名前での呼びかけ言葉の感情をあらわす音響特徴量の分析
Analyses of acoustic features that shows emotion of hail word in name

伊藤　博昭
Hiroaki Ito
法政大学情報科学部ディジタルメディア学科
E-mail:hiroaki.ito.2a@cis.hosei.ac.

Abstract

A purpose of this research is "Emotional speech evaluation system make for the training support of the young actor and the young voice actor. This system has a function to evaluate how cloth to emotion which a speaker intends by recognizing emotional speech with a computer. First, I collected.and, 600 samples of emotional speech of the hail word were collected from movies, dramas and animations .And, the sample was classified into 11 patterns "Pleasure", "Anger", "Sorrow", "Calmness", "Impatience (anxiety)", "Surprise" , "Doubt", "Trouble", "shock", "Request", and "Others". "Pleasure" and" Anger" and" Calmness" and "Impatience (anxiety)" were appeared a lot in that was done to analyze feature of pitch by "straight". The result, the pitch of "Calmness" was most smaller, and "Impatience" most bigger. It was bigger that the dynamic range of "Anger" was 3.8877.It was smaller that the dynamic range of "Calmness" was 2.2883.This result by using probability model was considered, the distinction rate of "Pleasure" was70% ,And " Anger" was 10%," Calmness" was 80%,"Impatience (anxiety)" was 30%.

1. まえがき

現在，声優，俳優の育成学校では，ヴォイストレーニングを通して発声方法の練習している．さらに，エチュード，ダンス等の授業で，さまざまな表現力を身につけ「役になりきる」ための技術も習得している．声優では，アフレコ演習を行い，声優としてのテクニックを身につけている．

演者は，感情を表現するため，表情，セリフ，動作などを用いる．その中で「セリフ」は，演じる感情を表現することが非常に難しい．それは，自分の演じている感情が，見ている人間に的確に表現出来ているのかを知る方法は限られているからである．先生（プロ俳優，声優）または，周りの人間に評価してもらう．つまり，人間の耳で聞き，それを人間が評価する手段しかないのが現状で,客観的に評価する方法が無い．

そこで，本研究の最終目的は，「若手俳優，声優の育成支援のための感情音声評価システム作成」である．そのシステムとは，発話者が感情音声((怒り，喜びなど)をコンピュータに認識させることで，その感情が，発話者の意図した感情とどれほどの整合性があるかを評価するものである．それにより，個人での演技練習も可能となり，具体的な結果を示すことにより，より正確な感情表現が可能になると考えられる．

音声情報には，韻律，音韻，声質の 3 つの要素が存在する．その中で音の時間的リズム，強弱，高低，あるいは長短等を表わす韻律情報は，人の感情表現を最も特徴付けている[7].

本研究は，感情音声の韻律的特徴パラメータを調べていく．それにより感情ごとに差別化を図り，その情報を確率モデルで分類することにより，発話者の感情を評価する．さらに，本研究では，テレビ番組，映画中に登場した「呼びかけ言葉」を感情音声サンプルとして使用する．経験を積み，あらゆる練習を行っている．プロの声優，俳優のセリフは，感情ごとにはっきりとした違いが見られると考えられる．そして「呼びかけ言葉」は，視聴者のストーリー理解を深めるために多く登場する．プロの声優，俳優は，その言葉により演じている感情を表現することが多いためこの言葉を研究の対象として選んだ．

2. 呼びかけ言葉

2.1. 名前での呼びかけ言葉呼びかけ言葉

任意の相手に話かけたい時に文頭に使う言葉である．呼びかけ言葉には以下の二つのパターンが存在する．1)「ねぇ，なぁ，おい，やあ」などの呼びかけ 2)「～君．それを取って．」という～君という名前での呼びかけ

前者の呼びかけは，出現頻度が低く，サンプル収集に適していない．そこで本研究では，後者の「名前での呼びかけ言葉」を使用する．

2.2. 感情音声評価システム(実行の流れ)

まず，発話者が感情音声(怒り， あせり，平静，喜び), どの感情での発話をトレーニングしたいかを決定する．そして，マイクを使い(本研究では，ダイナミックマイクを使用)，発話者がその音声をコンピュータに入力する．そして，入力された感情音声の pitch と pitch の dynamic range を抽出する．その値を確率モデルにより，一番近しい感情に分類する．最後に，決定した感情を出力する．

2.3. Pitch による特徴量抽出

上記で記したように，音の時間的リズム，強弱，高低，あるいは長短等を表す，韻律情報は，人間の感情を最も特徴付けている[7]．それにより，過去の研究では，韻律的特徴情報を用いて，感情音声を感情ごとに分類する研究が多く行われている．韻律的特徴情報に，声の高低を表す pitch 情報が存在する．Pitch 情報では，喜びの平均 pitch が高い．悲しみは，平均 pitch が低いなど，感情ごとにはっきりとした特徴が抽出できることが分かってい

Supervisor: Prof. Katunobu Ito

る[2][3]．さらに、感情認識では，70%程の精度が求められている．そこで，本研究では最大 pitch，最小 pitch，平均 pitch，pitch の dynamic range を使用し，感情ごとの特徴を求める．dynamic range は，pitch の最大周波数を最小周波数で割ることにより計算した．

3. 感情音声の収録

過去の研究の音声サンプル収集法では，実環境で，「演劇部員の方，一般の方に収集したい言葉を指定し，発話してもらう方法」，「自然な会話を録音し，その中でサンプルを集める方法」が多くみられた．

しかし，本研究は，「若手俳優，声優の育成支援」を目的としているため，テレビ番組，映画での音声サンプル収集を行った．

さらに，過去研究では収集する音声サンプル数は 50～200 個程度のものを使用していた．しかし，本研究では，より信頼度の高いものを得るため，サンプルを約 600 個収集する．

3.1 テレビ番組，映画での収録

本研究は，アニメ番組，ドラマのテレビ番組と映画を使い，その言葉を調査研究する．サンプルを収集する中で，どのジャンルでの収集が一番効率良いのかを調べるため，アニメ，ドラマ，映画のジャンルは考えず，ランダムに作品を選んでいく．

テレビ番組，映画の中で「名前での呼びかけ言葉」が登場した場合にその場面の映像と音声を切り離し，音声のみを保存する． 保存する音声の形式は，モノラルの 48khz 16bit 音声とする．

4. 感情音声の分類

感情音声のサンプルを集め，感情音声一つ一つに，感情の特徴を付加する．感情の種類は，「喜び，怒り，哀しみ，平静心，あせり(不安)，驚き」など一般的なものを使用する．新たな感情は，登場するたびに追加する．

感情音声は，複数の人間により，分類を行うことにより，出来るだけ客観的に分類する．

4.1. 感情の分類

「ケイゾク(ミステリードラマ 2 話分で 45 分＊2，計 90 分」 (収集サンプル数 22 個)

「となりのトトロ(ファミリー映画 90 分作品)」 (48 個)

「金田一少年の事件簿~魔術列車編(ミステリーアニメ 4 話分で 23 分＊4，計 92 分)」 (44 個)

「新世紀エヴァンゲリオン（SF アニメ 4 話分で 23 分＊4，計 92 分） (33 個)

「天空の城ラピュタ((ファミリー映画 124 分作品)」(147 個)

「魔女の宅急便(ファンタジー映画 103 分作品)」(81 個)

「千と千尋の神隠し(ファミリー映画 124 分作品)」(149 個)

「耳をすませば(ファミリー映画 110 分作品)」(103 個)

以上の計 625 個

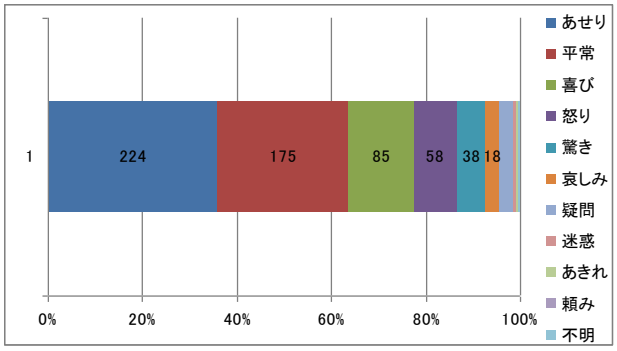

図 2　感情の分類結果(比率が多い順)

これらの作品で登場した感情音声は，「喜び」，「怒り」，「哀しみ」，「平静」，「あせり(不安)」，「驚き」，「疑問」，「迷惑」，「あきれ」，「頼み」，「その他」の，計 11 パターンに分類した．

表 1　感情の分類結果（個数）

あせり	平静	喜び	怒り	驚き	悲しみ
224	175	85	58	38	18
疑問	迷惑	あきれ	頼み	その他	
18	4	2	2	1	

4.2 考察

感情の分類では，ファミリー映画である「となりのトトロ」に感情がバランスよく出現していた．これより，ファミリードラマは，多くの感情のサンプルを集めることに適しているのではないかと考え，「ファミリー映画」での調査を中心に行った．

ファミリーであるジブリ作品では，感情の出現がバランス良く，サンプルの出現も多かった．平静，あせりでの感情は，各作品に常に多く登場していた．全作品 625 個中で，平静は 175 個，あせりは 224 個登場した．これにより，平静，あせりの感情音声は，サンプル収集がしやすいことが明らかになった．さらに，喜びの感情も 85 個登場し，怒りの感情も 58 個登場し，収集することが出来た．しかし，他の感情音声サンプルは，あまり多く収集出来ず，十分な信頼結果が得られないと考え，除外した．

そのため，以後は「喜び，怒り，平静，あせり」の感情音声サンプルについて分析していく．

5.音響特徴分析

5.1.1 pitch の分析

韻律情報には，声の大きさ，速度，高さ，長短などがある．過去，声の高さを表す pitch 情報を用い感情を評価する研究が多く行われ，数多くの成果を挙げている[2][3]．そのため本研究では，韻律情報として，pitch を使用する．

5.1.2 pitch 抽出方法

Pitch を求める方法には自己相関法，ケプストラム法，改良ケプストラム法，straight(YIN)，瞬時周波数心腹スペクトルを用いた pitch 抽出法[5]など様々な方法が存在する．ケプストラム法，自己相関法は，雑音環境下での影響を受けやすく，本研究には不適切である．高品質の

vocoder，straight-tempo は，高精度で安定なパラメータの抽出にこだわっている[9]．そのため，雑音環境下でも雑音の影響をあまり受けずに，正確な pitch を求めることが出来る．そのため，今回の研究では，straight 法を用いて行う．

5.1.3. pitch の正規化

発話者の声の高さにより，pitch の値は変化してしまう．それを防ぐために，pitch の正規化を行った．Pitch の正規化とは，発話者ごとに基本周波数を求め，その値と，発話した単語の pitch を比較する．基本周波数からの値の変化率により，発話者により変化する pitch を評価した．なお，発話者ごとの基本周波数は，その声優の「喜び」「怒り」「平常」「あせり」，それぞれのセリフを収録し，そのセリフ全ての pitch の平均を求めることで抽出した．さらに dynamic range も求め，さらに感情ごとの違いを明確にした．

5.1.4 pitch の抽出結果

表 2　pitch 抽出結果

	怒り	平静	あせり	喜び
Pitch(正規化後)	1.5658	0.7696	1.2860	1.0871
Dynamic range	3.877	2.2883	3.4686	2.4653

平静は，Pitch が，基本ピッチとほぼ同じ値を示し，Dynamic range は 2 倍程度の値を示した．怒りは，Pitch が最大の値を示した．Dynamic range は 3 倍を超える数値を示した．あせりは，Pitch が基本ピッチを超える値を示し，Dynamic range が最大の値を示した．
平静は，Pitch，Dynamic range ともに最低の値を示した．

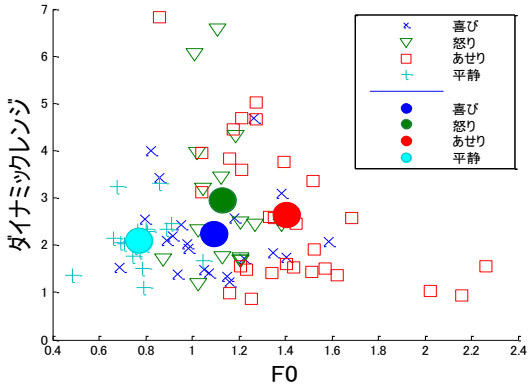

図 3　pitch と dynamic range 抽出結果

感情の分布を示すとともに,感情別にピッチ，ダイナミックレンジの平均の値も示した。(○：青が喜び、緑が怒り、赤があせり、水色が平静)

5.2.1 確率モデル

Pitch と，dynamic range の分布を正規分布と仮定し，平均と共分散を求め，確率モデルとした．モデルに感情が分類されていない音声を入力し，各カテゴリに対する尤度を求めて，最も尤度の大きいカテゴリに分類をした．

尤度は次の式で求める．

$$Emotion = \arg\max_{i}[p(x \mid C_i)]$$

$p(x \mid C_i)$ で尤度を求め，尤度が最も大きかった感情に分類する．

5.2.2 感情音声評価システム

作成した確率モデルを使用し，感情評価システムを作成した．

感情音声評価システムの実行例: 怒りの感情音声を分類
発話者の平均の基本周波数:240HZ
発話したセリフ「お前ふざけんなよ.

1位　怒り=0.8133
2位　喜び=0.4972
3位　あせり=0.1813

図 4　感情音声評価システム実行結果

このシステムを実行すると，感情が近い順に上から表示される．さらに感情ごとの尤度も表示される．今回の例では，1 位に怒りが表示されており，正しく発話できたことになる．

5.2.3 感情の認識率

発話者Ａ，Ｂ，Ｃ，Ｄの四人の発話を感情ごとに合計 10 個ずつ集めた．それにより,以下のような認識結果が得られた．

表 3　感情音声評価システムの感情分類結果

	怒り	平静	あせり	喜び
認識率	40%	80%	30%	70%

5.2.4 感情ごとの境界

確率モデルでは，入力された音声のピッチとダイナミックレンジの値により分類される感情が変化する．本研究では，学習データとして使用した Pitch と Dynamic range の値を，散布図として視覚化した(図 8 参照)．しかし，何の値が入れば，どの感情に分類されるのかを示す具体的な範囲が，図を見ただけでは認識できない．そのため，感情ごとの具体的な境界を示すことで，それぞれの感情が分類される範囲を視覚化した．さらに，表 10 により，誤認識されやすい感情を示したが，その結果が正しかったのかを調べる．

感情の境界線は，喜びとあせり，喜びと平静の様に 2 個 1 組で引く．確率モデルを使用し，尤度が同一になるピッチとダイナミックレンジの値を収集する．その値を曲線で結ぶことにより，境界を示す．

喜びは紫，怒りは水色，あせりは灰色，平静は黄色，により分類を示す．

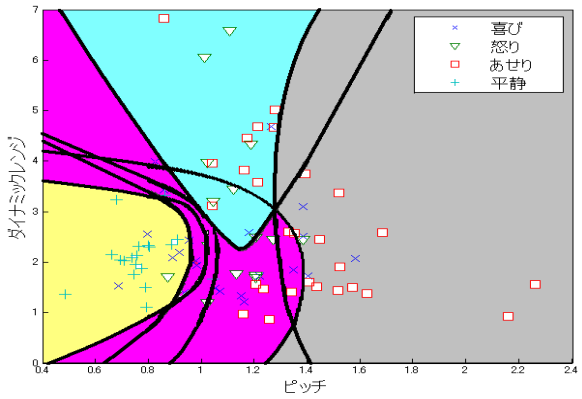

図 5 Pitch と Dynamic range の分布(改善後)

5.2.5 誤認識結果

誤認識された感情音声は，どの感情に分類されたのかを調査する．それにより，どの感情がどの感情に近いのかを調査する．

表 4 確率モデルでの感情分類結果
(縦軸：入力感情音声，横軸：分類された感情)

	喜び	怒り	あせり	平静
喜び	70%	10%	10%	10%
怒り	30%	40%	0	30%
あせり	30%	30%	30%	10%
平静	20%	0	0	80%

5.2.6 考察

境界線を引くことにより，認識される範囲をより明瞭にすることが出来た．認識率が，最も高くなった平静は，値もまとまっている．さらに，学習データが認識範囲に平静の値がほとんど収まっている．そのため，認識率が高くなったと考えられる．喜びは，値にばらつきは見られるが，ほとんどの値が，喜びの認識範囲に収まっている．70%の認識率を得られたのはこのためだと考えられる．学習データを追加することにより，値が安定し，認識率はさらに向上すると考えられる．それに対し，あせり，怒りは，認識率が低い．怒り，あせりは，学習データが認識範囲に収まっておらず，別の認識範囲に出現しているものが多々見られる．この値のばらつきが認識率低下の原因と考えられる．これより，怒り，あせりは，pitch，dynamic range 情報だけでは，特徴がうまく抽出できないと考えられる．これは，新しい特徴パラメータ(パワー，長短，速度など)を考慮し，確率モデルを作成することで向上すると考えられる．さらに，表 10 での結果と，図 8 での結果を合わせて検証する．喜びは，他の感情に平均的にご認識されていた．図 8 でも喜びの Pitch，Dynamic range の分布図の値は，中心に散布しており，他の感情に認識されやすいことが分かる．さらに，平静，あせりはそれぞれが誤認識されにくいという結果になった．これは図 8 からも値，認識範囲が離れているためだと考えられる．図 8 より，怒り，あせりは，近い値が多いため，誤認識される確率は高い．しかし，表 10 の結果では，あせりが怒りに，誤認識される確率が 30%，怒りがあせりに誤認識される確率は 0%となっている．評価データのサンプルを増やすことにより，あせりと怒りの誤認識の確率は上がる可能性がある．

6.結び

本研究は，「若手俳優，声優の育成支援のための感情音声評価システム作成」を目的として行った．

まず，感情音声サンプルを映画，テレビ番組から約 600 個収集した．STRAIGHT を用いることで，プロの俳優の感情音声の pitch，dynamic range を抽出した．そして求めた韻律パラメータを学習データとして使用することで確率モデルを作成した，さらに，確率モデルを使用し，感情音声評価システムを作成した．つぎに，学習データを視覚的に把握するため，Pitch と Dynamic range の値を散布図に表した．作成した確率モデルでは，喜び 70%，平静は 80%の認識率を得ることが出来た．あせり 30%，怒りは 40%と，十分な認識結果が出せなかった．認識率が低い理由を考察するため，Pitch，Dynamic range の特徴量の分布に対し，確率モデルのそれぞれの感情の分類範囲を示した．これにより，あせり，怒りの学習データは，それぞれの認識範囲に収まっておらず，値がまとまっていないことが分かった．これは，評価データを追加し,新たな特徴パラメータ（パワー，長短，速度など）を追加することで，さらに向上出来ると考えた．確率モデルの認識精度を向上させていき，感情のカテゴリを追加していくことにより，本研究の目的である，育成システムを作成することが可能であると考えられる．

文献

[1] 西香織，”映像作品におけるシーンの印象別背景音楽選定のための音響特徴量分析”，法政大学情報科学部ディジタルメディア学科卒業論文，p28，2007
[2] 平賀裕，斎藤善行，森島繁生，原島博，”音声に含まれる感情情報抽出の一検討”，電子情報通信学会技術研究報告．HC，ヒューマンコミュニケーション，vol.93，No.439，p8，
[3]平館　郁雄，”怒りの感情音声における音響特徴量と感情知覚との関係に関する研究”，北陸先端科学技術大学院大学情報研究科，修士論文，p67，202
[4]P．Ekman，W．V．Friesen(工藤力訳):表情分析入門 表情に隠された意味をさぐる，工藤力訳，第 7 版，誠信書房 1987
[5]田中智宏，益子貴史，小林隆夫，”瞬時周波数振幅スペクトルに基づく pitch 抽出法の検討”，電子情報通信学会技術研究報告. SP, 音声 vol.100，No726，p8
[6]深澤友貴”バランス歌唱コーパスを用いた歌唱音高コントロールの統計的分析”，法政大学情報科学部ディジタルメディア学科卒業論文，2007，p19
[7]佐藤秀明，赤松則男，”ニューラルネットワークによる感情音声の分類”，電子情報通信学会技術研究報告. NC, ニューロコンピューティング，vol.101，No154，p5
[8]河原英紀，”聴覚の情景分析が生み出した高品質 VOCODER:STARAIGHT”，日本音響学会誌，Vol.154，No7，p15