用い、各音響特徴量を単峰の正規分布でモデル化する単純な混合正規分布を尤度関数を用いた場合の実験結果を報告する。

## Classification of Articulation Manner for Realizing Personalized Hearing Aid

### 1. はじめに

難聴とは、聴力が障害された状態であり、外耳から中耳に至る経路に起因する伝音難聴と、内耳から聴神経に至る経路に起因する感音難聴に分類される [4]。これらのうち感音難聴では、主に加齢に伴って音に含まれる高周波成分の聞こえが低下する。図 1 では、周波数が 2,000Hz を超える辺りから高齢なほど聴力レベル (dB) が低下していくことがわかる [10]。その他、従来の難聴に関する研究では、特

定の語音が違う語音に聞き間違えられること（異聴）が報告されており、例えば、/t/や/h/が/k/に異聴されたりと

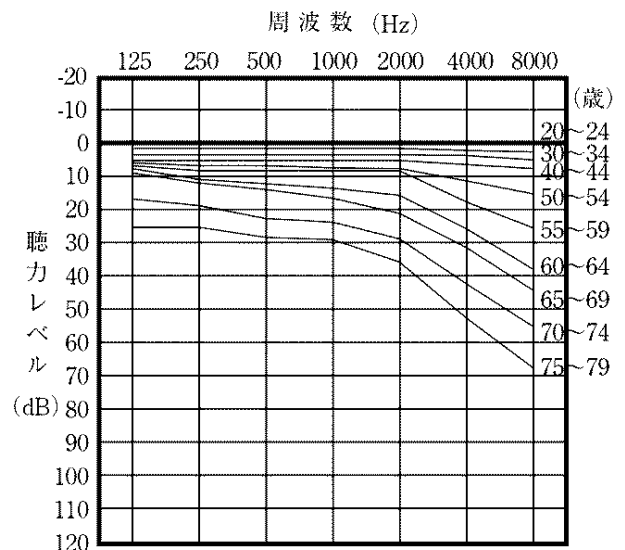


図 1 年齢別の平均聴力 [10]

Fig. 1 Average hearing ability across ages [10].

<sup>1</sup> 東京工芸大学  
1583 Iiyama, Atsugi, Kanagawa 243-0297, Japan  
<sup>2</sup> 福岡大学  
8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan  
<sup>3</sup> 西九州大学  
4490-9 Osaki, Kanzaki-machi Kanzaki-shi, Saga-ken 842-8585 Japan  
<sup>4</sup> 九州工業大学  
1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, Japan  
<sup>5</sup> 九州大学  
4-9-1 Shiobaru, Minami-ku, Fukuoka 815-0032, Japan  
<sup>†1</sup> 現在、メディアジョイント株式会社  
Presently with Media Joint Corporation  
<sup>a)</sup> moriyama@t-kougei.ac.jp

いった一定の傾向が指摘されている [6]。このような音韻の聴き取りにくさや異聴の傾向（難聴の個性）は、語音弁別測定によって知ることができる [3]。

一方、難聴者は聴き取りを改善するために補聴器を用いる。現在の補聴器は、上述のように、難聴によって聴き取りにくくなる高周波成分を増幅する機能や、空調の音など声以外の周辺のノイズを低減する機能を有する。しかし、このような一律な音強調処理では、上述の難聴の個性に対応することはできず、補聴器を使う人によってはその効果が発揮されない場合がある [5]。そこで、今後の補聴器では語音弁別測定の結果を反映して、難聴の個性に対応する補聴を実現することが求められている。

難聴の個性に合った補聴を実現するためには、まず、語音弁別測定により補聴器のユーザーが聴き取りにくい語音を特定する。そして、入力された音信号から強調すべき語音を検出し、あらかじめ定められたアルゴリズムによりその語音を強調して（ユーザーの耳へ向かって）出力するような補聴器の実現が有効であると考えられる。語音弁別測定については、従来の方式を用いれば良い。また、語音の強調に関する研究成果も既に報告されている [5]。しかし、補聴に有効な音韻を検出する手法に関しては未だ報告がないため、新たな検討が必要である。

補聴においては、一定の音響特徴を有する音韻（すなわち、その人物が聴き取りにくい音韻のグループ）を一律に強調する処理が有効であると考えられるため、従来の音韻検出技術ではなく、補聴に特化して音韻グループの検出を行う技術が必要となる。このような新たな音韻のグルーピングに関する検討を行うと共に、入力された任意テキストの音声信号からその音韻グループを検出する手法を提案する。

## 2. 音声強調のための音韻のグルーピング

### 2.1 音韻クラス

音声を構成する多くの音韻は、いくつかの調音様式で発音されている。母音は、声道を閉鎖せずに生成され、大きなエネルギーを有する。またその周波数特性は低域に集中する傾向がある。子音は、呼気が上気道を通過する際に抵抗にあい、それに打ち勝つ時に発生する音である。子音のうち、破裂音、摩擦音、鼻音に分類されるものは補聴において重要である [6]。破裂音は、口腔内の声道が閉鎖されたのちに突発的に開放される際の連続の音である。前半の無音部分に続いて声道が開放される際は雑音様の広帯域音が生ずる。摩擦音は、声道内に狭い隙間が作られ、空気がその隙間を通る際に発生する乱流の音である。摩擦音は、エネルギーは小さく、広帯域の雑音である。鼻音は、口腔が閉じられ、気流が鼻に抜ける有声音である [11]。したがって、鼻音は母音と音響的に類似している。

本研究では、同じ調音様式を有する音韻群を音韻クラス

表 1 音韻クラスと各音韻クラスに属する音韻

Table 1 Phoneme classes with the member of phonemes.

音韻クラス	属する音韻
母音クラス	/a/, /e/, /i/, /o/, /u/
破裂音クラス	/k/, /g/, /p/, /b/
摩擦音クラス	/s/, /z/, /f/, /sj/, /cj/, /hj/, /zj/
鼻音クラス	/m/, /n/

と呼び、同じ音韻クラスに属する音韻は、同じ方法で強調すれば聴き取りが改善すると仮定し、音声から個別の音韻を検出する代わりに、母音、破裂音、摩擦音、鼻音の各音韻クラスを検出することとする。各音韻クラスとそれらに属する音韻を表 1 に示す。

## 2.2 音響的特徴の確率モデル

### 2.2.1 音韻クラスを弁別する音響特徴量

2.1 節で述べた音韻クラスを音声中から検出するために、それらを弁別するような音響特徴量を算出する。母音クラスと摩擦音クラス及び破裂音クラスは、そのエネルギーに大きな差があり、また、周波数特性についても、母音クラスが低域に集中するのに対して、摩擦音クラスや破裂音クラスは広帯域に分布する（高周波域にも成分を持つ）。従って、信号のエネルギー及び周波数特性に関する音響特徴量により、音韻クラスを弁別できると考えられる。

### 2.2.2 音響特徴量の確率モデル

同じ音韻であっても、前後に接続する音韻に影響を受け、その音響的特徴はわずかに変化する（調音結合と呼ばれる）。そこで、未知の音声から算出された音響特徴量が、いくつかある音韻クラスの音響特徴分布のいずれと近いかを算出するために、音響特徴量の確率分布を数学的モデルで表現する必要がある。ここでは、音韻クラスの音響特徴量の確率分布が、平均とそこからのばらつきを有する正規分布に従うと仮定し、その確率分布のモデルとして混合正規分布（Gaussian Mixture Model; 以下、GMM）を用いる。GMM とは、データの分布を複数の正規分布の重ね合わせで表現する確率密度関数である。

## 3. 個人差に対応する補聴方式

### 3.1 概要

1 節で述べた補聴器のユーザーの個人差に対応するために、聴き取りにくい語音を特定した後、入力された音声波形が属する音韻クラスを識別し、あらかじめ定められたアルゴリズム（振幅や継続時間の変更）によって音声強調して出力する補聴方式を想定している。音韻クラスの検出には、最大事後確率（Maximum A Posteriori; 以下、MAP）推定法を用いる（3.3 節）。まず、入力された音声に対して、分析フレーム内で窓関数を乗じて切り出される短時間音声波形から音響特徴量を算出し、次に、算出した音響特徴量

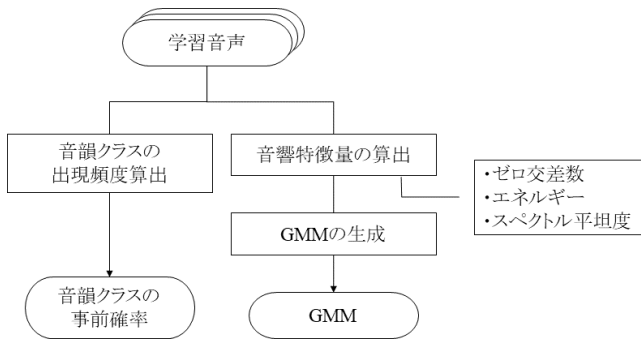


図 2 事前確率及び GMM 推定の流れ

Fig. 2 Estimation of prior probability and GMM.

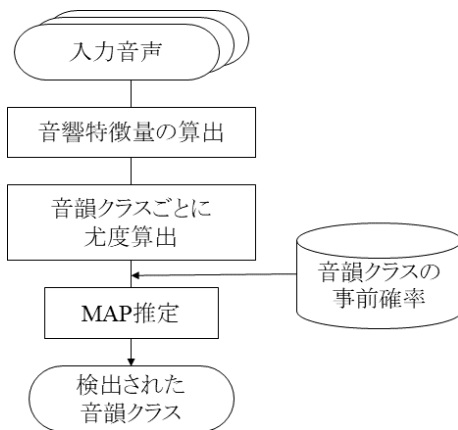


図 3 音韻クラス検出の流れ

Fig. 3 Detection of phoneme classes.

から推定される各音韻クラスの尤度と各音韻クラスの事前確率から事後確率を算出し、すべての音韻クラスに対する事後確率のうち最大の事後確率を与える音韻クラスを、当該音声波形が属する音韻クラスとして出力する (図 3)。これを分析フレームを時間軸方向にシフトしながら繰り返す。なお、音韻クラスの事前確率は、種々の音韻環境を含むような音声データベース (学習音声) から全音韻数に対する各音韻クラスの出現比率を求め、音韻クラスの尤度関数は、各音韻クラスの音響特徴量分布に GMM を当てはめる (図 2)。

### 3.2 音響特徴量

音響特徴量として、ゼロ交差数、エネルギー、スペクトル平坦度を算出する。ゼロ交差数とは、波形が一定時間内に時間軸をまたいで符号が反転する回数である (式 (1))。ここで、 $x_i$  は第  $i$  番目のサンプル値、 $\chi$  は分析フレームの開始時刻、 $\hat{\chi}$  は分析フレームの終了時刻、 $\theta$  は 0 との交差の有無を表している (交差ならば 1、そうでなければ 0)。エネルギーとは、ある時刻において周波数解析を行った区間のパワースペクトルの合計値である (式 (2))。ここで、 $F_s$  はサンプリング周波数、 $X(f)$  は周波数スペクトルを表す。スペクトル平坦度とは、パワースペクトルを周波数軸方向に任意のサブバンド (例えば 24 バンド) に分け、全サ

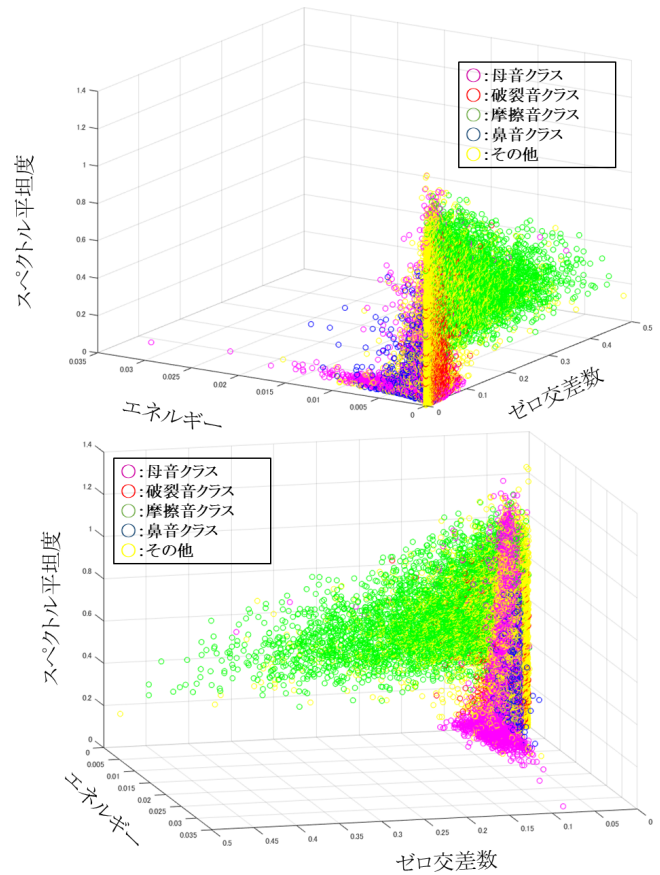


図 4 音響パラメータを算出した散布図

Fig. 4 Scatter plot of calculated acoustic parameters.

ブバンドのパワーの相乗平均を相加平均で除算した値である (式 (3))。ここで、 $\Lambda$  はサブバンドの数、 $\rho(\lambda)$  は  $\lambda$  番目のサブバンドにおけるパワーの合計値を表す [9]。

$$\text{ゼロ交差数} = \sum_{i=\chi}^{\hat{\chi}} \theta \begin{cases} 1 (x_i x_{i-1} < 0) \\ 0 (x_i x_{i-1} > 0) \end{cases} \quad (1)$$

$$\text{エネルギー} = \sum_{f=0}^{F_s/2} X^2(f) \quad (2)$$

$$\text{スペクトル平坦度} = \frac{\sqrt{\prod_{\lambda=0}^{\Lambda} \rho(\lambda)}}{\left(\frac{1}{\Lambda} \sum_{\lambda=0}^{\Lambda} \rho(\lambda)\right)} \quad (3)$$

後述する音声データベースに含まれる音韻波形すべてに関して、音響特徴量を算出し散布図にプロットしたものを図 4 に示す。図 4 から、母音クラスと鼻音クラスは、エネルギーが大きく、スペクトル平坦度が小さいことがわかる。また破裂音クラスは、エネルギーとゼロ交差数がともに小さく、スペクトル平坦度が幅広く分散する。また摩擦音は、ゼロ交差数が大きく、スペクトル平坦度が大きい。これらの図を比較すると、母音クラスと鼻音クラスは共にゼロ交差数が小さいが、鼻音クラスの方が分散が小さいことがわかる。

表 2 各音韻クラスの出現数と出現比率

Table 2 Frequency and the ratio of each phoneme class.

音韻クラス	出現数	出現比率
母音クラス	35,454	0.5256
破裂音クラス	6,000	0.0889
摩擦音クラス	5,266	0.0781
鼻音クラス	4,696	0.0696
その他	20,737	0.3074

### 3.3 ベイズ推定に基づく音韻検出

$X_i$  を  $i$  番目の音韻クラス,  $Y$  を未知の音声から算出された音響特徴量としたとき, その音声音が音韻クラス  $X_i$  である確率は  $P(X_i|Y)$  と表すことができ, これを事後確率と呼ぶ. 3.1 節で述べた MAP 推定法は, この事後確率  $P(X_i|Y)$  が最大となるような音韻クラス  $X_{\hat{i}}$  を推定する方法である (式 (4)) [8]. ベイズ理論により, 事後確率  $P(X_i|Y)$  は, 式 (4) の最右辺に示すように, 音韻クラス  $X_i$  の事前確率 ( $i$  番目の音韻クラスの出現比率)  $P(X_i)$  と,  $i$  番目の音韻クラスの音響特徴量の確率密度関数  $P(Y|X_i)$  の積の形に変形でき, これらの推定値より事後確率最大となる音韻クラス  $X_{\hat{i}}$  を求める (分母の  $P(Y)$  は音韻クラスによらず一定).

$$\hat{i} = \arg \max_i P(X_i|Y) = \arg \max_i \frac{P(Y|X_i)P(X_i)}{P(Y)} \quad (4)$$

## 4. 提案法による音韻検出実験

### 4.1 用いた音声データベース

種々の音韻環境を含み, かつ, 各音韻の始点及び終点の時刻を含むような音声データベースとして, 日本語 MULTEXT 韻律コーパス (MULTEXT-J) を用いた [7]. 同データベースには, 男性・女性各 3 人の計 6 人の話者による日常会話について, 音声波形と音素ラベル (音素名, 始点, 終点) が保存されている.

本データベースに含まれる音韻に関して, 音韻クラスごとに出現数並びに出現比率を求めた結果を表 2 に示す.

### 4.2 事前確率及び尤度関数推定の結果

各音韻クラスの前確率は, 4.1 節の音声データベースから求めた各音韻クラスの出現比率をそのまま用いることとした.

各音韻クラスの前確率は, MATLAB Statistics and Machine Learning Toolbox の fitgmdist 関数を使用し, 音響特徴量分布に混合正規分布を当てはめることで推定した [1]. このとき, 最大繰り返し数を 1,000 回, GMM の成分数を 1 とした. 事後確率は, 事前確率と尤度の積で算出されるが, ここでは pdf 関数を用いた.

なお, 音声データベースに含まれる全音韻数 72,153 個を 2 分割し, 前半を GMM の推定用に用い, 後半を評価用に用いた.

表 3 音韻検出結果の混合行列

Table 3 Confusion matrix of phoneme classification result.

		検出結果				
		母音	破裂音	摩擦音	鼻音	その他
正解	母音	9,301	870	96	5,001	2,460
	破裂音	317	568	15	387	1,713
	摩擦音	3	128	1,098	9	1,395
	鼻音	632	469	15	909	323
	その他	1,526	1,301	328	1,487	5,727

表 4 音韻クラス検出結果の適合率と再現率 [%]

Table 4 Accuracy evaluation of the proposed method.

	母音	破裂音	摩擦音	鼻音	その他
適合率	78.96	17.03	70.75	11.66	49.29
再現率	52.47	18.93	41.70	38.71	55.23

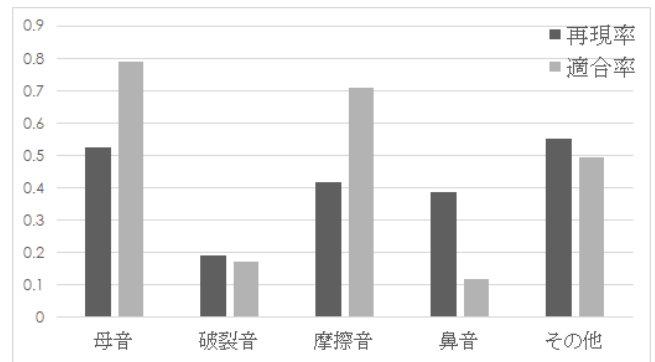


図 5 音韻検出結果の適合率と再現率

Fig. 5 Accuracy evaluation of the proposed method.

### 4.3 音韻識別の結果

4.2 節より, 音韻クラスごとに推定した GMM と事前確率を用いて, 評価用音声に含まれるすべての音韻波形に対し, 3.2 節の音響特徴量を算出し, 3.3 節で述べたように各音韻クラスの後確率を推定した. 表 3 に, 正解の音韻クラスを列方向, 事後確率が最大となった (識別された) 音韻クラスを行方向に集計したものを示す (対角要素が正解). 音韻クラス識別の精度を評価するために, 適合率と再現率を算出した結果を表 4 と図 5 に示す. 真陽性 (True-Positive; 以下, TP) と偽陽性 (False-Positive; 以下, FP), 偽陰性 (False-Negative; 以下, FN) を用いて, 適合率及び再現率は, それぞれ式 (5) のように算出される [2].

$$\text{適合率} = \frac{TP}{TP + FP} \quad \text{再現率} = \frac{TP}{TP + FN} \quad (5)$$

### 4.4 結果に関する考察

図 5 から, 母音クラス及び摩擦音クラスでは, 適合率が比較的高かったのに対して再現率は低くなっている. これは, 他の音韻クラスを誤って当該クラスとして検出したもの (FP) は少なかったが, 当該クラスであるにも関わらず他の音韻クラスとして検出されたもの (FN) が多かったこ

とを意味している。一方、鼻音クラスは適合率と再現率と共に低い結果となったが、その内訳は、母音クラスや摩擦音クラスとは対照的に、再現率の方が大きかった。これは、鼻音クラスではないもの（特に、母音クラス）を鼻音クラスと識別してしまったケースが多かったためである。

母音クラスのおよそ 28.2%が鼻音クラスとして誤って識別されてしまったこと、鼻音クラスのおよそ 27%が母音クラスとして誤って識別されてしまったことは、3.2 節の音響特徴量の散布図（図 4）において、母音クラスと鼻音クラスの音響特徴量分布が互いに多く重なりあっていることから予測され得る。したがって、これらを正しく識別するためには、母音クラスと鼻音クラスを弁別する新たな特徴量を考慮することが必要であると考えられる。

破裂音クラスについては、母音クラス及び鼻音クラスと識別されたものが大半であったが、これは、破裂音クラスの時間波形前半に存在する無音区間において、ゼロ交差数が少なく、母音クラスや鼻音クラスと類似した傾向を示すことが原因と考えられる。したがって、無音区間に続いてバースト状の波形が続くという破裂音クラスに特有の時間構造を反映するような音響特徴量を考慮することが有効であると考えられる。

摩擦音クラスについては、その他の音韻と識別されてしまったものが多かった。今回、明に母音、破裂音、摩擦音、鼻音の各クラスに分類しなかったものはすべて「その他」に分類してその音響特徴量分布を用いたが、1 節より、これら 4 クラスと類似する音響的特徴を有する音韻は GMM 推定処理から除外することを検討する。

一方、上述の改良を行うことは別に、今回用いたベイズ推定において、音韻クラスの出現比率を事前確率として MAP 推定に用いたが（3.3 節）、2 節より、鼻音クラスは、出現比率が 0.0696 と小さかったことが精度低下に繋がった可能性についても検討を行う。具体的には、事前確率をすべての音韻クラスで等しく仮定することで、ベイズ推定ではなく、最尤推定法によって音韻クラスを決定する方法を検討する。

## 5. おわりに

本研究では、ユーザーの難聴の個性に対応して特定の調音様式に従う語音を強調する補聴器の実現を目指し、入力音信号中から特定の音韻クラスを検出する手法を提案した。提案法は、信号のエネルギー及び周波数分布に関する音響特徴量に混合正規分布を当てはめ、最大事後確率推定法により、母音クラス、破裂音クラス、摩擦音クラス、鼻音クラス、その他を識別するアルゴリズムである。実験において、音響特徴量の確率モデルに最も単純な混合ガウス分布を用い、学習及び評価に用いた音声データも単純に音声データベースを半分に分けたものを用いた結果、ある程度の精度で識別に成功した。一方で、用いる音響特徴量や

学習に用いる音声データの選択に改良が必要なことも明らかになった。

本手法により、眼鏡をその人の視力に合わせて作るように、語音検査で聞き取りにくい語音を明らかにし、補聴器をユーザーの難聴特性に適応させることができるようになる。さらに補聴器に留まらず、スマートフォンやテレビ会議等の遠隔音声通話をユーザーにとって聴こえやすいものにするこへも応用が可能である。

今後は、最大事後確率推定法の諸元や用いる音声特徴量に関する検討を行う。また、ベイズ推定ではなく最尤推定法によって音韻クラスを決定する方法も検討する。さらに、単語理解度の観点からは、必ずしも個々の音韻の明瞭度は完全でなくても良いことを考慮し、補聴システム全体の性能から求められる音韻識別の要求精度を明らかにする。

謝辞 本研究は JSPS 科研費 JP18K12163 の助成を受けたものです。

## 参考文献

- [1] : MATLAB 入門, <https://jp.mathworks.com/help/matlab/getting-started-with-matlab.html>.
- [2] : scikit-learn でクラス分類結果を評価する, <https://pythondatascience.plavox.info/scikit-learn/>.
- [3] : 聴力検査・測定, <https://www.widexjp.co.jp/hearing/audiometry/>.
- [4] フリー百科事典『ウィキペディア (Wikipedia)』: 難聴.
- [5] 啓一 安, 隆行荒井, 敬 小林, 美津子進藤: 高齢者における聴覚特性の劣化が無声摩擦音・破擦音の識別に及ぼす影響, 日本音響学会誌, Vol. 68, No. 10, pp. 501–512 (オンライン), 入手先 (<https://ci.nii.ac.jp/naid/110009518043/>) (2012).
- [6] 小寺一興, 赤井貞康, 廣田栄子, 三浦雅美, 矢部 進: 感音性難聴における日本語音節の異聴の研究, 日本耳鼻咽喉科学会会報, Vol. 96, No. 9, pp. 1404–1409, 1573 (オンライン), 入手先 (<https://ci.nii.ac.jp/naid/130000996652/>) (1993).
- [7] 音声資源コンソーシアム: 日本語 MULTTEXT 韻律コーパス (MULTTEXT-J), <http://research.nii.ac.jp/src/MULTTEXT-J.html>.
- [8] 生田 顕, 折本寿子: 有限レベル変動域を考慮したファジィ・ベイズフィルタに基づく音環境の一状態推定法, 計測自動制御学会論文集, Vol. 52, No. 7, pp. 385–392 (2016).
- [9] 山田真司: 音楽情報処理で用いられる音響パラメータによる音楽理解の可能性, 日本音響学会誌, Vol. 70, No. 8, pp. 440–445 (2014).
- [10] 日本聴覚医学会編: 聴覚検査の実際 第 4 版, 南山堂 (2017).
- [11] 木村琢也, 小林篤志: IPA (国際音声記号) の基礎, 日本音響学会誌, Vol. 66, No. 4, pp. 178–183 (2010).