E-042

韻律的特徴の分析に基づく感情制御規則の導出と合成音声による評価

Rules to Control the Prosodic Features of Emotional Utterances at Several Degrees

河津 宏美[†] *Hiromi Kawatsu*

大野 澄雄[†] *Sumio Ohno*

1. はじめに

感情が表現された音声を合成することを目的とし，複数の程度の感情情報を含む音声を対象に，感情の程度に対する音声の韻律の変化について検討を行っている．韻律の制御は，より自然な音声を合成するために重要であると考えられており，感情音声の分析 [1] のほか，対話調音声の分析 [2] などの研究も行われている．

本稿では，感情の種類として「怒り」「恐れ」を取り上げ，弱・中・強の 3 段階の程度でそれぞれ感情を表現した発話を分析の対象とした．韻律的特徴のうち，基本周波数パターン（以後，F_0 パターン）および発話速度について，感情の程度との関係を検討し，感情制御規則の導出を行った．導出した規則を適用した音声を合成し，合成音声が表現する感情の程度を主観評価した結果について述べる．

2. 音声資料

2.1 録音条件

発話テキストとして，4 文節からなる文を「怒り」「恐れ」について 10 文ずつ用意した.「怒り」「恐れ」の感情を表現しやすい文を用意し，それぞれの感情が表現された典型的な発話の収集を目指した．表 1 に発話テキストの例を示す．話者は演劇経験のある成人 8 名（男性 6 名，女性 2 名）で，弱・中・強の 3 段階の程度でそれぞれの感情を表現した発話を収録した．発話の際，指定した感情を表現しやすくするために，発話の状況設定を行い発話テキストと共に話者に提示した（表 2 参照）．また，比較のために，同一の発話テキストに対し，感情を込めない中立な発話を収録した．録音は，簡易防音室内において「中立」→「中」→「弱」→「強」の順に 3 回繰り返し行った．

表 1. 発話テキストの例

怒り	1. 私の家族の悪口を言うな
	2. まじめに練習をしないなら帰れ
	3. あんたに何の権限があるのですか
恐れ	1. 霧で前が少しも見えないよ
	2. 滑りそうな凍った階段を下りるのか
	3. 何故か写真の音楽家が泣いている

表 2. 発話の状況設定の例

怒り	テキスト	その命令に従うのは無理だ
	発話状況	理不尽なことばかりされ，いくら上司とはいえ，我慢できずに反抗．怒って一言．
恐れ	テキスト	百獣の王に襲われたら勝ち目はないな
	発話状況	動物園から逃げ出したライオンを偶然発見し，目が合ってしまった．襲われることを想像して一言．

2.2 主観評価

話者の意図によって数段階の程度で感情を表現した収録音声に対し，聴取実験を行った．この聴取実験は，話者の意図した感情表現の程度と，聞き手が受容した感情の程度との一致度を把握することが目的である．被験者は大学 4 年生の男女 6 名である．話者の意図した感情の程度ごとに聞き手側の観点から感情の程度の主観評価値の分布を求め，それぞれの程度間で片側検定による有意差検定を行った．その結果，MTI (男性話者) の音声資料において，話者の意図での感情の程度が最も的確に聞き手に伝達されたことが分かった．以下，MTI の音声資料について韻律的特徴の分析を行った結果について示す．

3. 韻律的特徴の分析手法

3.1 F_0 パターンの分析

F_0 パターンの分析には，藤崎らによって提案された F_0 パターン生成過程モデル [3] を用いた．図 1 に F_0 パターン生成過程モデルを示す．このモデルは，対数 F_0 パターンが句頭から句末に向かって上昇とその後の緩やかな下降を示すフレーズ成分と，語のアクセントに対応して局所的な起伏を示すアクセント成分，および発話単位中でほぼ一定値をとるベースライン成分（基底周波数）の総和として表現できるとするものである．このモデルの入力パラメータは，音声を生成する人間の生理的・物理的な特性を捉えたもので，言語的内容とも整合した制御パラメータが得られることが確認されている．入力パラメータである基底周波数，フレーズ指令の大きさとその生起位置，および，アクセント指令の大きさとその生起タイミングについて，録音した音声の F_0 パターンを対象に分析した．

具体的にはまず，収録した音声資料を 10 kHz・16 bit でデジタル化し，LPC 予測残差に対する変形自己相関関数法を用いて 10 ms 間隔で F_0 の値を抽出した．その F_0 パターンに対し，視察によりモデルのパラメータを定め，次にそれを初期値として，AbS 法に基づき最良近似を与えるパラメータを求めた．

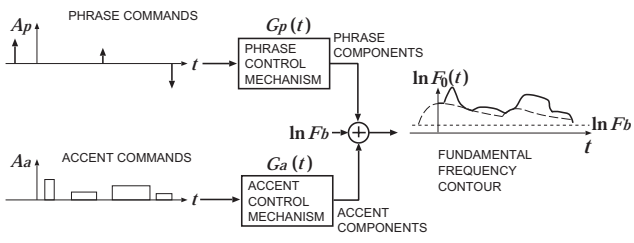

図 1. F_0 パターン生成過程モデル

[†]東京工科大学, Tokyo University of Technology

3.2 発話速度の分析

発話速度に関して，感情の程度の影響を検討するため，文節の継続時間長比の変化について検討を行った．まず，音声認識ソフトウェア Julian [4] を用いて，文節の継続時間長を求めた．その際，言語モデル・音響モデルおよび辞書に関しては，付属の標準モデルを使用した． Julian では，音声ファイルと共に書き下しファイルを与えることで，書き下しの内容に従ったセグメンテーションを行うことができる．ただし，セグメンテーションの単位は 10 ms である．Julian での自動セグメンテーションの後，目視による手動での修正を行った．ここで，継続時間長比 (R) は，以下のように定義した．

$$R = d_{\mathrm{target}} / \bar{d}_{\mathrm{neutral}} \quad (1)$$

ただし，$\bar{d}_{\mathrm{neutral}}$ は同一テキストに対して収録した 3 回の中立発話の文節の継続時間長の平均を示し，d_{target} は対象発話の継続時間長を示す．なお，発話速度の分析において休止は重要な要素となり得るが，大部分の音声データである休止を含まない発話のみを扱うこととし，休止の有無による影響を除外した．

4. 分析結果

韻律的特徴である F_0 パターンや発話速度は，感情の種類とその程度のほか，表 3 にあげる文の言語的要因の影響を受けると考えた．そこで，感情の程度の違いが韻律的特徴にどのように影響するかについて，言語的要因との関連から検討した．図 2 に感情の程度に対する各韻律パラメータの変化をまとめた．

4.1 F_0 パターンの分析結果

(a) 基底周波数

図 2(a) には，感情の程度ごとに基底周波数の平均値と標準偏差を示した．図中の直線は，すべての発話を対象に求めた基底周波数の回帰直線を示す．なお，回帰直線を求める際，中立発話を感情の程度 0 とし，弱の発話を 1，中の発話を 2，強の発話を 3 と尺度化した．

「怒り」「恐れ」いずれの感情においても，感情の程度が強くなるにしたがって，基底周波数は高くなる傾向にある．ただし,「怒り」では感情の程度が小さいときはほぼ一定あるいはわずかに高くなる傾向がみられた．一方,「恐れ」では，感情の程度の影響が大きく，感情の程度が強くなるにつれて，ほぼ一様に高くなる傾向があった．

表 3. 韻律パラメータの変化に影響を与える要因

	制御要因（とり得る値）
1.	感情の程度（無，弱，中，強）
2.	文頭からの文節位置（1，2，3，4）
3.	後続の文節境界の深さ（1，2，3，文末）
4.	文節のモーラ数（2〜7）
5.	アクセント型（平板型，頭高型，起伏型）
6.	促音の数（0〜1）
7.	撥音の数（0〜2）
8.	長母音の数（0〜2）

(b) フレーズ指令

文中でのフレーズ指令の生起に関しては，文節境界の枝分かれ種別ごとに，感情の程度による影響の受け方に違いがみられた．その他，先行フレーズ指令からのモーラ数と，直後の韻律語のモーラ数とに影響を受けることを確認した．そこで，文節境界においてフレーズが生起するか否かを予測することを目的に，感情の程度，文節境界の枝分かれ種別，先行フレーズ指令からのモーラ数，直後の韻律語のモーラ数に基づいて判別分析を行い，フレーズ指令の生起に関して規則化をした．

フレーズ指令の大きさは，その生起位置が文頭の場合と文中の場合とで，感情の程度に対する変化の傾向に違いがみられた．図 2(b) には，文頭・文中のフレーズ指令ごとに，感情の程度に対するフレーズ指令の大きさの平均値と標準偏差を示した．図中の直線は，回帰直線である．中立の発話の場合，フレーズ指令の大きさは，文中に比べて文頭のほうが大きい．いずれの感情においても，感情の程度が強くなると，文頭のフレーズ指令の大きさは小さくなるのに対し，文中のフレーズ指令は大きくなり，感情の程度が強において，発話内のすべてのフレーズ指令の大きさの差が小さくなった．

(c) アクセント指令

表 4 に，河井ら [5] によって導入されたアクセント指令の大きさの種別を示す．図 2(c) には，感情の程度に対するアクセント指令の大きさについてすべての発話を対象に回帰直線を求め，その種別ごとに示した．感情の程度が変化しても，D_H, F_H のアクセント指令の大きさの変化は少なく，D_M, F_M に関して，感情の程度が強くなると，アクセント指令の大きさに増大の変化傾向がみられた．その傾向は，特に「怒り」の感情において顕著にみられた．

アクセント指令の生起タイミングに関して，分節境界上の基準点 [5] に対し，F_0 パターン生成過程モデルを用いた分析によって得られたアクセント指令の立ち上がり・立ち下がりのタイミングの相対時間が，感情およびその程度の影響を受けるかについて検討を行った．その結果，感情の種類やその程度の影響はほとんど見られず，アクセント指令の立ち上がり・立ち下がりのタイミングはアクセント型にのみ依存することを確認した．表 5 に中立発話のアクセント指令の生起タイミングに関して基準点との相対時間の平均値を示した．

表 4. アクセント指令の大きさの種別

	記号	説明
起伏式	D_H	句中，最初に現れる起伏式韻律語
	D_M	D_H 以外の起伏式韻律語
平板式	F_H	句中，D_H 以前の平板式韻律語
	F_M	F_H 以外の平板式韻律語

表 5. アクセント指令の生起タイミング

	立ち上がり [s]	立ち下がり [s]
平板型	−0.088	−0.064
頭高型	−0.068	0.040
起伏型	−0.111	−0.030

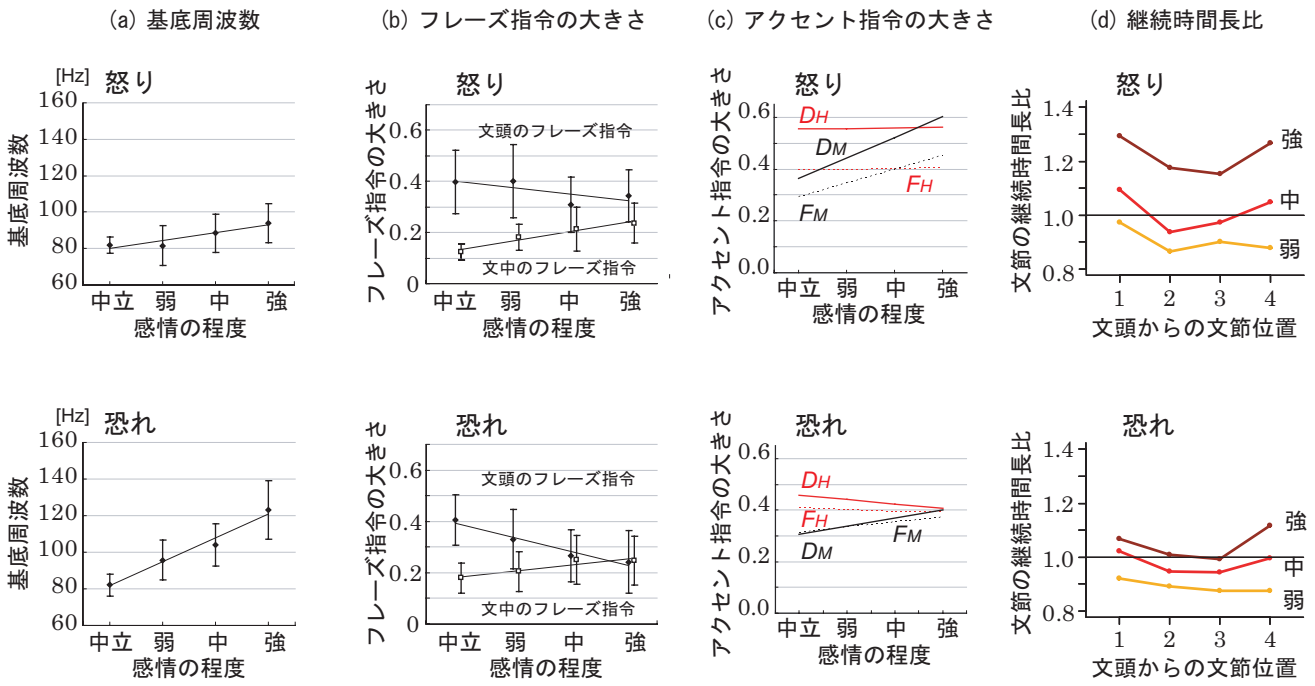

図 2. 感情の程度に対する韻律パラメータの変化

4.2 発話速度の分析結果

表 3 に示す制御要因の影響に関して，発話速度の変化を検討するため，ステップワイズ回帰モデルの変数増加法に基づき，変数の選択を行った．その結果，感情の程度および文頭からの文節位置が上位で選択され，その他には，当該文節のモーラ数が選ばれた．図 2(d) に文頭からの文節位置に着目し，感情の程度に対する継続時間長比を示した．いずれの感情においても，すべての文節において感情の程度の影響がみられ，特に第 4 文節において，感情の程度の影響が大きくみられた．また，「恐れ」よりも「怒り」で，感情の程度の影響が強くみられた．

4.3 感情制御規則の導出

基底周波数，フレーズ指令の大きさ，アクセント指令の大きさに関して，図 2(a)〜(c) に示す回帰直線に基づき，F_0 パターンの各モデルパラメータの大きさを決定する．フレーズ指令の生起位置に関しては，判別分析により得られた予測値に基づき，文節境界にフレーズ指令が生起するか否かを決定する．アクセント指令の立ち上がり・立ち下がりのタイミングは，アクセント型ごとの基準点との相対時間として一定の値を与える．

また，発話速度に関しては，文節単位に制御を行い，図 2(d) に示す継続時間長比の平均値に基づき中立発話の発話速度を線形に伸縮することにより決定する．

5. 合成音声による評価

5.1 合成音声の作成

4.3 節で導出した規則を適用した合成音声を作成し，その合成音声の感情の程度を主観評価することにより規則の妥当性を確かめる．評価に用いる音声を合成するために，closed データとして感情制御規則の導出に使用したテキスト 10 文からランダムに 5 文を選択し，open データとして別のテキスト 5 文を用意した．closed データと open データは同一話者によるものであり，いずれも中立および 3 段階で感情を表現した実発話が収録されている．着目した特徴以外の変動要因を排除するため，波形合成エンジン MBROLA [6] を用いて，F_0 パターンと各音の時間構造を指定することにより合成音声を作成した．

- F_0 パターンについては，生成過程モデルのパラメータを次の 3 通りの方法で指定した．
  (1) 中立の実発話に対して最良近似を与えるパラメータ
  (2) 感情を込めた実発話に対して最良近似を与えるパラメータ
  (3) 感情制御規則に基づいて決定したパラメータ
- 各音の時間構造に関しては，次の 3 通りの方法で音素の継続時間長を決定した．
  (1) 中立の実発話の発話速度を模擬
  (2) 感情を込めた実発話の発話速度を模擬
  (3) 感情制御規則に基づいて各音の継続時間長を決定

以上，F_0 パターンと時間構造を独立にコントロールした合成音声を感情ごとに計 9 セット作成した．感情を込めた音声として，それぞれの感情に対して 3 段階の感情の程度に対応する合成音声が含まれる．

5.2 作成した合成音声に対する聴取実験

2.2 節で行った聴取実験と同じ方法で，作成した合成音声に対して，各合成音声が表現する感情の程度を主観評価した．聴取実験により，導出した感情制御規則の妥当性について検証することが目的である．

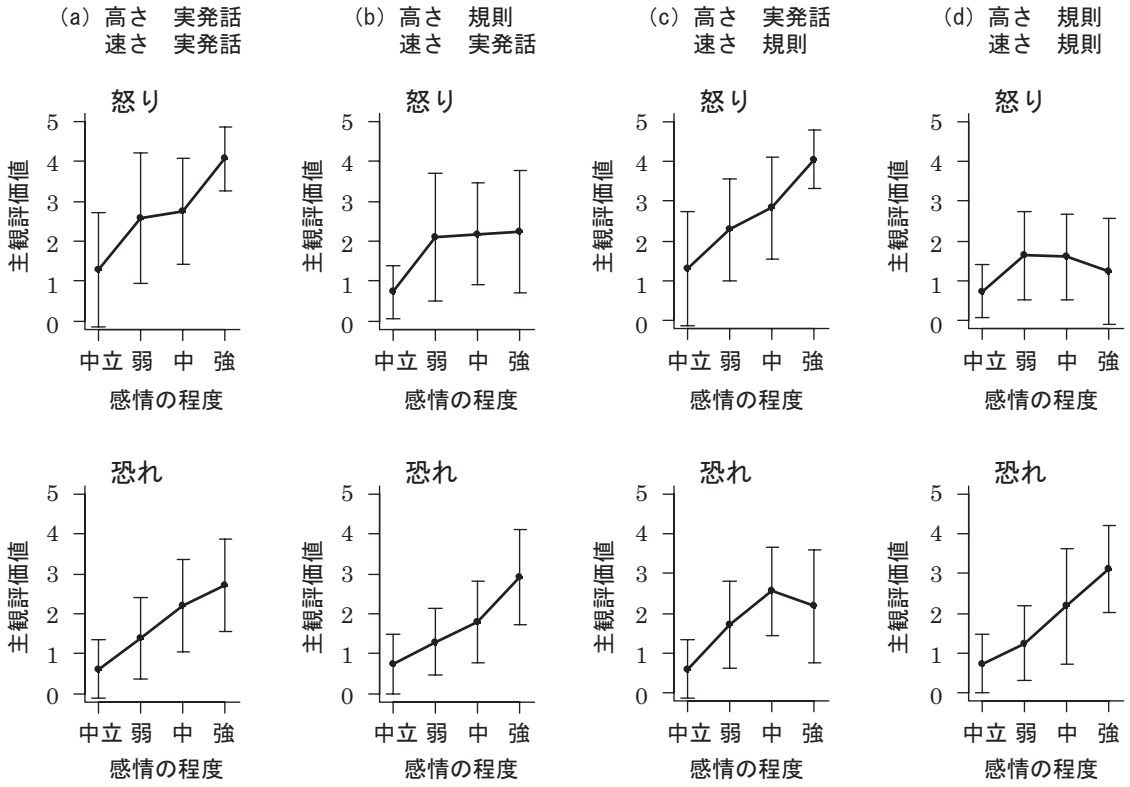

図 3. 合成音声に対する主観評価値の分布

5.3 実験結果および考察

聴取実験の結果から，合成音声のセットごとに合成音声の感情の程度に対する主観評価値の分布を求めた．図 3 に，以下の 4 つの場合の open データに対する結果を示す．

(a) 高さ，速さともに実発話を模擬したもの

(b) 高さのみ感情制御規則を適用したもの

(c) 速さのみ感情制御規則を適用したもの

(d) 高さ，速さとも感情制御規則を適用したもの

「怒り」では，感情制御規則に基づき F_0 パターンを生成した場合，感情が表現された音声の時間構造を用いても，中立と感情ありの区別がされるにとどまり，感情の程度は伝達されなかった．closed データに対する主観評価の結果も，open データに対する結果と同様の結果となっており，十分な F_0 パターンのパラメータ制御規則が得られていないものと考えられる．一方，速さを感情制御規則に基づき変化させた場合には，期待通りに感情が伝達された．

「恐れ」では，感情制御規則に基づき F_0 パターンを指定した場合にも，感情の程度が伝達された．さらに，感情制御規則に基づき時間構造を変化させた場合にも，感情が伝達された．このことから，「恐れ」について，有効な感情制御のための指針を得たといえる．

また，「怒り」「恐れ」ともに，F_0 パターンに関する効果が大きく，中立の時間構造であっても，感情が表現された F_0 パターンを利用することによって，感情が伝達された．一方，時間構造に関する影響は小さく，時間構造のみを変化させても，中立の F_0 パターンに対し最良近似を与えるパラメータを用いた場合，意図した感情は伝達されなかった．

6. おわりに

本稿では，感情の種類として「怒り」「恐れ」の 2 つを取り上げ，それらの感情が表現された発話が音声の基本周波数パターンに及ぼす影響を，F_0 パターン生成過程モデルに基づき分析し，感情制御規則を導出した．また，中立発話の文節の継続時間長と対象発話の継続時間長との継続時間長比を分析し，感情が表現された音声の発話速度に関する特徴について検討した．さらに，導出した感情制御規則を音声合成時に適用し，合成したそれぞれの音声が表現する感情の程度の伝達性を評価した．その結果，特に「恐れ」に関して有効な感情制御規則を得たことを確認した．また，「怒り」「恐れ」ともにその感情を表現するためには F_0 パターンが重要であり，発話速度の影響は小さいことを確認した．

今後は，本稿では取り扱わなかった韻律的特徴である音声のパワー変化や声質の特徴が感情の程度に及ぼす影響を検討する予定である．

参考文献

[1] 河津, 大野, “程度の異なる感情音声に対する韻律的特徴の分析,” 音講論（春）, pp. 233-234, 2007.

[2] 川波, 広瀬, “韻律構造に基づく対話調音声の発話速度の分析と規則化,” 信学技報, SP99-9, pp. 1-8, 1999.

[3] 藤崎, “音声の韻律的特徴における言語的・パラ言語的・非言語的情報の表出,” 信学技報, HC94-09, 1994.

[4] Julian, http://julius.sourceforge.jp/

[5] 河井他, “日本文章音声の合成のための韻律規則,” 音響誌, 50(6), pp. 433-442, 1994.

[6] The MBROLA Project, http://tcts.fpms.ac.be/synthesis/mbrola.html