感情が聴取による話者認識に及ぼす影響 —聴取結果の考察ならびに韻律的特徴との比較—*

☆川本啓輝，荒井隆行（上智大・理工），網野加苗（科警研）
安啓一（上智大・理工）

1　はじめに

音声から話者を識別する話者認識に有効な特徴量に関する研究が現在まで数多く行われてきた。例えば網野（2006）は，調音の際に生理学的特徴が関与する鼻音の周波数特性が個人性を特定するためには有効であると述べている[1]。しかし話者を認識する際には，個人性以外の要因にも注目すべきである。例えば，話し手の心理状態・感情といった情緒性は個人性の知覚に影響を及ぼしていると考えられる。犯罪捜査において音声から犯人を識別するなど，法科学への応用を実現する場合は犯人の情緒性（その状態での感情，詐称の意図など）を音声から把握し，考慮しなければならない[2]。すなわち，話者を正確に識別するためには，話者の音声に含まれている感情の影響を考える必要がある。Shahin（2011）は話者認識システムに感情認識のフェーズを組み込むことで，話者認識率を向上させるシステムの構築を実現した[3]。しかし，感情がどのように聴取による話者認識率に影響を及ぼすかは未だに解明されていない。そこで本論文では，感情が聴取による話者認識に及ぼす影響を調べ，その影響がどのような要因により生じているのかを調査するため，実験結果と音声の韻律的特徴との比較を行った。

2　実験

2.1 録音

日本語母語話者である演劇経験の無い男性14 名に感情を含んだ音声を発話してもらった。今回用いた感情は，平静（nor），喜び（joy），悲しみ（sad），怒り（ang）の 4 種類とした。それぞれの感情には場面設定を行い，使用文はいずれの感情で発話されても不自然にならない文を選定した[4]。場面設定，文について以下に表 1, 2 を示す。発話者 1 名当たり 96 文（文 8 種類，感情 4 種類，3 回ずつ発話）を録音し，合計で 1344 文（発話者 14 名，1 名当たり 96 文発話）の資料が得られた。

表 1　感情表現のための場面想定

感情	想定される場面
平静	できる限り棒読みで
喜び	仲の良い友達との会話
悲しみ	明日 8 時に会社に 行くなんて憂鬱だ…
怒り	虫の居所の悪い上司のつもりで

表 2　発話文
*は聴取実験に用いた文を表す。

No	文
1*	いつ来るかなあ
2*	そこ閉めてくれる？
3*	雨が降ったね
4	大丈夫だよ
5	よく変わるね
6	待っています
7	明日空いてる？
8*	書いておきます

2.2 聴取実験

適切に感情移入ができている話者の文を用いるため，著者を含む 4 名で感情判定を行い，音声が 4 つの感情のうち，どの感情で発話されているかを判断し，選択してもらった。その結果，意図した感情が含まれていると判断された音声のみを使用した。実験には，聴取者の負担を考慮し，表 2 のうち No. 1, 2, 3, 8 の文を用いた。最終的に聴取実験で用いた音声は，80 文（発話者 5 名（per1 ~ 5），文 4 種類，感情 4 種類）となった。

聴取実験は全 16 名（男性 8 名，女性 8 名）に対して ABX 法により行った。刺激数は 240

*Effects of emotions on human speaker identification: Comparison of identification results and prosodic properties of the stimuli, by KAWAMOTO, Hiroki, ARAI, Takayuki (Sophia Univ.), AMINO Kanae (NRIPS), YASU Keiichi (Sophia Univ.)

刺激（話者組み合わせ 10 通り，文組み合わせ 6 通り，感情 4 種類）とした。以下に詳細な条件を示す。

・A, B は同じ文，X は異なる文（聴取者は X の話者が A，B のいずれと同じかを回答）。

・3 回の発話のうち 1 回目の発話を使用（感情判定の平均一致率が最も高かったため）。

・発話の際の口とマイクロホンとの距離の違い，発話時の音量の違いによるばらつきの影響を排除するため，提示音声の音圧レベルは 50 dB に統一。

・順序効果に対するカウンターバランスを取るため，ABX と BAX を同数呈示。

2.3 実験結果

感情，話者ごとの話者認識正答率の結果をそれぞれ図 1, 2 に示す。図 1 より nor の正答率に対して，joy の正答率が下降，sad, ang の正答率が上昇する結果となった。感情間の有意差を調べるため，Tukey の多重比較検定を行ったところ正答率が高い sad, ang と正答率の低い nor, joy のグループ間で有意差（$p < 0.01$）が得られた。図 2 より per5 の音声のみ正答率が高く，他の 4 名の話者は差が無いことが分かった。また，聴取者に対して実験後に行ったアンケートにも「特徴的な声の話者が 1 名いた」という意見が多くあった。その後，感情の分析と同様に，話者について一要因の分散分析を行った結果，話者間に有意差が認められた（$p < 0.01$）。また，per5 と各話者の間には有意差が認められたが，他の話者との間には有意差がなかった。感情により正答率に差が生じた原因は，話者による感情表現の違いにあると考え，感情・話者の 2 要因に注目して分析を行った。結果を図 3 に示す。図 3 より，話者・感情によって正答率に差が生じていることが分かった。特に，sad, ang はどの話者に対しても nor より正答率が高い傾向にあり，joy は低い傾向となった。これは図 1 と同様の結果であった。

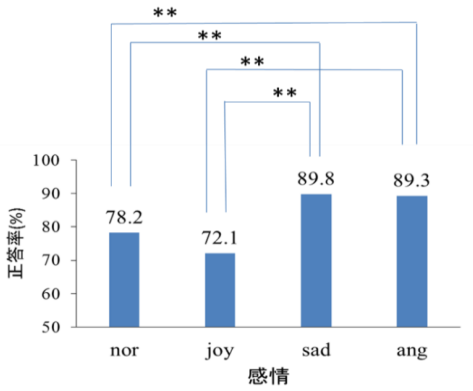

図 1　聴取実験の正答率（感情ごと）
**は有意水準 1%での有意差を示す。

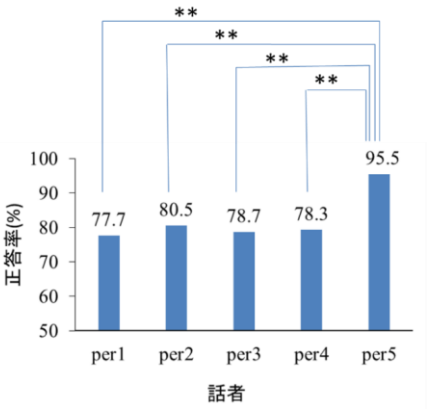

図 2　聴取実験の正答率（話者ごと）
**は有意水準 1%での有意差を示す。

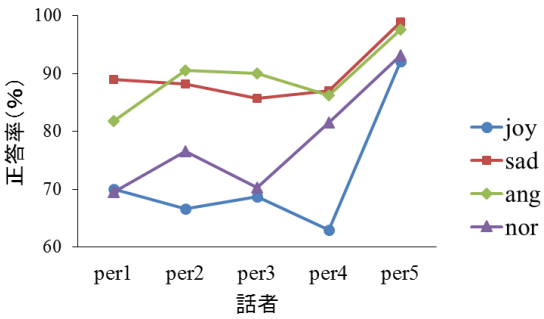

図 3　感情×話者の正答率

3　音響分析

各感情，話者で音声の音響的な特徴に傾向が見られるかを調べた。注目したのは基本周波数 F0 に関係する特徴量 $F0_{avg}$（F0 の発話内平均），$F0_{range}$（発話内の F0 の最大値-発話内の F0 の最小値），1 モーラあたりの平均持続時間である。注目する特徴量は文献[4]を参考に選定し，ソフトウェア Praat[5]を使用して分析した。それぞれの特徴量について，話者ごとの感情表現の違いを調べた。その結果を図 4-6 に示す。

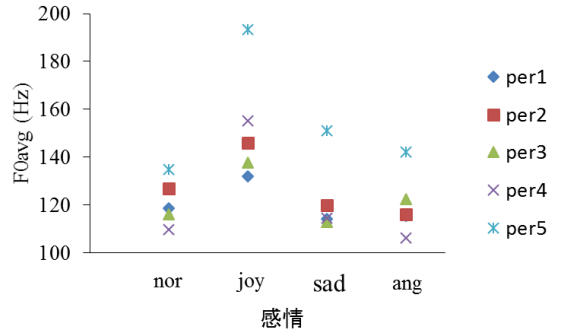

図 4　感情，話者ごとの $F0_{avg}$ の平均

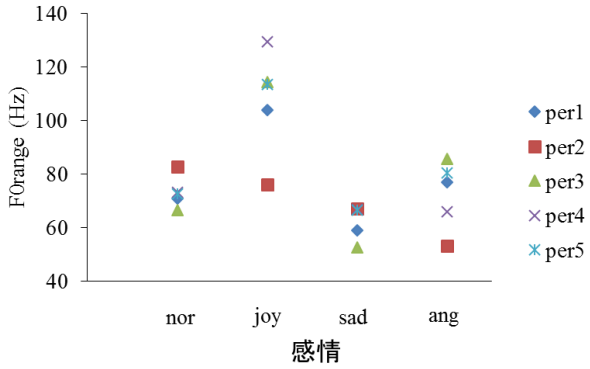

図 5　感情，話者ごとの $F0_{range}$ の平均

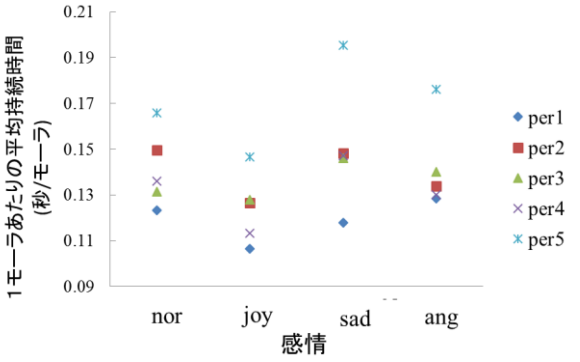

図 6　感情，話者ごとの 1 モーラあたりの平均持続時間の平均

4　考察

4.1 考察（実験）

1 要因の分散分析から，joy の音声を用いた際の話者認識正答率が nor に比べて低い結果となった。これは，joy の感情を移入することで話者間の特徴の差が小さくなったためと考えられる。また，sad と ang の正答率が上昇したのは sad, ang の感情を移入することで話者間の特徴の差が顕著になったためと考えられる。2 要因の分散分析の結果（図 3）から nor と joy，sad と ang の正答率に差が生じることが分かった。また，per5 の正答率はすべての感情において高いことが分かる。これらは 1 要因の分散分析と同様の結果となっている。per1～per3 は joy と nor，sad と ang の間に正答率の差が生じている。しかし per4 に関しては，joy の正答率が低く，nor の正答率が高くなっている。したがって，per4 は他の話者に比べて joy と nor の感情表現の方法に差があることが分かる。

4.2 考察（音響分析）

図 4 より joy の音声は，他の感情に比べて $F0_{avg}$ が大きいことが分かる。また，図 5 より話者によって $F0_{range}$ の差が大きく開いていることが分かる。$F0_{avg}$ は声の高さ，$F0_{range}$ は抑揚に相当する[6]ので，joy の音声は他の感情に比べて，声が高くなり抑揚が大きくなることが予想される。図 6 より，sad は他の感情に比べて 1 モーラあたりの平均持続時間が話者によってばらついている。以上のことから joy の音声では F0 が変化し，sad の音声では 1 モーラあたりの平均持続時間が変化することが分かった。

図 4 より，per5 の $F0_{avg}$ は全ての感情において，他の話者に比べて大きくなっている。さらに図 6 より per5 の 1 モーラあたりの平均持続時間は他の話者に比べて長くなっている。per5 の話者認識正答率が高かったことや聴取者に対するアンケート結果を考慮すると，$F0_{avg}$ と 1 モーラあたりの平均持続時間は，感情が含まれた音声の話者認識に有効な特徴であることが示唆される。今回の聴取実験では，ABX 法を用いたので，話者 A と話者 B の $F0_{avg}$，1 モーラあたりの平均持続時間によって話者 X を識別しているということになる。すなわち，話者 A, B の $F0_{avg}$ 差，1 モーラあたりの平均持続時間差が正答率に影響を及ぼしていると考えられる。そこで，聴取実験に用いた 240 通りの音声資料に対して，話者 A, B の $F0_{avg}$ 差，1 モーラあたりの平均持続時間差と正答率の相関関係を調べた。結果を図 7, 8 に示す。ピアソンの積率相関係数を求めた結果，$F0_{avg}$ 差と正答率の間には正の相関が認められた（$r = 0.39$, $p < 0.01$）。また，1 モーラあたりの平均持続時間差と正答率の間にも正の相関が認められた（$r = 0.36$, $p < 0.01$）。有意差が認められたものの高い相関係数が得られなかった理由は正答率が全般的に高く，分布に天井効果が見られたためだと考えられる。

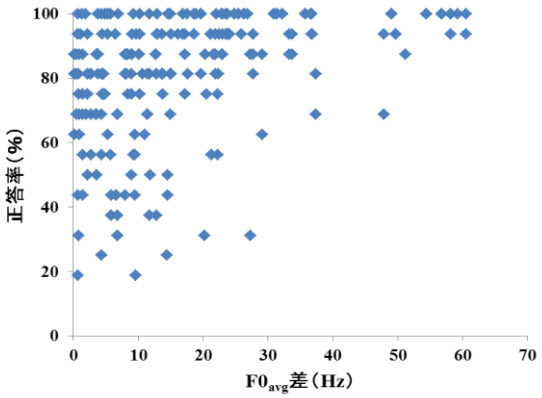

図 7　$F0_{avg}$ 差と正答率の相関関係

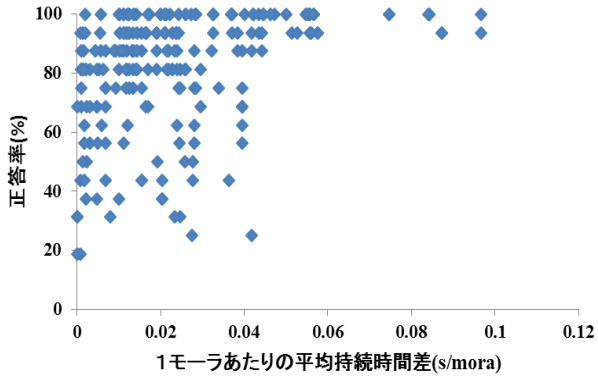

図 8　1 モーラあたりの平均持続時間差と正答率の相関関係

今回の ABX 法による実験では同話者を判断する状況と，別話者を判断する状況が存在する。例えば，A = X（話者 A, X が同話者）であるとき，聴取者は A = X を判断すると同時に B≠X を判断することになる。すなわち，聴取者は同話者判断と別話者判断を同時に行うことになる。それぞれの判断をする際に有効な韻律的特徴を調査するため，韻律的特徴と正答率の相関を求めた。相関係数を求めたところ次のような結果になった。

・$F0_{avg}$ について

同話者判断：A = X（r = 0.006），B = X（r = 0.081）

別話者判断：A≠X（r = 0.429），B≠X（r = 0.246）

・1 モーラあたりの平均持続時間差について

同話者判断：A = X（r = 0.146），B = X（r = 0.187）

別話者判断：A≠X（r = 0.241），B≠X（r =-0.006）

以上より，これらの韻律的特徴は同話者判断に比べて別話者判断に有効であることが示唆された。

5　おわりに

本実験では，音声の録音に協力してもらった話者の中には，演劇経験のある話者はいなかった。つまり，感情をこめる演技には慣れていなかったということになる。2.1 節の感情判定の際に，適切に感情表現ができている話者のみに音声を厳選したものの，感情表現の能力に個人差があることは音声から明らかであった。そのため，感情表現によって話者認識率に影響が及ぼすことを示すためには，発話者を演劇経験のある人物に統一することや，演技ではない自発的な感情音声を使うことなどが必要であると考えられる。

また，話者を識別する際に人は，様々な特徴量を複合的に用いることが分かっている[7]。本研究で注目した特徴量は F0 と持続時間だけであったので，感情表現と話者認識の関係を示すためには不十分である。今後，さらなる特徴量に注目することで話者認識の研究を深めていくことが課題となる。

謝辞

本研究を行うにあたり，実験参加者として協力して下さった方に感謝致します。本研究の一部は JSPS 科研費（26870865）の助成を得た。

参考文献

[1] Amino *et al.*, Acoust. Sci. Tech., 27(4), 233-235, 2006.

[2] J. P. Campbell, IEEE Signal Processing Magazine, 26, 95-103, 2009.

[3] I. Shahin, Int. J. Speech. Technol., 14, 89-98, 2011.

[4] 門谷他，音声言語情報処理, 34(8), 43-48, 2000.

[5] P. Boersma and D. Weenink, “Praat: doing phonetics by computer,” Version 5.3.51, Retrieved from “http://www.praat.org/”.

[6] 森山他，電子情報通信学会論文誌, 82(4), 703-711, 1999.

[7] M. R. Sambur, IEEE Transactions on Acoustic, Speech and Signal Processing, 23(2), 176-182, 1975.