Vol.2011-SLP-89 No.17
2011/12/19

感情音声の感じ方の違いを考慮した感情推定

仁平 佳宏[†1]　橘 完太[†2]

近年，音声からの感情推定というテーマの研究が盛んである．従来研究では，話し手の発話した音声から話し手の感情を推定することを主としたのに対し，筆者らは聞き手が音声に含まれる感情をどう受け取るかの推定をおこなう．感情音声の感じ方は聞き手に依って大きく異なり，各聞き手に合った推定システムが望まれる．そこで筆者らは，実際に聞き手が感情音声をどう感じているかを調べ，それをもとに聞き手の感じ方の違いをいくつかの個性に普遍化し，個性に基づく推定システムの創造を目指す．本論文ではその第一歩として，聞き手が感情音声を聞いてどう感じるかの推定問題を定式化する．

Estimation of Individual Listener's Guess at Speaker's Emotion

Yoshihiro Nihei[†1] and Kanta Tachibana[†2]

Nowaday, estimation of emotion from speech is a hot research domain. Most conventional researches focus on speaker's emotion. In contrast, our interest is listener's guess at speaker's emotion. Guess at speaker's emotion is defferent among listeners. So, an emotional speech estimation system considering differences among listeners is needed. The aims of our research is to make aprobability model of listener's personality, to establish an identifiation method of personality model parameters, and to create an estimation system based on personality. As a first step of these aims, in this paper we formulate the estimation of individual listener's guess at speaker's emotion problem.

†1 工学院大学大学院情報学専攻
Graduate School of Infomatics, Kogakuin University
†2 工学院大学情報学部
Faculty of Infomatics, Kogakuin University

1. はじめに

近年，音声から感情を推定するテーマの研究が数多く行われている[1)2)]．音声には話し手の感情情報が含まれていること[3)4)]から，音声は感情推定において重要な要因となる．特に，音声の特徴量として韻律情報を用いた感情推定は盛んであり，数々の実績がある[5)6)]．

従来研究では，話し手の発話した音声から話し手の感情を推定することを主とした．これら既存の研究では，主に演技による音声データが用いられている．ここで，演技による音声は，本当に演技者がある特定の感情になり発話しているのかという疑問が浮かぶ．私たちは上手な演技を見ると，上手い演技だと感じるのである．演技者がどのような気持ちで演技しているかも関係はあるだろう．しかし，演技者がいくら気持ちを込めて演技しても，聞き手にその気持ちを感じてもらえなくては意味が無い．つまり，演技とは聞き手ありきなのではないだろうか．

そこで筆者らは，演技において重要なことは，演技者が本当に特定の感情になっていることではなく，聞き手がどう感じるかである，と位置づけた．つまり，話し手に演技をしてもらう実験においては，聞き手がどう感じるかを推定した方が直接的であると考えた．そこで本研究では，話し手が発した感情音声を聞き手が聞いたとき，聞き手が話し手の感情をどう感じるか，の推定をおこなうこととした．これを以下では，聞き手が感情音声を聞いてどう感じるかの推定，と表現する．

感情音声を聞いてどう感じるかは，聞き手に依る所が大きい．つまり，感情音声の感じ方は聞き手によって様々ということである．例えば，話し手が怒りながら，もしくは怒りの演技で発話した音声を，ある聞き手は怒りと認識し，ある聞き手は喜びと認識するという具合である．実際に，後述の感情音声主観評価実験においても，そのような傾向があることがわかる．そこで筆者らは，聞き手による感情音声の感じ方の違いに着目した．文献[7)]では聞き手間の類似度を測り，類似度の近い人の回答を重視した推定法を提案し，結果として既存の手法と同程度の推定結果を得た．文献[7)]では聞き手間の類似度の測定法に改良の余地があると述べた．改良の一案として本論文では聞き手間の感情音声の感じ方の違いをどう扱うかを検討し，定式化していく．

2. ドイツ語感情音声データベースと感情音声主観評価実験

2.1 ドイツ語感情音声データベースと特徴抽出

本研究では，“ 平静 ”，“ 怒り ”，“ 喜び ”，“ 悲しみ ”の 4 つの感情を扱う．各感情に属す

る感情音声は 25 個ずつ，合計 100 個をドイツ語感情音声データベース[8] から選出した．ドイツ語感情音声データベースから感情音声を選出する手順を述べる．文献[8] では各感情で同一文章を読み上げている．感情音声の評価段階で，認識率テストで 80%未満，かつ，自然さテストで 60%未満の感情音声はデータベースから除外される．本研究では，4 つの感情音声が 1 つも除外されていない文章を優先的に選出した．その後，1 つの感情音声が除外されている文章から 3 つの感情音声を選出した．以下同様の手順で各感情の感情音声が 25 個ずつになるまで続けた．ドイツ語感情音声データベースには，男声と女声の感情音声が収録されているが，本研究では男声のみを用いた．男声と女声では音声の特徴に大幅な違いがあるためである．

また，選出した感情音声からそれぞれ表 1 に示す 5 つの特徴量を抽出する．また，基本周波数のイメージを図 1 に示す．音圧についても同様のイメージである．抽出した特徴量を $\mathbf{x} \in \mathbb{R}^5$ と表現する．

2.2 感情音声主観評価実験

本研究の目的は，聞き手が感情音声を聞いてどう感じるかを推定することである．そのため，感情音声主観評価実験としてドイツ語感情音声データベースから抽出した 100 個の音声を実際に被験者に聞いてもらい，それぞれの音声はどの感情として発話されていると思うかを回答してもらった．感情音声主観評価実験では，1 つの音声に対して必ず 1 つの感情を指定してもらった．なお，感情音声主観評価実験時，100 個の感情音声はランダムな順番で再生した．

感情音声主観評価実験で集まった感情音声 n に対する感情 j の票数を w_{nj} とする．例えば，感情音声 3 は平静 1 票，怒り 16 票，喜び 4 票，悲しみ 0 票であったので，w_{31}=1, $w_{32} = 11$，$w_{33} = 3$，$w_{34} = 0$ である．$n \in \{1, ..., 100\}$ は感情音声番号で，$j \in \{1, 2, 3, 4\}$ である．j の 1，2，3，4 はそれぞれ平静，怒り，喜び，悲しみを表す．

3. 感情推定問題の定式化

3.1 ベイズの定理

3 章では感情推定問題の定式化をしていく．定式化をおこなうにあたり，まずベイズの定理について触れておく．ベイズの定理は以下の式で表される．

$$p(H \mid D) \propto p(D \mid H)p(H) \tag{3.1}$$

ここで，H は仮説，D はデータを表す．つまり，データが得られたもとでの仮説の確からしさは，その逆確率と仮説の確率との積に比例するということである．式 (3.1) の左辺を事後確率，または事後分布，$p(D \mid H)$，$p(H)$ をそれぞれ尤度関数，事前確率，または事前分布という．さらに，H を H_2 に従う確率変数とみなせば，式 (3.1) は次のように表せる．

$$p(H \mid D, H_2) \propto p(D \mid H)p(H \mid H_2) \tag{3.2}$$

H_2 は仮説 H の仮説である．一般的に H_2 はハイパーパラメータと呼ばれる．

表 1 感情音声から抽出する特徴量
Table 1 Feature value extracted by emotional speech

特徴量	表記方法
基本周波数に関する特徴量 [Hz]	
最大値	$f0$max
平均値	$f0$mean
ダイナミックレンジ	$f0$range
音圧に関する特徴量 [dB]	
最大値	$p$max
平均値	$p$mean

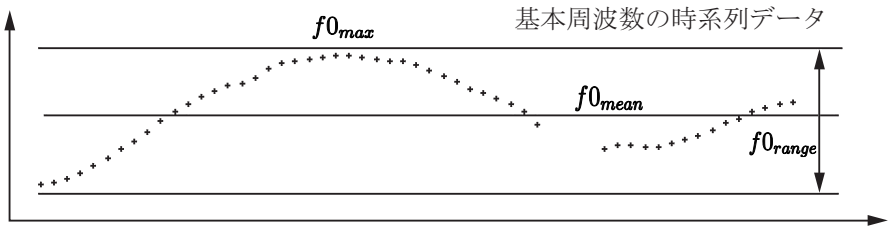

図 1 基本周波数に関する特徴量のイメージ
Fig. 1 Image of fundamental frequency

3.2 感情推定問題の定式化

聞き手が感情音声を聞いてどう感じるかの推定の定式化にあたり，まず考えることは，実際に聞き手が感情音声をどう感じているかである．聞き手が感情音声をどう感じているかは，感情音声主観評価実験での得票数 w_{nj} をもとに得られる．w_{nj} と感情音声の特徴量 $\mathbf{x}_n$ から次式により，感情 j と感じる特徴量の平均ベクトル $\boldsymbol{\mu}_j \in \mathbb{R}^5$ を求める．

$$\boldsymbol{\mu}_j = \frac{1}{\sum_{n=1}^{N} w_{nj}} \sum_{n=1}^{N} w_{nj}\mathbf{x}_n \tag{3.3}$$

さらに，$\boldsymbol{\mu}_j$ を用い，感情 j と感じる特徴量の精度行列 $\mathbf{T}_j$ を得る．

$$\mathbf{T}_j^{-1} = \frac{1}{\sum_{n=1}^{N} w_{nj}} \sum_{n=1}^{N} w_{nj}(\mathbf{x}_n - \boldsymbol{\mu}_j)(\mathbf{x}_n - \boldsymbol{\mu}_j)^{\mathrm{t}} \tag{3.4}$$

ここで，精度行列は共分散行列の逆行列と定義される．また，t は転置を表す．

次に，聞き手の感じ方の違いを考慮したモデルを考える．まず，聞き手 i の個性を表す α_{ik} を導入する．α_{ik} は，聞き手 $i \in \{1, ..., 14\}$ が個性グループ $k \in \{1, ..., 4\}$ に属する確率である．α_{ik} は，聞き手 i の N 個の音声に対する回答 $\delta_{nij} \in \{0, 1\}$ に応じて決めて固定する．ここで δ_{nij} は感情音声 n を聞き手 i が聞いて感情 j と感じたら 1 を，それ以外の場合には 0 をとる．α_{ik}，δ_{nij}，さらに感情音声特徴量 $\mathbf{x}_n$ より，個性 k を持つ人が感情 j に聞こえる特徴量の平均ベクトル $\boldsymbol{\theta}_{jk} \in \mathbb{R}^5$ を次式で求める．

$$\boldsymbol{\theta}_{jk} = \frac{1}{\sum_{i=1}^{I} \alpha_{ik} \sum_{n=1}^{N} \delta_{nij}} \sum_{i=1}^{I} \alpha_{ik} \sum_{n=1}^{N} \delta_{nij} \mathbf{x}_n \tag{3.5}$$

次に $\boldsymbol{\theta}_{jk}$ を用い，個性 k を持つ人が感情 j に聞こえる特徴量の精度行列 $\boldsymbol{\Lambda}_{jk} \in \mathbb{R}^{5\times5}$ を得る．

$$\boldsymbol{\Lambda}_{jk}^{-1} = \frac{1}{\sum_{i=1}^{I} \alpha_{ik} \sum_{n=1}^{N} \delta_{nij}} \sum_{i=1}^{I} \alpha_{ik} \sum_{n=1}^{N} \delta_{nij} (\mathbf{x}_n - \boldsymbol{\theta}_{jk})(\mathbf{x}_n - \boldsymbol{\theta}_{jk})^{\mathrm{t}} \tag{3.6}$$

上述の変数から，感情音声を聞いてどう感じるかの推定をおこなう．本論文では，次のように推定式を定義した．また，ν_j はガウス・ウィシャート分布の自由度パラメータ，$\mathbf{W}_j$ は同じくガウス・ウィシャート分布の尺度行列である．$\mathbf{W}_j$ の初期値は $\mathbf{W}_j = \mathbf{T}_j$ とする．$\boldsymbol{\mu}_j$，ν_j，$\mathbf{W}_j$ はそれぞれガウス・ウィシャート分布のパラメータとなる．

$$\begin{aligned} &p(\boldsymbol{\theta}_{jk}, \boldsymbol{\Lambda}_{jk} \mid \mathbf{x}_n, \delta_{nij}, \boldsymbol{\mu}_j, \mathbf{T}_j, \nu_j, \mathbf{W}_j) \\ &= p(\mathbf{x}_n, \delta_{nij} \mid \boldsymbol{\theta}_{jk}, \boldsymbol{\Lambda}_{jk}, \boldsymbol{\mu}_j, \mathbf{T}_j, \nu_j, \mathbf{W}_j) p(\boldsymbol{\theta}_{jk}, \boldsymbol{\Lambda}_{jk} \mid \boldsymbol{\mu}_j, \mathbf{T}_j, \nu_j, \mathbf{W}_j) \end{aligned} \tag{3.7}$$

本論文では，式 (3.7) の尤度関数を多変量ガウス分布で表す．多変量ガウス分布は

$$\mathcal{N}(\mathbf{y} \mid \mathbf{m}, \mathbf{L}^{-1}) = \frac{1}{(2\pi)^{D/2}} |\mathbf{L}|^{1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{L}(\mathbf{y} - \mathbf{m})^{\mathrm{t}}\right\} \tag{3.8}$$

と書ける．ただし，$\mathbf{m}$ は D 次元の平均ベクトル，$\mathbf{L}$ は $D \times D$ の精度行列，そして $|\mathbf{L}|$ は $\mathbf{L}$ の行列式を表す．

また，尤度関数が多変量ガウス分布で，なおかつ平均ベクトル，精度行列が未知の場合，共役事前分布がガウス・ウィシャート分布になることが知られている．そこで本論文でも，式 (3.7) の事前分布をガウス・ウィシャート分布で表す．ウィシャート分布は

$$\mathcal{W}ish(\mathbf{L} \mid \mathbf{V}, v) = B|\mathbf{L}|^{(v-D-1)} \exp\left\{-\frac{1}{2}\mathrm{Tr}(\mathbf{V}^{-1}\mathbf{L})\right\} \tag{3.9}$$

と書ける．v は分布の自由度パラメータ，$\mathbf{V}$ は $D \times D$ の尺度行列，そして Tr(·) はトレースを表す．正規化係数 B は次式で与えられる．

$$B(\mathbf{V}, v) = |\mathbf{V}|^{-v/2} \left\{2^{vD/2} \pi^{D(D-1)/4} \prod_{d=1}^{D} \Gamma\left(\frac{v+1-d}{2}\right)\right\}^{-1} \tag{3.10}$$

これらから，式 (3.7) の右辺は次式のように表すことができる．

$$\mathcal{N}(\mathbf{x}_n, \delta_{nij} \mid \boldsymbol{\theta}_{jk}, \boldsymbol{\Lambda}_{jk}) \mathcal{N}(\boldsymbol{\theta}_{jk} \mid \boldsymbol{\mu}_j, \mathbf{T}_j) \mathcal{W}ish(\boldsymbol{\Lambda}_{jk} \mid \nu_j, \mathbf{W}_j)) \tag{3.11}$$

式 (3.11) から $\boldsymbol{\theta}_{jk}$ と $\boldsymbol{\Lambda}_{jk}$ の確率分布が求められる．本論文では，個性を考慮した感情推定確率モデルのパラメータ同定問題を定式化した．今後は，今回定式化した $p(\boldsymbol{\theta}_{jk}, \boldsymbol{\Lambda}_{jk} \mid \cdot)$ を用いて，感情音声を聞いてどう感じるかの推定実験をおこなう．

参 考 文 献

1) Burrows, D., Ghadiyaram, A., Jordan, M., Nwana, A. and Xu, A.: Emotion Recognition and Synthesis in Speech (2010).
2) Neiberg, D., Elenius, K. and Laskowski, K.: Emotion Recognition in Spontaneous Speech Using GMMs, *INTERSPEECH2006-ICSLP* (2006).
3) 門谷信愛希，阿曽弘具，鈴木基之，牧野正三：音声に含まれる感情の判別に関する検討，電子情報通信学会研究報告，Vol.100, No.522, pp.43–48 (2000).
4) 重永　實：感情の判別分析からみた感情音声の特性，電子情報通信学会論文誌，pp. 726–735 (2000).
5) 多田和彦，矢野良和，道木慎二，大熊　繁：感情遷移における急激な韻律特徴変化の検出による感情遷移判別法，日本知能情報ファジィ学会誌，Vol.22, No.1, pp.90–101 (2010).
6) 矢野良和，鈴木克典，野田哲矢，道木慎二，大熊　繁：個人音声情報を反映させた感情音声認識機の構築，日本知能情報ファジィ学会誌第 17 回インテリジェント・システム・シンポジウム講演論文集，pp.393–398 (2007).
7) 仁平佳宏，橘　完太：感情音声の感じ方の類似度を考慮した感情推定，第 13 回日本感性工学会大会 (2011).
8) F.Burkhardt, Paeschke, A., Rolfes, M., W.Sendlmeier and Weiss, B.: A Database of German Emotional Speech, *Proc. Interspeech2005* (2005).