

話者認識技術を用いた性同一性症者の音声に対する 男声度・女声度の自動推定*

丸山和孝(東大) 櫻庭京子(さくら草学園) 峯松信明(東大) 広瀬啓吉(東大)
田山二郎(国立国際医療センター) 今泉敏(県立広島大) 山内俊雄(埼玉医科大)

1 はじめに

音声情報処理技術の発展に伴い、言語情報以外のパラ・非言語情報の抽出技術について研究が行なわれている。例えば [1] では、話者認識技術を応用することで、話者の知覚的年齢の自動推定を試みている。

2003 年度の法改正により、性同一性症 (Gender Identity Disorder, GID) 者は、自らにとって本来の性で生活ができるようになった。GID には、女性から男性に性別の移行を希望するもの (Male to Female, MtF) と男性から女性に性別の移行を希望するもの (FtM) がある。容姿に関しては外科手術を行うなどの対処が可能であるが、MtF 者の声には男性としての特徴が強く残る。FtM の場合、ホルモン投与により声を低くすれば男性であると判定されることが多いが、MtF に対しては効果が少ない [2]。その結果、MtF は書類上は女性でも、日常生活の中で女性として行動することが難しくなる場合が少なくない。本稿では、MtF の「声の女性化」を指導するボイスセラピーでの応用を目的とし、彼女等の音声に対する女声度の自動推定技術を検討した。

2 声の女性化と女声度の工学的定義

どのような音響的特徴によって、女性らしさは表現されるのだろうか？一般に MtF は女性の高い声を望む傾向にあり、中には外科手術 (声帯手術) を施す方々もいる。しかし、聴取実験によればその効果は十分ではない [2]。第二著者の櫻庭が行なっている女声化のアプローチは、自然性が損なわれない程度で高い声の生成を習慣化させ、その上で“喉を絞る”こと (即ち、声道形状の変化) を要求する。声道長の短長によりフォルマント周波数は上下するが、この変化を実現することが狙いである。「声を高くする」訓練と比較し「喉を絞る」訓練は、そもそも「正しい絞り方が出来ているのかどうか」に対する MtF 自身の判断が困難であり、試行錯誤となる場合が多い。そこで音声を入力とし、適切な判定を自動的に行なう装置が望まれている。話者認識技術は基本的に話者の声道形状の音響的モデル化であるため、上記の臨床応用は技術的に十分可能であると考えられる。

混合ガウス分布 (GMM) を用いた話者モデリングは、ケプストラム系列へと変換された音声データ全体を一つの GMM でモデル化する。音声全体を対象とすることで音韻によるスペクトル変動はキャンセルされ、音声全体に対して静的に影響を及ぼす要因 (即ち、話者性やマイクの特性など) のみがモデル化されることになる。話者 s のモデルを M_s とした場合、話者認識 (観測量 o が予め用意された話者集団のどれかを判定する) は $P(o|M_s)$ を算出することによって、話者照合 (o が話者 s の音声であるか否かを判定する) は、 $\frac{P(o|M_s)}{P(o|M_{\bar{s}})}$ を算出することで可能となる。本稿では基本的に話者照合の枠組みでの女声度の推定を考え、以下の式でこれを定義する。

$$F(o) = \log P(o|M_F) - \log P(o|M_M) \quad (1)$$

M_M 、 M_F は各々男声モデル、女声モデルを表す。

3 実験

3.1 使用した音声

モデル作成には JNAS 音声、男女それぞれ 114 名ずつ、各話者 30 発声を用いた。評価用の音声は GID 話者の音声 143 発声 (MtF111 名、FtM2 名の計 113 話者) である。GID 話者の音声には別途聴取実験を行い、25 名ほどの聴者に男性か女性の 2 値で回答してもらった。以下、この聴取実験において女性と判定された割合を女性判定率とする。本実験では、計算機により算出した女声度と女性判定率との相関係数を求め、性能の評価を行った。

3.2 音響パラメータ抽出と無音区間・子音区間除去

分析条件を表 1 に示す。音声はパラメータ系列化したのちに、学習用、評価用音声の双方に対し無音区間除去を行った。話者性が現れるのは主に母音の部分であるため、子音区間も同時に除去した。以上の操作はパワーに着目した処理で行った。また、 F_0 の抽出できなかったフレームも除去した。

Table 1 分析条件

サンプリング	16bit / 16kHz
窓	窓長 25msec、シフト長 10msec
パラメータ	MFCC12 次元+ Δ 12 次元+ Δ E 計 25 次元 log F_0

* Automatic Femininity Estimation of MtF's Speech Based on Speaker Recognition Techniques., by Kazutaka Maruyama, Kyoko Sakuraba, Nobuaki Minematsu, Keikichi Hirose, Toshio Yamauchi, Niro Tayama, Satoshi Imaizumi

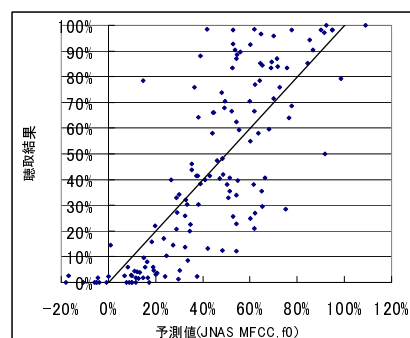
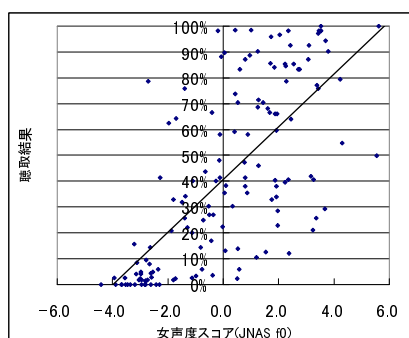
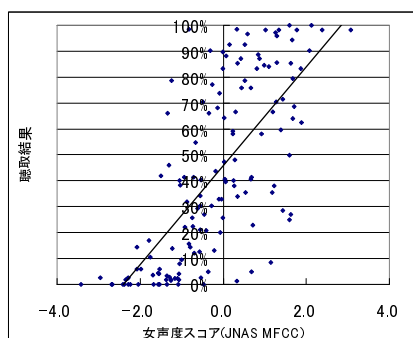


Fig. 1 女声度 (MFCC) 対 聴取結果 Fig. 2 女声度 ($\log F_0$) 対 聴取結果 Fig. 3 予測値 (JNAS) 対 聴取結果

3.3 モデル作成とスコア算出

続いて MFCC と $\log F_0$ それぞれに、女声モデルと男声モデルを 16 混合 GMM として作成した。 $\log F_0$ についてもこのようなモデル化をすることで、通常の女性よりも極端に高い音声ではスコアが低くなる。

GID 話者の音声の女声度と女性判定率の散布図は、MFCC、 $\log F_0$ についてそれぞれ図 1、2 のようになった。また、相関係数はそれぞれ 0.717、0.704 となった。より女性らしい声を出すためには声の高さだけではなく、声道形状が重要であることが示された。 $\log F_0$ のモデルに比べ MFCC では女声度と女性判定率が乖離する音声が多く、より安定した評価が可能である。

3.4 予測値の算出

次に、線形回帰分析により入力音声の聴取実験で女性と判定される率の予測値を算出した。ただし、説明変数としては MFCC、 $\log F_0$ それぞれの男性モデルおよび女性モデルの対数尤度、4 つすべてを用いた。このときの散布図を図 3 に示す。相関係数は 0.799 となり、声道形状の情報と音源情報をあわせることで高い相関が得られた。この場合でも依然として外れサンプルが存在する。図左上の音声は聴取実験では女性、計算機では男性と判定されたものであるが、これらの音声は高齢の女性に聞こえるものが多く、今回用いたデータベースでは正しく評価できなかったと考えられる。図右下はその逆であるが、声が不自然に高く、計算機がそれを女性らしいと高く評価してしまった音声が多く集まった。

3.5 GID 話者の音声によるモデル

前節までは JNAS のデータからモデルを作成したが、GID 話者の音声を用いて、女性判定率の低い音声から男性モデル、女性判定率の高い音声から女性モデルを作成することも可能である。これは、MtF の話者にとっては、ボイスセラピーによって女性判定率の高くなった声こそ目指すべき声であり、逆に男性と判定される音声からは遠い方が望ましいからである。そこで JNAS 音声を用いた場合と同様に GID 話者の音声から男性モデルと女性モデルを作

成した。男性モデルの学習データには女性判定率が 60% 以下の音声、女性モデルの学習データには女性判定率が 60% 以上のものを用いた。ただし評価用音声と同一話者の音声は学習用データから除外した。また音響パラメータは MFCC のみを試した。このときの女声度と女性判定率の相関係数は 0.749 となった。JNAS モデルの場合に比べ高い相関が得られた。つぎに、第 3.4 節の予測値計算における、4 つの対数尤度に GID 話者モデルの対数尤度 2 つを加えた 6 つの対数尤度を説明変数として重線形回帰分析を行い、予測値を算出した。この予測値と女性判定率の相関係数は 0.807 となった。GID 話者モデルを合わせることで、わずかながらより高い相関が得られた。

4 まとめ

話者認識技術に基づく、MtF を対象とした音声の女声度の評価手法を提案した。人間による男女の判断には、声の高さだけではなく、声道形状の制御も重要であることが示された。また臨床応用のため、線形回帰分析を用いて聴取実験による女性判定率を予測した。予測値と実際の聴取結果との相関係数は 0.807 と良好な結果を得ることができた。なお、実際の臨床応用に関する検討は論文 [4] を参照されたい。

参考文献

- [1] 峯松信明他, “話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定”, 情報処理学会論文誌, vol.43, no.7, pp.2186-2196, 2002
- [2] 櫻庭京子他, “女性と判定された性同一性障害者 (MtF) の声の基本周波数”, sp2002-187, 信学技法 vol.102 No.749, 2003
- [3] J.P.Campbell, “Speaker Recognition: A Tutorial,” Proc. IEEE, vol.85, No.9, pp.1437-1462, 1997
- [4] 櫻庭京子他, “Transsexual voice therapy における話者認識技術の臨床応用”, 音講論 (春), 491-492, 3-P-20, 2006