

感情表現音声の韻律的特徴 ―聴取実験による検討―

鶴 真紀子・武 田 昌 一^{*1}

^{*1} 帝京平成大学

A Study of Prosodic Features of Emotional Speech

TSURU Makiko and TAKEDA Shoichi^{*1}

^{*1} : Teikyo Heisei University

This paper is concerned with the prosodic features of "anger" expressions in Japanese speech. We compared aural impressions of speech with the physical parameters of the prosodic features. One female announcer uttered 4-mora and 6-mora words with four degrees of anger: neutral, displeasure, anger, and fury. Listening tests were conducted using these words as speech samples to quantify aural impressions. Prosodic-feature parameters were speech rate and F0 parameters. Analysis results were as follows: (1) The prosodic-feature parameters that correlate to the rate of coincidence between the listening test result and the speaker's intention were the magnitude of accent command and minimum and maximum fundamental frequencies, (2) The prosodic-feature parameters that correlate to the rate of identification as a specific degree of emotion were the magnitude of accent command and maximum fundamental frequency, and (3) The coincidence rates for displeasure (35.0%) and anger (23.3%) were low compared with that for fury (70.8%).

Key words : Anger, Prosody, Emotional Speech, Listening Test

キーワード : 怒り, 韻律, 感情音声, 聴取実験

はじめに

現代のマルチメディアの技術開発はとどまることがなく, 研究, 開発そして実用化へと進められている。そもそも, “マルチメディア” とは “複合媒体”, “複数媒体” といわれることが多い。マルチメディア技術開発の目的は人間と機械のより良いコミュニケーション手段を得ることである。

“マルチメディア” 研究の目的の一つに, われわれ人間そのものの感覚を再現することがある。人間が目, 耳, 鼻, 口, 皮膚をもって他の人間や機械などのあらゆる環境・情報とのコミュニケーションをとる場合, それに種々の感

情と判断が伴う。これらの技術はユニバーサルコンピューティングへの一助と考えられている。

そこで, 筆者らはマルチメディアのうち耳, つまり音声について研究を進めている。近い将来には, 機械の合成音声と人間, あるいは人間の感情音声と機械とが互いに感情の交流がもてるような再現性のあるコミュニケーションツールとしたいと思っている。

これまでも音声合成をより自然音声に近づけるために, 人間の感情音声を分析した研究が数多くされてきた⁽¹⁾⁻⁽³⁾。しかし, 人間の感情に近づけるためには, 聞き手である人間がどのように聞き取るかを把握することは不可欠である。そのための聴取実験も報告されているが, その

多くは、数種類の感情（怒り・喜び・悲しみ・感謝など）の聞き分けが中心であり、その正解率や判断に起因するパラメータなどが解析されたものである。そして、それらの研究成果は合成音声の実現へと導くものである⁽⁴⁾⁻⁽⁹⁾。

今回、最も感情が表現されやすいとされている「怒り」について、感情の度合いを聞き分ける聴取実験を試みた。今回の目的は、話者が発する感情音声と聞き手の受けた感情の違いが、韻律のどのパラメータに起因しているのか確かめることに重点をおき、聴取実験を行う。

聴取実験では音声を分析したときの韻律的特徴パラメータとして、「発話速度」「最高・最低基本周波数」「パワー」「アクセント指令」「フレーズ指令」などが多く用いられる。ここで「韻律」とは「声の大きさ」、「声の高さ」、「発話速度」の総称である。

声は「肺に蓄えられた空気を気流として喉頭に送り込み、声帯を振動させることによって作られる。」⁽¹¹⁾

「声の高さ」は声帯の振動回数が関与しており、高い声は振動回数が多く、低い声は少ないことが知られている。声帯開閉の繰り返しの基本周期から「基本周波数」が得られる。「基本周波数」は声の高さの物理的特徴量である。

「声の大きさ」は声帯の振幅の大きさに関係している。声帯の振幅が大きいほど声が大きくなり、このときに「パワー」の値は大きくなる。

音 声 試 料

- ・話者：NHK 女性アナウンサー 1 名。
- ・単語：「まなまな」。4 モーラ、アクセントは平板型、字面が感情に影響しない無意味音。感情のない「平常」および「軽い怒り」「怒り」「激怒」の感情を込めた音声を各 5 回発声。発声の際の状況設定と感情の度合いの基準は話者に一任。今回は、「軽い怒り」「怒り」「激怒」各 5 サンプル計 15 音声と比較対象となる「平常」音声 1 サンプルを使用。

実 験 方 法

- ・最初に「平常」サンプルを聞かせ、続いてランダムに並べた「感情」サンプルを 2 回聞かせる。これは、森山ら⁽⁹⁾や向笠ら⁽¹⁰⁾を参考にした。感情音声のみを聞かせると、被験者が耳馴れし、均等な評価ができなくなる恐れがあるので、各感情音声の前に平常音声を流し、比較評価ができるようにした。
- ・被験者には次のように説明する。『今から、「まなまな」という単語を聞いてもらいます。その単語には感情のない「平常」と「軽い怒り」「怒り」「激怒」の感情がそれぞれ込められています。まず、感情のない「平常」音声を 1 回流し、続けて「感情音声」を 2 回流します。皆さんは感情音声を聞いて、手元のシートにその段階にマークしてください。少し間をおいて次に音声を同じように「平常」音声を 1 回、「感情」音声を 2 回流していきます。なお、音声は一回しか聞けません。』
- ・同じメーカーのヘッドホンで最大音量に設定して実験を行った。
- ・被験者数 12 人
- ・静かな環境で、ひとり一台の機器で同時に音声を聞かせた。
- ・同じ被験者に 15 サンプルの聴取実験を 2 回行った。1 回目と 2 回目はサンプル順を変更した。

結 果

聴取実験に使用した 16 サンプル（平常 1 サンプル含む）の韻律的特徴を表 1 に示す。

次に、聴取実験の結果（24 データ＝12 人×2 回）について述べる。

(1) 一致率

音声試料を聴取した被験者たちの一致率と回答および度合別の一致率を表 2 に示す。

表 2 から、話し手の感情音声聞き手に伝わったとき、その感情の度合いを正確に受信されることの少ないことが分かる。特に「怒り 2」と

「怒り4」は4.2%ときわめて低い一致率である。一方、「激怒2」を除く「激怒」音声を見ると65%以上の一致率になっている。感情の度合別で見ると、一致率は「激怒」(70.8%)>「軽い怒り」(35.0%)>「怒り」(23.3%)とな

る。中間の「怒り」については判別しにくいことがわかった。

次に、このような一致率の相違を、韻律的特徴から明らかにしていく。

表1 実験に使用したサンプルの韻律的特徴

怒りの度合い	平均発話速度 (mora/s)	フレーズ指令 の大きさ	アクセント指令 の大きさ	最低基本周波数 (Hz)	最高基本周波数 (Hz)
平常	5.7	0.29	0.15	160.5	271.0
軽い怒り1	6.4	0.51	0.28	120.0	240.2
軽い怒り2	6.4	0.44	0.31	118.7	229.3
軽い怒り3	6.0	0.44	0.37	116.9	237.6
軽い怒り4	6.6	0.55	0.36	107.2	233.3
軽い怒り5	6.1	0.47	0.44	109.3	234.2
怒り1	6.1	0.49	0.57	112.6	275.1
怒り2	5.9	0.61	0.47	110.1	264.3
怒り3	6.2	0.53	0.44	115.0	261.8
怒り4	5.9	0.55	0.45	113.4	258.5
怒り5	6.0	0.59	0.40	115.3	265.6
激怒1	5.9	0.51	0.77	118.7	303.1
激怒2	6.2	0.55	0.55	113.2	289.4
激怒3	6.0	0.54	0.63	115.6	300.0
激怒4	6.2	0.48	0.58	113.6	296.5
激怒5	6.3	0.51	0.66	114.7	300.8

表2 一致率と被験者の回答

感情の度合	各サンプルの 一致率 (%)	被験者の回答数				平均一致率 (%)
		平 常	軽い怒り	怒 り	激 怒	
軽い怒り1	25.0	17	6	1	0	35.0
軽い怒り2	45.8	11	11	2	0	
軽い怒り3	33.3	13	8	3	0	
軽い怒り4	41.7	12	10	2	0	
軽い怒り5	29.2	14	7	2	1	
怒り1	41.7	2	11	10	1	23.3
怒り2	4.2	15	8	1	0	
怒り3	16.7	7	13	4	0	
怒り4	4.2	14	8	1	1	
怒り5	50.0	1	11	12	0	
激怒1	87.5	0	1	2	21	70.8
激怒2	37.5	0	3	12	9	
激怒3	83.3	1	0	3	20	
激怒4	66.7	0	2	6	16	
激怒5	79.2	0	1	4	19	

(2) 正解率と韻律的特徴の相関

本稿で扱う韻律的特徴は表1のとおり「平均発話速度」「フレーズ指令の大きさ」「アクセント指令の大きさ」および「最低・最高基本周波数」とする。また「一致率」および一致率との「相関係数」を追加したものを表3に示す。

前述のとおり、一致率は「怒り」の度合間でばらつきがあるため、ここでは、度合別に韻律的特徴を分析していく。そうすることによりそれぞれの特徴が顕著に現れると思われる。(以下、各韻律をパラメータと記す)

・「軽い怒り」

一致率との相関において、「平均発話時間」以外のパラメータは負の傾向にある。特に顕著なものは「最高基本周波数」の $r = -0.87$ である。「最高基本周波数」が低くなればなるほど一致率が上がっていることになる(図1)。「最高基本周波数」が最も低い「軽い怒り2」の一致率が高いことが分かる。

・「怒り」

「フレーズ指令」以外は正の傾向にある。強い傾向はないものの「最高基本周波数」に中程度の強い傾向が見られる(図略)。

表3 各サンプルの韻律的特徴と一致率および相関係数

怒りの度合	平均発話速度 (mora/s)	フレーズ指令 の大きさ	アクセント指令 の大きさ	最低基本周波数 (Hz)	最高基本周波数 (Hz)	一致率 (%)
軽い怒り1	6.4	0.51	0.28	120.0	240.2	25.0
軽い怒り2	6.4	0.44	0.31	118.7	229.3	45.8
軽い怒り3	6.0	0.44	0.37	116.9	237.6	33.3
軽い怒り4	6.6	0.545	0.36	107.2	233.3	41.7
軽い怒り5	6.1	0.47	0.44	109.3	234.2	29.2
相関係数	0.48	-0.12	-0.12	-0.16	-0.87	

怒りの度合	平均発話速度 (mora/s)	フレーズ指令 の大きさ	アクセント指令 の大きさ	最低基本周波数 (Hz)	最高基本周波数 (Hz)	一致率 (%)
怒り1	6.1	0.49	0.57	112.6	275.1	41.7
怒り2	5.9	0.61	0.47	110.1	264.3	4.2
怒り3	6.2	0.53	0.44	115.0	261.8	16.7
怒り4	5.9	0.55	0.45	113.4	258.5	4.2
怒り5	6.0	0.59	0.40	115.3	265.6	50.0
相関係数	0.57	-0.24	0.11	0.49	0.68	

怒りの度合	平均発話速度 (mora/s)	フレーズ指令 の大きさ	アクセント指令 の大きさ	最低基本周波数 (Hz)	最高基本周波数 (Hz)	一致率 (%)
激怒1	5.9	0.51	0.77	118.7	303.1	87.5
激怒2	6.2	0.55	0.55	113.2	289.4	37.5
激怒3	6.0	0.54	0.63	115.6	300	83.3
激怒4	6.2	0.48	0.58	113.6	296.5	66.7
激怒5	6.3	0.51	0.66	114.7	300.8	79.2
相関係数	-0.45	-0.41	0.80	0.76	0.99	

・「激怒」

「アクセント指令」($r=0.80$), 「最低基本周波数」($r=0.76$)および「最高基本周波数」($r=0.99$)にかなり強いプラスの相関がみられる(図2)。これについては, 武田ら⁽¹⁾の報告にもあるように, 聞き手にとっても度合いを感知するパラメータとして「最高基本周波数」が強く関与していることが再確認された。

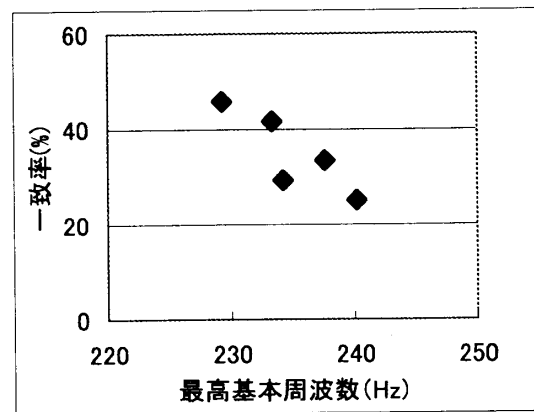
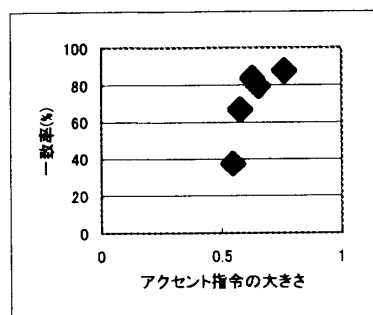
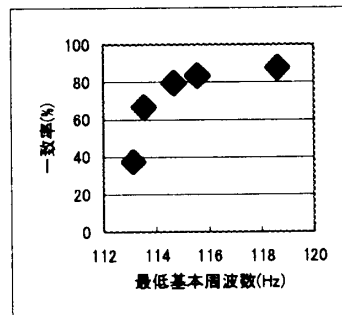


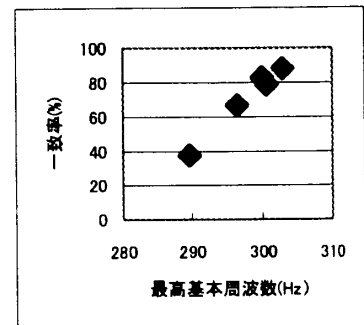
図1 一致率と最高基本周波数の相関
(軽い怒り)



(a) アクセント指令の大きさ



(b) 最低基本周波数



(c) 最高基本周波数

図2 一致率とパラメータの相関(激怒)

(3) 韻律的特徴と同定数

これまでの, 話し手の感情の度合いを中心に「一致率」との相関を見てきたが, 本実験の目的である「聞き手」の受信した情報を中心に, 各パラメータをみる必要がある。そこで, 表1に被験者の回答を追加したものをパラメータごとにソーティングして, パラメータの数値変化による同定数の推移に規則性があるか確認した。その結果, 「フレーズ指令の大きさ」, 「平均発話時間」および「最低基本周波数」において同定数との関連性は見られなかったが, 「アクセント指令の大きさ」と「最高基本周波数」には関連性は見られた(表4)。また, 両パラメータにおいてそれぞれ被験者の同定数間に有意な差($p<0.01$)が認められ, 度合の認識は, パラメータの数値変化に伴うものであることが明らかになった。

次に, 表4におけるパラメータの具体的な境界値については, 「アクセント指令の大きさ」が0.37以下であればほぼ半数は「平常」と認識され, 0.58以上であれば「激怒」と認識され, その一致率も65%以上である。「最高基本周波数」の258.5Hz以下であれば「平常」と認識された。ただ今回の実験で比較のために使用した「平常」サンプル音声の最高基本周波数が271.0Hzだったため, このような認識になったとも推察される。また, 296.5Hz以上であれば, ほぼ75%以上の一致率で「激怒」と認識される。

今回の聴取実験の結果から「軽い怒り」と「怒り」を識別することが困難であることが判明した。

表4 サンプルの韻律的特徴と被験者の同定数（最多同定数に網掛け）

	アクセント 指令の大きさ	平常	軽い 怒り	怒り	激怒
軽い怒り 1	0.28	17	6	1	0
軽い怒り 2	0.31	11	11	2	0
軽い怒り 4	0.36	12	10	2	0
軽い怒り 3	0.37	13	8	3	0
怒り 5	0.40	1	11	12	0
軽い怒り 5	0.44	14	7	2	1
怒り 3	0.44	7	13	4	0
怒り 4	0.45	14	8	1	1
怒り 2	0.47	15	8	1	0
激怒 2	0.55	0	3	12	9
怒り 1	0.57	2	11	10	1
激怒 4	0.58	0	2	6	16
激怒 3	0.63	1	0	3	20
激怒 5	0.66	0	1	4	19
激怒 1	0.77	0	1	2	21

	最 高 基 本 周波数	平常	軽い 怒り	怒り	激怒
軽い怒り 2	229.3	11	11	2	0
軽い怒り 4	233.3	12	10	2	0
軽い怒り 5	234.2	14	7	2	1
軽い怒り 3	237.6	13	8	3	0
軽い怒り 1	240.2	17	6	1	0
怒り 4	258.5	14	8	1	1
怒り 3	261.8	7	13	4	0
怒り 2	264.3	15	8	1	0
怒り 5	265.6	1	11	12	0
怒り 1	275.1	2	11	10	1
激怒 2	289.4	0	3	12	9
激怒 4	296.5	0	2	6	16
激怒 3	300.0	1	0	3	20
激怒 5	300.8	0	1	4	19
激怒 1	303.1	0	1	2	21

む す び

今回は、話者が発する感情音声と聞き手の受けた感情の相違と、起因するパラメータを確かめることに重点をおいて聴取実験を行った。女性アナウンサー1名の音声を対象に実験を行った結果、以下の結論が得られた。

- (1) 一致率：「激怒」(70.8%)>「軽い怒り」(35.0%)>「怒り」(23.3%)となる。中間にあたる「怒り」については判別しにくいことがわかった。
- (2) 一致率とパラメータの相関：「軽い怒り」において「最高基本周波数」が $r = -0.87$ と強い負の相関が認められ、「怒り」において「最高基本周波数」($r = 0.68$)に中程度に強い傾向が見られる「激怒」においては、「アクセント指令」($r = 0.80$)「最高基本周波数」($r = 0.99$)および「最低基本周波数」($r = 0.76$)にかなり強いプラスの相関がみられた。
- (3) 韻律的特徴と同定数：「アクセント指令の大きさ」と「最高基本周波数」には被

験者の同定数に有意な差が認められた($p < 0.01$)。

今回の1名の話者が発声する音声については、怒りの度合を識別するパラメータとして、「最高基本周波数」と「アクセント指令の大きさ」が大きく関与していることが判明した。しかし、更にその境界値を明らかにするために実験を重ね、データを補完していく必要がある。

今後の問題点としては、実験の際に比較対象として提示する「平常」サンプルの韻律的特徴に配慮する必要がある。更には、合成音声を使った聴取実験で起因するパラメータの境界値を明らかにしていく考えである。

謝 辞

今回の聴取実験の被験者として快く協力してくれた情報社会学科1年生の有志の学生にお礼を申し述べたい。

参 考 文 献

- (1) 武田昌一・大山玄・朽谷綾香・西澤良博, 日本語音声における「怒り」を表現する韻律的特徴の解析, 日本音響学会誌, 58巻, 9号, pp561-568, 2002
- (2) 武田昌一・西澤良博・大山玄, 「怒り」の音声の特徴分析に関する1考察, 信学技報, SP2000-164, pp33-40, 2001
- (3) 川波弘道・広瀬啓吉, 態度・感情音声における韻律的特徴の考察, 信学技報, SP97-67, pp73-80, 1997
- (4) 飯田朱美・伊賀聡一郎・安村通晃, 音声に見られる感情表現について: 聴取実験結果の分析, 情報処理学会, 音声言語処理15-20, pp113-118, 1997
- (5) 飯田朱美・伊賀聡一郎・安村通晃, 韻律にみられる感情表現について, 情報処理学会第54回全国大会, 1997
- (6) 河津恵・藁科綾子・石橋俊, 「怒り」と「喜び」の感情の音声合成と聴取実験による評価, 日本音響学会講演論文集, pp191-192, 1998
- (7) 飯田朱美・ニック キャンベル・安村通晃, 感情表現が可能な合成音声の作成と評価, 情報処理学会論文集, Vol.40, No.2, pp479-486, 1999
- (8) 小池和仁・斎藤博昭・中西正和, 感情音声の合成, 信学技報, NLC 98-43, pp83-88, 1998
- (9) 森山剛・斎藤英雄・小沢慎治, 音声における感情表現ごと感情表現パラメータの対応付け, 電子情報通信学会論文誌, Vol.J82-D-II, No.4, pp703-711, 1999
- (10) 向笠美幸・目加田慶人・春日正男・松本修一・小池淳, 感情を含む音声信号の特徴解析について, 信学技報, HCS 2000-3, pp17-22, 2000
- (11) 日本音響学会編, 新版音響用語辞典, p309, コロナ社, 東京, 2003

(2004年3月13日 受稿)