社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

スーパーベクトルとSVRに基づくMtF話者のための女声度推定

王　程碩† 鈴木　雅之† 峯松　信明† 櫻庭　京子†† 広瀬　啓吉†

† 東京大学，〒113–8656 東京都文京区本郷 7–3–1
†† 獨協医科大学越谷病院，〒343–8555 埼玉県越谷市南越谷 2-1-50
E-mail: †outeiseki@gavo.t.u-tokyo.ac.jp

あらまし　MtF 話者の音声を対象に，女声度推定の高精度化を試みた。GMM スーパーベクトルあるいは HMM スーパーベクトルを話者特徴量として，識別的な回帰モデルである SVR を用いて推定した。性別依存 GMM 尤度スコアと線形回帰を用いた従来手法よりも高い精度が得られた。
キーワード　GID，MtF，女性度，聴取実験，GMM，HMM，スーパーベクトル，SVR

Voice femininity estimation for MtF patients using supervectors and SVR

C. WANG†, M. SUZUKI†, N. MINEMATSU†, K. SAKURABA††, and K. HIROSE†

† The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113–0033 Japan
†† Dokkyo Medical University Hospital, 2–1–50, Minami-koshigaya, Koshigaya, Saitama, 343–8555 Japan
E-mail: †outeiseki@gavo.t.u-tokyo.ac.jp

Abstract　Femininity estimation of MtF (Male to Female) voices is technically implemented. Speaker characteristics are extracted as GMM supervectors or HMM supervectors and femininity estimation is realized as discriminative regression using support vector regression. Experiments show that the proposed method shows higher performance than our previous method of using likelihood scores of gender dependent GMMs and linear regression.
Key words　GID, MtF, femininity, listening test, GMM, HMM, supervector, SVR

1. はじめに

GID（Gender Identity Disorder，性同一性障害）者，特に，生物学的には男性であるが，社会的には女性としての生活を望むMtF（Male to Female）の方々にとって，声の女らしさ（女声らしさ）の獲得は困難であると言われている [1]。この場合，外科治療やホルモン投与に頼るよりも，ボイスセラピーがより効果的である [1]。MtF 者は，自分の声がどの程度女声として知覚されるのかを気にするが，これを客観的に調査するには，ある程度の規模の聴取者を対象にした聴取実験が必要である。これは多大の労力を要するので，セラピストの意見に頼ることになる。しかし当該患者の声を聞き続けると，セラピストも馴化してしまい，純粋に客観的な判断が難しくなる。このような背景から，従来より筆者らは判定のぶれがない，「女声度聴取実験」シミュレータとしての女声度推定器の開発を行なってきた。先行研究では，男声モデル，女声モデルを GMM を用いて構築し，入力音声に対してこれらのモデルから計算される尤度スコアと線形回帰の枠組みを用いて女声度を予測していた。本研究では GMM スーパーベクトル（GS），あるいは HMM スーパーベクトル（HS）を話者特徴量として，識別的な回帰モデルである SVR（Support Vector Regression）を使って女声度推定の高精度化を試みたので報告する。

2. MtF 音声に対する女声度ラベリング

「女声度聴取実験」シミュレータを作成する場合，様々な MtF 音声に対して女声度スコアが付与されたコーパスが必要である。本研究では先行研究において構築されたコーパスを用いるが [2]，これは MtF 話者 111 名から収録した 142 発声（セラピー前後の音声が別途収録されている話者もいる）に対する女声度ラベリング結果（−3〜+3 の 7 段階）である。音声は全て「ジャックと豆の木」の冒頭の 2 文（39 単語，135 音素）である。18 名の評価者の中，各評定者と，それ以外の評定者の平均値で定義される平均女声度との相関の平均は 0.87 であった。

3. 従来手法

筆者らの一部は [2] において，JNAS の全男性話者，全女性話者を用い，MFCC（s）及び F_0（f）を特徴量として，男声モデル（$P(s|M_M^s)$，$P(f|M_M^f)$），女声モデル（$P(s|M_F^s)$，$P(f|M_F^f)$）を GMM を用いて構築し，入力音声に対する声の女性らしさ（Voice Femininity, VF）を下記で定義した。

$$VF(s,f) = \alpha \log P(s|M_F^s) + \beta \log P(f|M_F^f) + \gamma \log P(s|M_M^s) + \epsilon \log P(f|M_M^f) + C \quad (1)$$

各係数や定数項は，女声度ラベリング結果を用いて最小二乗誤差基準で推定した。女声モデルに対する係数（α，β）は正の値を，男声モデルに対する係数（γ, ϵ）は負の値を持つ。

4. 提案手法

先行研究における男声・女声モデリングは GMM を用いた話者認識技術の一応用として位置づけられる。近年の話者認識研

表 1　実 験 条 件

サンプリング	16bit/16kHz
フレーム長，周期	25ms, 10ms
プリエンファシス	$1-0.97z^{-1}$
特徴パラメータ	12MFCC + 12ΔMFCC
	+Δ パワー，及び，対数 F_0
GMM	対角分散共分散行列，混合数 128
文 HMM	left-to-right 型，状態数 135，単一ガウス分布

究ではスーパーベクトルを話者特徴量とする手法が主流となっている [3]。多数話者の音声から学習した GMM，即ち UBM（Universal Background Model）を事前に構築し，これに対して各話者の音声サンプルを用いて UBM を話者適応し（平均ベクトルのみの適応），得られた UBM-based 話者依存 GMM から平均ベクトルを連結したスーパーベクトル（GS）を話者特徴量とする。本研究でもまず，これを話者特徴量として利用する。次に，使用する音声がテキスト固定であることを考慮し（2. 節参照），GS では無視される時系列情報（テキスト情報）も利用したスーパーベクトルの構成を検討する。具体的には，不特定話者 HMM を UBM とし，話者適用により得られた特定話者 HMM の平均ベクトルからスーパーベクトルを得，これを使う。

女性度推定の枠組み（推定器）としては，先行研究では MFCC や F_0 に関する GMM 尤度スコアに対する線形回帰を行ったが，本研究では，上記の話者特徴量に対して識別的な回帰である SVR を直接応用する。ここで F_0 の利用は，有声無声区間が発声によって異なるため，GS 利用時のみ考慮した。

5. 女声度推定の実験

5.1 音声資料と実験条件

GMM-UBM 及び HMM-UBM の構築は JNAS を用いた。なお，以下の実験では音声中の無音区間は，自動検出により排除して実験を行なっている。表 1 に音響分析条件を示す。

5.2 実 験 手 順

以下の手順に従って GS-SVR，HS-SVR による女声度推定実験を行なった。なお，女声度の予測は 142 個の MtF 発声に対して leave-one-out cross-validation を行い，推定された 142 個の予測値と聴取実験結果とを比較する。

5.2.1 GS-SVR に基づく推定

- JNAS 中の全発話を用いて GMM-UBM を構築する（混合数 128）。MFCC，F_0 に対して別個に構築する。
- 各 MtF 話者の音声（39 単語）を用いて GMM-UBM を話者適応（MAP 適応）し，各 MtF 話者の GMM を，MFCC, F_0 に対して得る。
- 128 個のガウス分布からスーパーベクトルを得る。MFCC が 25 次元，F0 が 1 次元であり，各々の GMM の混合数は 128 であるので，スーパーベクトルの次元数は 3,328 となる。
- このスーペーベクトルを話者特徴量として，SVR の枠組みで女声度を予測する。

5.2.2 HS-SVR に基づく推定

- JNAS 全発話から MFCC を抽出し，不特定話者音素 HMM を構築（1 状態/音素，全 135 状態）し，それを連結して文 HMM-UBM を得る。
- 各 MtF 話者の音声を用いて HMM-UBM を話者適応（MAP 適応）し，各 MtF 話者の文 HMM を得る。

表 2　女声度推定結果 (MFCC)

	相関	二乗誤差
GMM-LR	0.74	1.17
GS-SVR	0.84	1.04
HS-SVR	0.81	1.13

表 3　女声度推定結果 (MFCC と F_0)

	相関	二乗誤差
GMM-LR	0.85	0.70
GS-SVR	0.88	0.64

- 135 状態の文 HMM から，各状態の平均ベクトルを連結し，スーパーベクトルを得る。MFCC が 25 次元，各状態は単一ガウス分布であり，スーパーベクトル次元数は 3,375 となる。
- このスーペーベクトルを話者特徴量として，SVR の枠組みで女声度を予測する。

以上の手順により得られた結果を，同一条件で行なわれた従来手法の結果と比較する。

6. 実験結果と考察

MFCC だけを使う場合，従来手法（GMM-LR）と提案手法（GS-SVR，HS-SVR）によって予測された女声度と，聴取実験により得られた女声度との相関係数及び平均二乗誤差を表 2 に示す。MFCC と F0 両方を用いる場合の GMM-LR と GS-SVR の推定結果を表 3 に示す。提案手法は，特徴量が MFCC 及び MFCC と F_0 両方の場合も，相関，二乗誤差の両者において先行研究よりも高い精度で女声度を推定できることが分かる。今回のタスクはテキスト固定であり，系列モデリングである HMM に基づいたスーパーベクトルの方がより適切に「女らしさ」を表現できていると期待されたが，実験の結果は GMM スーパーベクトルの方が精度が高い結果となった。なお F_0 も利用した GS-SVR の結果（相関係数）は 0.88 であり，0.87（2. 節参照）を越えている。女声度推定器は「もう一人の評定者」として十分に認定できると言える。

7. ま と め

MtF 者に対するボイスセラピーの技術的支援を目的として，女声度推定器の高精度化を検討した。先行研究では GMM による男声・女声モデルを構築し，尤度スコアの線形回帰で女声度を推定していた。本研究ではこれを，特徴量として GMM スーパーベクトルや HMM スーパーベクトルを利用し，回帰の枠組みを SVR に変更することで，聴取実験との相関値及び平均二乗誤差において，より高精度な結果を得ることができた。

文　献

[1] 櫻庭京子他，“性同一性障害者（MtF）の音声に対する知覚的性別の自動推定”，信学技報 SP2005-189, 29–34, 2006.

[2] N. Minematsu *et al.*, “Development of a femininity estimator for voice therapy of gender identity disorder clients,” in Speaker Classification II, LNAI4441, 22–33, Springer, 2007.

[3] T. Kinnunena *et al.,* “An overview of text-independent speaker recognition: from features to supervectors,” Speech Communication, 52, 12–40, 2010.

[4] W.Toshiya *et al.,* “Investigations of features and estimators for speech-based age estimation,” in Proc. of the Second APSIPA Annual Summit and Conference, 470–473, 2010.