感情音声合成方式の開発 －感性制御形 BMI の要素技術として－

武田 昌一[1], 株田 佳毅[2], 波床 政志[2], 鷲真紀子[3], 井内亮輔[4], 中川 秀夫[5], 桐生 昭吾[6], 西田 昌史[7], 山本 誠一[7]

要旨

「音声」は人間とロボットのコミュニケーションにおいて重要な役割を果たすようになってきた。また，音声は話者の感情の表出手段のひとつでもある。したがって，感情を付加した音声によるコミュニケーションは，より人にやさしいインタフェースの応用開発には不可欠である。このような感情音声形コミュニケーションシステムを実現するために，筆者らは，感情に関する脳情報が各種装置やロボットを制御するタイプの，音声を用いた感性制御形 BMI (Brain-Machine Interface) を確立することを本研究の目標としている。ここでわれわれは,「感情」は「感性」の一形態であると考えている。このような BMI の一要素として，本稿では，韻律変換および声質変換を用いて，種々の度合いの感情を表現することが可能な日本語音声規則合成システムを提案する。声質の特徴のうちで，高域周波数の強調の度合いを表す「スペクトル傾斜」と呼ばれるスペクトルの特徴量は，感情の種類と度合いに依存することを見出している。これまでの解析によれば,「怒り」,「喜び」および「泣きわめくタイプ（熱い）悲しみ」の場合は，感情の度合いが増大するほどスペクトル傾斜は増大する。それに対して,「意気消沈したささやくような（冷たい）悲しみ」では，感情の度合いが増大するほどスペクトル傾斜は減少する。そこで,「平常」のスペクトル傾斜をこれらの感情ごとに種々の度合いの値に変換する伝達関数を定式化し，音声合成システムにスペクトル傾斜変換規則として導入した。韻律変換規則もこれまでの知見に基づいて定めた。提案手法により合成した「熱い怒り」と「冷たい悲しみ」の音声サンプルを聴取したところ，合成音声は韻律と声質において，人間が発声した感情音声と類似の印象を持つことを確認した。更に，感情の度合いを変化させたときの感情の強さに対する印象の違いも明確に識別できた。

キーワード：ブレーン・マシン・インタフェース，音声合成，感情表現，感情の度合い，声質

1. はじめに

近年，脳機能計測装置や計測技術の進展に伴い，医療の現場のみでなく，様々な分野で脳研究が盛んになってきた。その中には，様々な視聴覚刺激に対して脳がどのように反応してどのような情報処理をしていくかを解明する研究から，最近では脳からの情報を指令として義手やコンピュータを制御するいわゆる BMI (Brain-Machine Interface) あるいは BCI (Brain-Computer Interface) の研究も盛んになり始めている[(1)－(3)]。

筆者らはこれまで感性情報解明研究の一環として，前者の研究を主として脳波計を用いて行ってきた。その中には，配色パターンの印象に関連する脳波成分の抽出[(4)]，ピアノ和音聴取時の脳の反応[(5)]，ビートトラッキング（音楽に合わせて手拍子を打つこと）時の脳の反応[(6)] などの研究がある。

脳波計を用いる研究では，以上のように視聴覚刺激を受動的に受容する活動が主対象であり，被検者の動作が伴うような場合は筋電雑音の混入が問題となり，測定が困難であった。上記のビートトラッキングを被検者に行わ

原稿受付 2012 年 11 月 29 日
本研究は近畿大学生物理工学部戦略的研究 No.10-IV-25, 2011 の助成を受けた。
[1] 近畿大学生物理工学部 システム生命科学科, 〒 649-6493 和歌山県紀の川市西三谷 930
[2] 近畿大学大学院生物理工学研究科 電子システム情報工学専攻, 〒 649-6493 和歌山県紀の川市西三谷 930
[3] 久留米信愛女学院短期大学 ビジネスキャリア学科, 〒 839-8508 福岡県久留米市御井町 2278-1
[4] 近畿大学生物理工学部 電子システム情報工学科, 〒 649-6493 和歌山県紀の川市西三谷 930
[5] 近畿大学生物理工学部 人間工学科, 〒 649-6493 和歌山県紀の川市西三谷 930
[6] 東京都市大学工学部 生体医工学科, 〒 158-8557 東京都 世田谷区 玉堤 1-28-1
[7] 同志社大学理工学部 情報システムデザイン学科, 〒 610-0321 京都府京田辺市多々羅都谷 1-3

せる場合も，手拍子を打つ代わりに指先だけでタッピングを行うという動作で代用せざるを得なかった。このように，脳波計は動作を伴う状況での脳活動の測定には不向きである。

「動き」を含めて，人間の知性，感性，運動など総合的に働かせる場面での脳の反応が計測できれば，芸術やスポーツなど，人類の高度な精神活動の根源を解明する道が開けると考えられる。筆者らは，人間の感性の問題を深く掘り下げ，本質に迫ることを目指している。

このように，更に深く踏み込んだ領域での研究を実現するために，脳波計に代わるものとして，現在では動きに対して雑音の影響を受けにくい光脳機能イメージング装置（脳血流の測定）を用いている。これまでに，音楽行為時の脳の情報処理過程 [7], [8] や，小倉百人一首かるた競技時の選手の脳の情報処理過程 [9] を解明する研究を進めている。

本研究では，感性の一形態である「感情」を対象として，これまでの筆者らが行ってきた「感情を込めて話すことができるロボット」を使って，実際に得られた脳情報データを用いてロボットを制御できることを実証することを最終目標とする。例えば，人間が怒ったときの脳情報データをロボットに入力したときに，ロボットが怒り顔になり，怒り音声を発声すれば，人間の感情が脳情報として取り出せたことを実証したことになる。

研究は，(1) 脳情報検出部，(2) ロボットの顔の感情表情生成および制御部，(3) 感情音声合成・出力部，および (4) 全体制御部（BMI 部），に分けて実施している。(1) については，可搬型光脳機能イメージング装置（“nIR HEG”）の BMI への応用可能性の評価と問題点の抽出を行う。現在その前段階として，据置型光脳機能イメージング装置（「光トポグラフィー装置」）を用いて，感情を頭の中でイメージしてそのときの声を発声したときの脳血流データを収集している。(2) については，ロボットが完成するに至っている。 (3) については，感情音声合成の基本枠組が完成し，「熱い怒り」と意気消沈したタイプの「冷たい悲しみ」表現は合成可能となっている。(4) については，今後 (1)〜(3) の技術を統合した上で構築する予定である。

本稿では，上記のうち (3) について報告する。

2. BMI システムの構想

筆者らが目指している感性制御形 BMI の最終形態を Fig.1 に示す。ヒト（左）が何らかの感情（例えば怒り）を抱いているときの脳情報を例えば可搬型光脳機能イメージング装置により抽出する。更にこの脳情報からヒトの感情を推定し，(1) 顔の表情を制御するパラメータと (2) 感情音声合成パラメータを抽出する。そして，(1) のパラメータに基づき顔の表情を制御する。例えば，ヒトが怒りの感情を抱いているときは，目玉を上につり上げる。また，(2) のパラメータに基づき感情表現を付加する発話文を作り出し，更に感情の種類と度合いを表すパラメータを決定し，音声合成部に送る。そして音声合成システムが感情音声を合成し，ロボットの口（スピーカ）から出力する。

以下の章では，この感性制御形 BMI システムのうち，感情音声合成部に焦点を絞って検討結果について述べる。

3. 感情音声合成研究の経緯

筆者らは，これまで，感情音声の合成を目的として「怒り」,「喜び」,「悲しみ」などを表現する自然音声について，感情の度合いを「平常」,「軽い」,「中程度」,「激しい」の４段階に分けて４人のアナウンサーや４人の声優が発声した音声の韻律の特徴 [10] および声質の特徴 [11][12] を調べてきた。

次に，これらの韻律の解析結果を基に，規則による感情音声の合成を試みたが，十分に感情を表現するには至らなかった。原因は韻律的特徴だけでは感情を表現するには不十分であるためと考えられる。韻律以外の情報にまで解析範囲を広げる必要があると考えた。そこで，その中で声質に着目して感情の種類・度合と声質の関係について検討してきた。

声質に影響を与える要因としては，音声，特に音源に含まれる雑音レベル，スペクトル形状 (傾きや山谷の起伏の大きさ）などが考えられる。筆者らは，これまで，音源情報として音声信号の予測残差に含まれる雑音量が感情の種類や度合いによって異なることを定量的に明らかにした [11]。

次に音声信号のスペクトル傾斜 [13][14] に着目し，感情の種類・度合いとの関係を調べた [12]。更に，これらスペクトル傾斜に関する知見を基に，「平常」音声から任意の感情・度合いの感情音声のスペクトル傾斜値に変換する方式を提案した [15]。

現在われわれは，これまでに得られた韻律的特徴および声質の特徴に関する知見に基づいて，平常音声から任意の感情の種類・度合いの音声を合成する方式を提案している。本稿では，その中で既に作成した「熱い怒り」と「意気消沈した表現の悲しみ」（以後「冷たい悲しみ」と呼ぶことにする）を表現する音声合成規則について述べる。

4. 感情音声合成方式の概要

平常音声の韻律および声質を変換して感情音声を合成する処理の全体ブロック図を Fig. 2 に示す。ここでの声質変換処理は，音源 (声帯波) と声道 (合成器のスペクトル包絡) における変形処理の組み合わせにより行う [15]。

本稿では，韻律変換については後の章で概要を述べるに留め，声質変換，その中でも特にスペクトル傾斜変換に焦点を当てて述べていく。

入力は平常音声から抽出した声帯波とする。声帯波は，簡便な処理としては平常音声を線形予測分析フィルタ (逆フィルタ) に通して予測残差波形を求め，更に積分処理を施すことにより求めることができる。

平常音声から抽出した声帯波 $v(t)$ に，感情音声の声帯波の形状に近づくような変形を施す。変形後の声帯波を $v_e(t)$，変形処理の伝達関数を $E(z)$ で表すこととする。

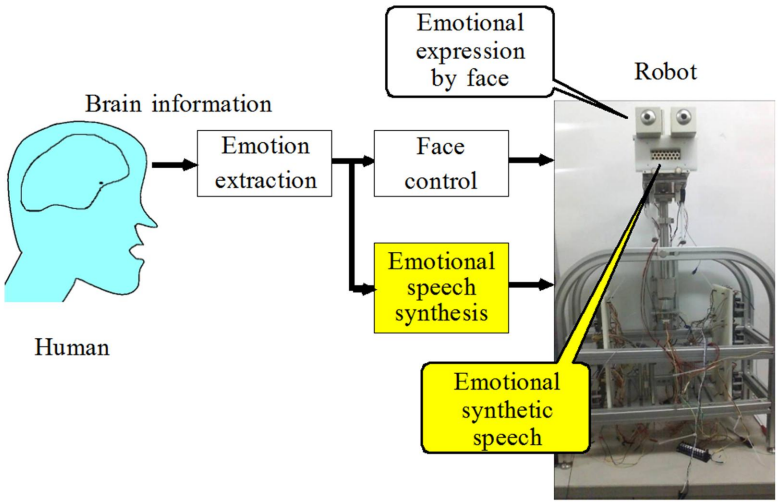

Fig. 1　Proposed BMI configuration.

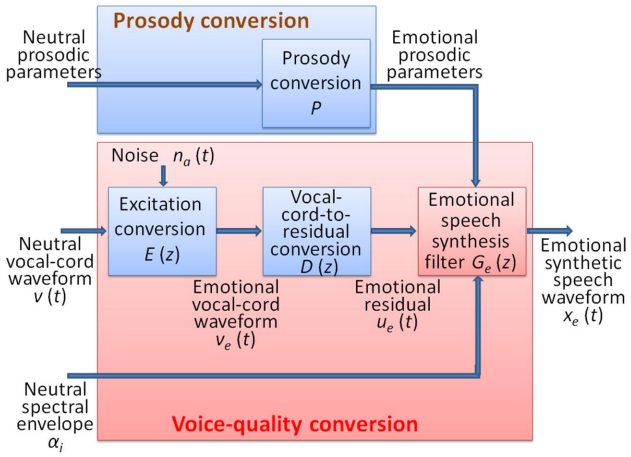

Fig. 2　Proposed rule-based emotional-speech synthesis system.

次に感情声帯波 $v_e(t)$ を感情残差波形に変換する。これは，入力音声残差波形から声帯波を求める処理と逆特性の処理である。この声帯波－残差変換処理を $D(z)$，感情残差波形を $u_e(t)$ で表す。声帯波を簡便な処理として，残差波形の積分処理により求めたとするならば，$D(z)$ は微分処理となる。

最後に，感情残差波形 $u_e(t)$ を感情合成フィルタ $G_e(z)$ に通過させることにより，感情合成音声 $x_e(t)$ を得る。ここで，感情合成フィルタ $G_e(z)$ については，次の節で詳しく述べる。

5. 感情音声合成フィルタの処理

声道における変形処理を加えて，感情音声に声質変換を施す方法の例を Fig. 3 に示す。この図は，Fig. 2 における感情合成フィルタ部 $G_e(z)$ の内容を詳細に示したものである。図中，$G(z)$ は通常の線形予測合成あるいは PARCOR 合成フィルタ，$S(z)$，$W(z)$ はそれぞれスペクトル傾斜変換部，スペクトル起伏変換部を表す。スペクトル傾斜変換部 $S(z)$ はスペクトル包絡全体の傾斜を制御し，スペクトル起伏変換部 $W(z)$ はホルマントの山谷の急峻さを制御することにより，声質を変換しようとするものである。

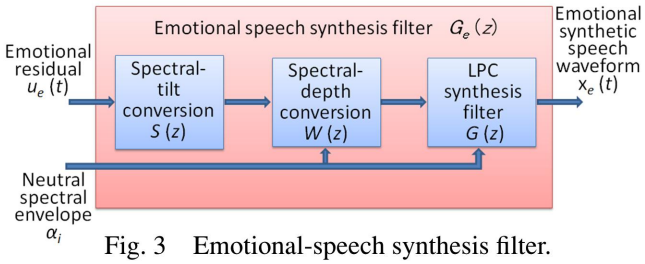

Fig. 3 Emotional-speech synthesis filter.

スペクトル起伏変換部 $W(z)$ については別稿に譲るものとし，本稿では，スペクトル傾斜変換部 $S(z)$ に限定して，その具体的な実現方法を述べる。

6. スペクトル傾斜変換フィルタ [15]

ここでは，Fig. 2，3 における各処理フィルタを具体的に決定する。一般論として，フィルタは自然発声の平常音声と感情音声を比較することにより得られる知見からの洞察により決定すべきである。知見が不十分な場合には，次善の策として推定によりフィルタの特性についての仮説を立て，仮説に基づくフィルタの特性を適用するものとする。

6.1. 音源変形 $E(z)$

「怒り」音声については，これまでの観察により，声帯波の形状が平常に比べると三角波に近づくような形で尖鋭化されていることがわかっている。ただし，平常であっても基本周波数が高いほど声帯波の形状が三角波に近づくので，三角波に近づくことが「怒り」特有の特性であるかどうかは，更に検討が必要である。

筆者らは，このような尖鋭化により，残差スペクトルの雑音レベルが増大することを見出し，N/S 比 [16] という指標を用いて各感情音声の残差波形に含まれる雑音レベルを定量化している [11]。残差スペクトルを詳細に観察すると，「怒り」の度合いが大きくなるにつれ，2 kHz 以上の周波数帯域で雑音レベルが顕著に増大していることがわかる。ただし，このような残差スペクトルにおける雑音特性と声帯波の尖鋭化特性との関係は未調査である。

そこで，声帯波にある特性の雑音が重畳されて残差スペクトルが変形すると考え，音源変形処理を次のように定式化する。雑音特性は一般には非線形であることが考えられるが，処理を簡単にするために，雑音としては乗法的雑音成分 $n_m(t)$ と加法的な雑音成分 $n_a(t)$ により構成されていると仮定する。そして，それぞれの雑音成分

の z 変換を $N_m(z)$，$N_a(z)$ で表すこととする。平常音声から抽出した声帯波 $v(t)$ の z 変換を $V(z)$，音源変形処理後の声帯波（感情声帯波） $v_e(t)$ の z 変換を $V_e(z)$ で表せば，音源変形処理は次式で表すことができる。

$$V_e(z) = V(z)N_m(z) + N_a(z) \tag{1}$$

ただし，音源変形の伝達関数 $E(z)$ は，$E(z) = V(z)N_m(z)$ で表されるものとする。

6.2. 声帯波ー残差変換 $D(z)$

声帯波を簡便な処理として，残差波形の積分処理により求めたとして，$D(z)$ を微分処理とすれば，もっとも単純な微分の近似計算は次式で表される 1 次差分である。

$$u_e(t) = v_e(t) - v_e(t - \Delta t) \tag{2}$$

ここで，Δt はサンプリング周期を表す。感情残差波形 $u_e(t)$ の z 変換を $U(z)$ で表せば，式 (2) の両辺の z 変換をとり，次式を得る。

$$U_e(z) = V_e(z) - z^{-1}V_e(z) = (1 - z^{-1})V_e(z) \tag{3}$$

更に， $D(z) = U_e(z)/V_e(z)$ であるから，これに式 (3) を代入して次式を得る。

$$D(z) = 1 - z^{-1} \tag{4}$$

6.3. スペクトル傾斜変換 $S(z)$

スペクトル包絡全体の傾斜を制御するフィルタ $S(z)$ は，一例として式 (5) で定義することができる。ここで，β はスペクトルの傾斜を制御するパラメータであり，β の値を $-1 \leq \beta \leq 1$ の間で変化させれば，スペクトル傾斜を -6 (dB/oct.) から 6 (dB/oct.) の間で連続的に変化させることができる。

$$S(z) = \frac{2 - (1 + \beta)z^{-1}}{2 - (1 - \beta)z^{-1}} \tag{5}$$

7. β – スペクトル傾斜特性の実測

式 (5) で示したスペクトル傾斜変換フィルタを実音声信号に適用して実測した結果より，以下の知見が得られている [(18)]。

ある感情音声信号のスペクトル傾斜の平常音声信号のスペクトル傾斜に対する差分値を ΔS_{tl} で表せば， ΔS_{tl} は単語，話者によらず常に

$$\Delta S_{tl} = 0.005\beta \tag{6}$$

で求めることができる。

過去の解析より，β 値は平常に比べ「怒り」の場合は大きく,「冷たい悲しみ」の場合は小さくなることがわかっている。たとえば $\beta = 0.5$ とすれば，平常音声を「怒り」の声質に変換でき，$\beta = -0.9$ とすれば,「冷たい悲しみ」の声質に変換できる [(15)]。

8. 感情音声合成規則

感情音声合成の第一段階の規則として，Fig. 2, 3 における韻律変換規則とスペクトル傾斜変換規則のみを採用し，音源変形などその他の規則は用いなかった。なお，韻律変換規則についてはここでは詳しく述べないが，Table 1, 2 に要約した特徴を反映する形で規則化する。

Table 1 Summary of the prosodic features of various emotions (speakers: announcers).

Emotion	Gender	Temporal structure	Fundamental frequency			Intensity
		Mean speech rate	Magnitude of phrase command A_p	Magnitude of accent command A_a	Maximum fundamental frequency F_{0max}	Maximum speech power
Anger	Male	Increase (anger) Reduction (fury)	Reduction	Increase	Increase	Increase
	Female	Increase	Increase			
Joy	Male	Reduction	No tendency	Increase	Increase	Increase
	Female	Increase				
Sadness	Male	Reduction	No tendency	No tendency	No tendency	Increase
	Female					Reduction

Table 2 Summary of the prosodic features of various emotions (speakers: radio actors/actresses).

Emotion	Gender	Temporal structure	Fundamental frequency			Intensity
		Mean speech rate	Magnitude of phrase command A_p	Magnitude of accent command A_a	Maximum fundamental frequency F_{0max}	Maximum speech power
Anger	Male	Reduction	Increase	Increase	Increase	(Increase)
	Female		No tendency			
Joy	Male	Reduction	Reduction	Increase	Increase	(Increase)
	Female					
Sadness	Male	Reduction	Reduction	No tendency	No tendency	(Increase)
	Female					

すべての規則は感情ごとに定め，規則を記述する諸量は「感情の度合い」の関数とした。例えば「熱い怒り」と「冷たい悲しみ」の場合，韻律パラメータである発話速度，基本周波数 F_0，音声パワー，および声質パラメータであるスペクトル傾斜を感情の度合い deg の関数として，これらの概略の変化パターンを示すと Fig. 4 のようになる。それぞれの規則は，この図で示したパターンのような変化をするように定式化する。

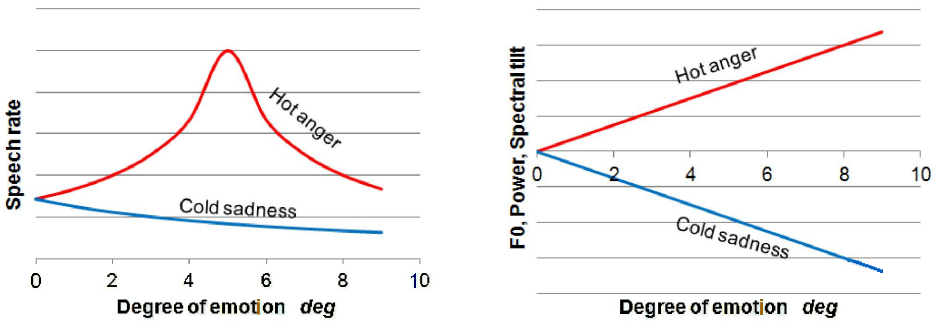

Fig. 4 Prosody- and spectral-tilt-conversion rules as functions of the degree of emotion.

8.1. 韻律変換規則

韻律変換規則は，現実の音声規則合成システムの入力として使われている韻律的特徴量を平常音声から感情音声に変換する規則として作成した。これらの韻律的特徴量は，モーラ持続時間 du，最低基本周波数 F_{0min}，アクセント指令の大きさ A_a，および音声パワー P である。

これまでに得られた解析結果の観察より，モーラ持続時間 du（発話速度），最低基本周波数 F_{0min}，アクセント指令の大きさ A_a，および音声パワー P は次の規則により変換される。これらの韻律的特徴量は話者依存性が高く，そのために厳密な定量化は現段階ではほとんど意味がないため，規則は Table 1, 2 で示した定性的な特徴のみを反映するように単純化した。

8.1.1. モーラ持続時間

「モーラ持続時間」は発話速度を制御するために用いる。

(1) **熱い怒り**

発話速度のモデルとして，男性アナウンサーが発話する音声に見られる，感情の度合いに応じて上昇後下降するいわゆる「への字パターン」を採用した。したがって,「熱い怒り」音声のモーラ持続時間は，発話速度が増大すれば発話時間長が短くなることから，発話速度とは逆に下降後上昇する「逆への時パターン」として次式で与えるものとする。

$$du_{ah} = du_n - cd_{ah}(5 - |deg - 5|) \tag{7}$$

ここで，du_n は対応する「平常」音声のモーラ持続時間を表す。この音声合成プログラムでは，感情の度合い deg は 0 から 9 までの整数値で与えるものとする。0 は最も弱い感情，すなわち「平常」，5 は中程度の強さの感情，9 は最も強い感情を表す。更に，cd_{ah} は変化量を決定する定数であり，実験的に $cd_{ah} = 10$ と定める。

(2) **冷たい悲しみ**

「冷たい悲しみ」音声のモーラ持続時間は,「悲しみ」の度合いと共に発話速度が減少する，すなわち時間長が増大するという解析結果をに基づき，次式で与えるものとする。

$$du_{sc} = du_n + cd_{sc}deg \tag{8}$$

ここで，cd_{sc} は変化量を決定する定数であり，実験的に $cd_{sc} = 10$ と定める。

8.1.2. 最低基本周波数

基本周波数に関する韻律的特徴パラメータとして,「最高基本周波数 F_{0max}」を用いる。しかしながら，規則合成における F_0 パターン生成には藤崎モデルを用いおり，その中で最低基本周波数値 F_{0min} を与えている。F_{0max} と F_{0min} の関係は複雑であり両パラメータ間の変換は容易ではない。そこで，F_0 値の制御には，暫定的に F_{0max} の代わりに F_{0min} を用いることとする。

(1) **熱い怒り**

「熱い怒り」音声の最低基本周波数 F_{0minah} は次式で与えるものとする。

$$F_{0minah} = F_{0minn}(1 + cf_{ah}deg) \tag{9}$$

ここで，F_{0minn} は対応する「平常」音声の最低基本周波数，cf_{ah} は変化量を決定する定数であり，実験的に $cf_{ah} = 0.05$ と定める。

(2) **冷たい悲しみ**

「冷たい悲しみ」音声の最低基本周波数 F_{0minsc} は次式で与えるものとする。

$$F_{0minsc} = F_{0minn}(1 - cf_{sc}deg) \tag{10}$$

ここで，cf_{sc} は変化量を決定する定数であり，実験的に $cf_{sc} = 0.05$ と定める。

8.1.3. アクセント指令の大きさ

基本周波数の他の韻律的特徴パラメータとして，藤崎モデルにより F_0 パターンを生成するのに不可欠な「アクセント指令の大きさ A_a」を用いる。

(1) **熱い怒り**

「熱い怒り」音声のアクセント指令の大きさ A_{aah} は次式で与えるものとする。

$$A_{aah} = A_{an}(1 + ca_{ah}deg) \tag{11}$$

ここで，A_{an} は対応する「平常」音声のアクセント指令の大きさ，ca_{ah} は変化量を決定する定数であり，実験的に $ca_{ah} = 0.05$ と定める。

(2) **冷たい悲しみ**

「冷たい悲しみ」音声のアクセント指令の大きさ A_{asc} は次式で与えるものとする。

$$A_{asc} = A_{an}(1 - ca_{sc}deg) \tag{12}$$

ここで，ca_{sc} は変化量を決定する定数であり，実験的に $ca_{sc} = 0.05$ と定める。

8.1.4. 音声パワー

(1) **熱い怒り**

「熱い怒り」音声のデシベルパワー dBP_{ah} は次式で与えるものとする。

$$dBP_{ah} = dbP_n + cp_{ah}deg \tag{13}$$

ここで，dbP_n は対応する「平常」音声のデシベルパワー，cp_{ah} は変化量を決定する定数であり，実験的に $cp_{ah} = 0.75$ と定める。

(2) **冷たい悲しみ**

「冷たい悲しみ」音声のデシベルパワー dBP_{sc} は次式で与えるものとする。

$$dBP_{sc} = dbP_n - cp_{sc}deg \tag{14}$$

ここで，cp_{sc} は変化量を決定する定数であり，実験的に $cp_{sc} = 0.75$ と定める。

8.2. スペクトル傾斜変換規則

β パラメータは 7 章で述べた方法により決定される。すなわち，次式のような感情の度合いを変数とする簡単な関数で決定される。

(1) **熱い怒り**

$$\beta = ctl_{ah}deg \tag{15}$$

ここで，ctl_{ah} は変化量を決定する定数であり，実験的に $ctl_{ah} = 0.1$ と定める。

(2) **冷たい悲しみ**

$$\beta = -ctl_{sc}deg \tag{16}$$

ここで，ctl_{sc} は変化量を決定する定数であり，実験的に $ctl_{sc} = 0.1$ と定める。

8.3. 規則合成実行例

Fig. 5 は，これまでに述べた規則により合成した種々の度合いの感情音声波形とその F_0 パターンの例を示している。合成音声は女声である。F_0 パターンは，筆頭著者が開発した，藤崎モデルに音素成分を加えた特殊なモデル [19] を用いて生成している。

図より,「熱い怒り」の度合いが増大するのに伴い，音声振幅（音声パワー）全体と基本周波数全体が増大し，アクセントレベルが強調されることがわかる。逆に,「冷たい悲しみ」の度合いが増大するのに伴い，音声パワー全体と基本周波数全体が減少し，アクセントレベルも減少している。

感情の種類と度合いにより，発話速度が変化している様子も見られる。すなわち,「熱い怒り」の度合いが増大すると発話速度は最初は増大し，感情の度合いが更に増大すると発話速度は逆に減少する，いわゆる「への字パターン」を呈している。他方「冷たい悲しみ」では，感情の度合いの増大と共に発話速度は減少している。

以上，提案手法により合成した音声サンプルを聴取したところ，合成音声は韻律と声質において，人間が発声した感情音声と類似の印象を持つことが確認できた。更に，感情の度合いを変化させたときの感情の強さに対する印象の違いも明確に識別できた。

9. おわりに

本稿では，感性制御形 BMI を構築することを目標として，その要素技術である感情音声合成方式について述べてきた。ここでわれわれは，韻律および声質変換を用いた，様々な感情の度合いを表現できる日本語音声合成システムを提案した。この中で，様々な度合いの「熱い怒り」と「冷たい悲しみ」を合成するための韻律変換および声質変換規則を提案した。

提案手法により合成した音声サンプルを聴取したところ，合成音声は韻律と声質において，人間が発声した感情音声と類似の印象を持つことが確認できた。更に，感情の度合いを変化させたときの感情の強さに対する印象の違いも明確に識別できた。

本研究を通じて得られた成果により,「熱い怒り」と「冷たい悲しみ」以外の種々のタイプと度合いの感情音声合成が実現すると期待される。

音声合成技術についての今後の課題は，種々のタイプと度合いの人間が発声した感情音声の特徴を解析することにより声質変換パラメータ値を決定し，これらの変換パラメータを用いて感情音声を合成し，更に心理評価実験などにより合成した感情音声の評価を行うことである。また，提案合成規則は単語音声のみから分析した結果に基づく限定的なものであるので，文音声への拡張も図って行く。

更に，これらの音声合成技術開発と平行して，ロボットを制御するための脳情報に関する知見を得て，実際に BMI を構築することも重要な今後の課題である。

謝辞

感情音声を発声して下さった劇団青年座の声優の方々に謝意を表する。

なお，本研究の一部は，近畿大学　生物理工学部　戦略的研究 No.06-I-3 (2007-2009), No.10-IV-25 (2011), および JSPS 科研費 21500209 の助成を受けて行ったものである。

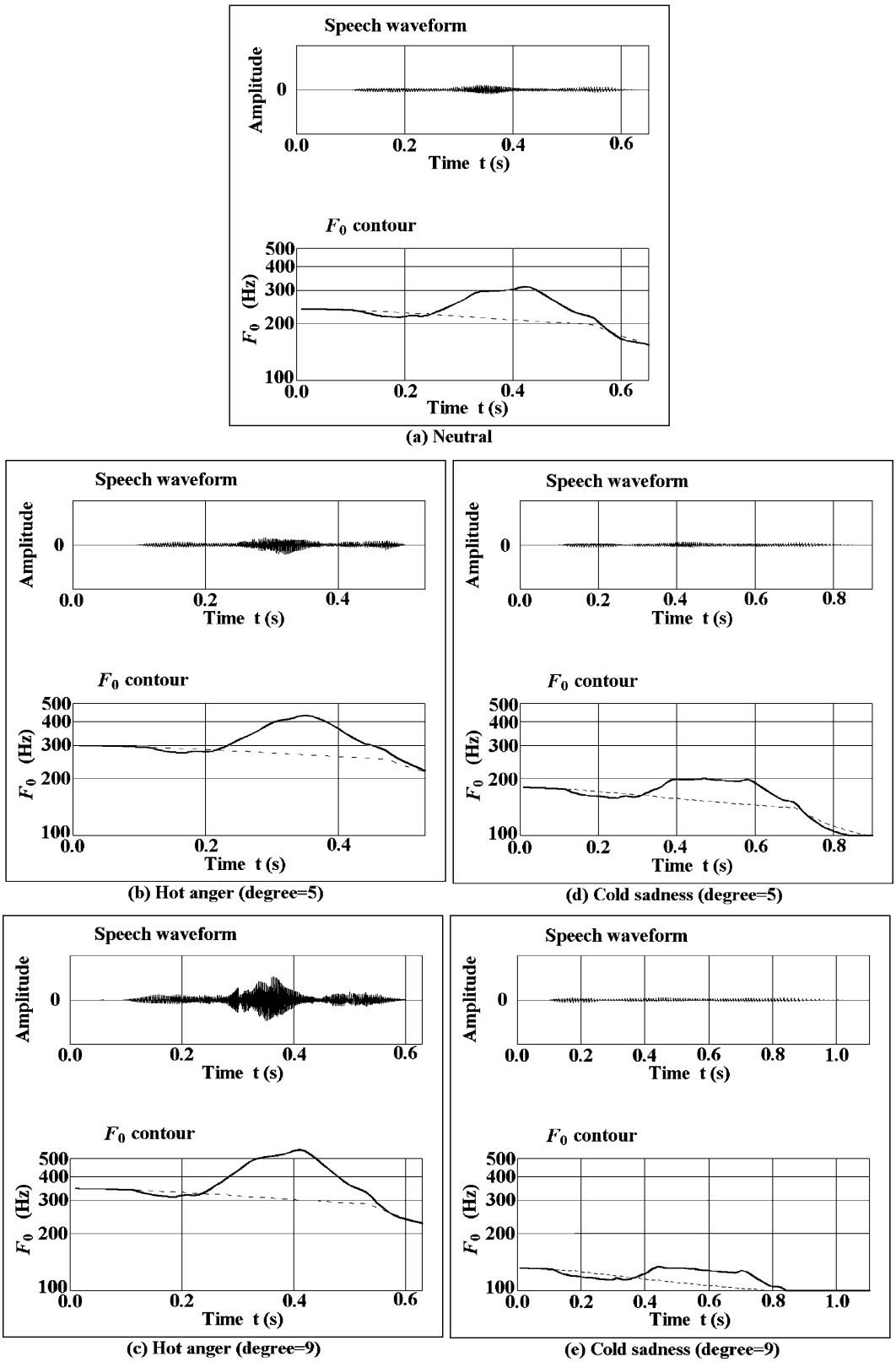

Fig. 5 Examples of synthetic speech waveforms (upper) and their F_0 contours (lower). Word: "Iyayo" meaning "No"; voice: female; emotion: "neutral", "hot anger", and "cold sadness"; degree: 5 (middle) and 9 (strongest)

参考文献

(1) Donoghue, J. P. (2008) Bridging the Brain to the World: A Perspective on Neural Interface Systems, Neuron 60, pp.511-521.

(2) Kim, S. *et al.*, J (2008) Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia, Neural Eng. vol. 5, pp.455-476.

(3) Truccolo, W. *et al.* (2008) Primary Motor Cortex Tuning to Intended Movement Kinematics in Humans with Tetraplegia, Journal of Neuroscience 28(5), pp.1163-1178.

(4) 武田昌一，山本佐代子，細川弥生，田中美枝子, 加藤修一　(2004)　配色パターン印象に関連する脳波成分の抽出, 感性工学研究論文集 Vol.5 No.1, pp.13-18.

(5) 山本佐代子，廣瀬百合子，佐川泰広，武田昌一　(2005)　ピアノ音色和音聴取時における脳波パワー変動, 感性工学研究論文集 Vol.6 No.1, pp.51-60.

(6) 山本佐代子，中西里果，佐川泰広，武田昌一　(2005)　ビートトラッキング時と音楽聴取時の脳波α波帯域パワー変動の比較検討, 感性工学研究論文集 Vol.5 No.3, pp.61-70.

(7) 廣瀬百合子，山本佐代子，藤井正子，大山 玄，井上正雄，武田昌一　(2007)　近赤外分光法による「音楽ドリル」実施時の前頭葉機能について　—高次脳機能障害リハビリテーションのためのソルフェージュ課題「音楽ドリル」の検証—, 日本音楽療法学会学術大会抄録.

(8) 廣瀬百合子，山本佐代子，藤井正子，井上正雄，武田昌一　(2007)　fNIRS による「音楽ドリル」実施時の前頭葉機能について　—高次脳機能障害リハビリテーションのためのソルフェージュ課題「音楽ドリル」の検証—, 第 11 回日本代替・相補・伝統医療連合会議 第 7 回日本統合医療学会 合同大会 in 松島 大会抄録 G4-2.

(9) 武田昌一, 長谷川 優, 平井祥之, 小杉 範, 津久井 勤, 山本誠一　(2009)　百人一首かるた選手の競技時の脳の情報処理に関する研究, 近畿大学生物理工学部紀要 No.24, pp.33-43.

(10) Hashizawa, Y., Takeda, S., Muhd Dzulkhiflee Hamzah., and Ohyama, G. (2004) On the Differences in Prosodic Features of Emotional Expressions in Japanese Speech according to the Degree of the Emotion, Proc. 2nd Int. Conf. Speech Prosody, Nara, Japan, pp.655-658.

(11) Takeda, S., Yasuda, Y., Isobe, R., Kiryu, S., and Tsuru, M. (2008) Analysis of Voice-Quality Features of Speech that Expresses “Anger”, “Joy”, and “Sadness” Uttered by Radio Actors and Actresses, Proc. Interspeech 2008, Brisbane, Australia, pp.2114-2117.

(12) Takeda, S., Ueno, Y., Nakasako, N., Nakagawa, N., Tsuru, M., Isobe, R., and Kiryu S. (2010) Spectral-Tilt Features of Emotional Speech -Research on Emotional-Speech Synthesis Based on Voice-Quality Conversion-, Proc. KEER2010, Paris, France, pp.1081-1090.

(13) Kasuya, H., Yoshizawa, M., and Maekawa, K. (2000) Roles of Voice Source Dynamics as a Converter of Paralinguistic Features, Proc. ICSLP2000, Paper #1283, Beijing, China.

(14) Liscombe, J., Venditti, J., and Hirschberg, J. (2003) Classifying Subjective Ratings of Emotional Speech Using Acoustic Features, Proc. Eurospeech 2003, Geneva, pp.725-728.

(15) 株田佳毅, 井上智広, 田口裕亮, 上垣内智美, 武田昌一 (2010) 種々の度合いの感情音声合成を実現するためのスペクトル傾斜制御方式の提案 —声質変換型感情音声合成の研究—, 日本音響学会 2010 年秋季研究発表会講演論文集 (CD-ROM) 3-P-19, pp.373-374.

(16) Muta, H., Baer, T., Wagatsuma, K., and Muraoka, T. (1989) A pitch-synchronous analysis of hoarseness in running speech, J. Acoust. Soc. Am., 84(4), pp.1292–1301.

(17) 靄真紀子，武田昌一 (2008) 声優が発声する感情音声の韻律的特徴と聴覚的印象の差異, 日本音響学会 2008 年春季研究発表会講演論文集 (CD-ROM) 3-Q-32, pp.445-446.

(18) Takeda, S., Kabuta, Y., Inoue, T., and Hatoko, M. (2012) Proposal of a Japanese-Speech-Synthesis Method with Dimensional Representation of Emotions Based on Prosody as well as Voice-Quality Conversion, Proc. KEER2012, Penghu, Taiwan, pp.462-470.

(19) Takeda, S. (1990) A Model for Generating Fundamental Frequency Contours Considering Phoneme Fluctuation and Rules for Speech Synthesis, IEICE Trans. Fundamentals, J73-A, pp.379-386.

英文抄録

Development of an Emotional Speech Synthesis System – As an Element of Technology for *kansei*-Based BMI –

Shoichi Takeda[1], Yoshiki Kabuta[2], Masashi Hatoko[2], Makiko Tsuru[3], Ryosuke Iuchi[4], Hideo Nakagawa[5], Shogo Kiryu[6], Masashi Nishida[7] and Seiichi Yamamoto[7]

"Speech" is becoming more and more important in communication between a human and a robot, and, moreover, speech is also a carrier of the speaker's emotion. Speech communication with emotions is therefore indispensable to develope more human-friendly applications. To achieve such an emotion-based speech communication system, we set the goal of our study to establish a *kansei*-based brain-machine interface (BMI) using speech, through which the brain information on emotions controls devices, robots, etc. Here, we consider "emotion" as one form of "*kansei*" or affect. As an element of such a BMI, this paper proposes a rule-based Japanese speech synthesis system that is capable of expressing variable degrees of emotions based on prosody as well as voice-quality conversion. Among voice-quality features, we find that a spectral feature called "the spectral tilt" that expresses the degree of empasis of the higher-frequency band depends on the type and degree of emotion. From our previous analyses, we found that the spectral-tilt quantities increased as the degrees of "anger", "joy", and "crying-type (hot) sadness" increased. On the other hand, the spectral-tilt quantities were found to decrease as the degree of "dispirited-and-whispering-type (cold) sadness" increased. We formulate a transfer function that converts spectral-tilt quantities of "neutral" speech to those of emotional speech in various degrees, and introduce this spectral-tilt conversion rule into our speech-synthesis system. The prosody-conversion rules are also determined based on our previous findings. Informal listening to "hot anger" and "cold sadness" synthetic-speech samples converted by the proposed method gives us impressions of those similar to natural emotional speech and the differences depending on the degrees of emotions are recognizable.

Key words : BMI, speech synthesis, emotional expression, degree of emotion, voice quality.

1. Department of Computational Systems Biology, Kinki University, Wakayama 649-6493, Japan
2. Program in Electronic Systems and Information Engineering, Graduate School of Biology-Oriented Science and Technology, Kinki University, Wakayama 649-6493, Japan
3. Department of Business Career, Kurume Shin-Ai Women's College, Fukuoka 839-8508, Japan
4. Department of Electronic Systems and Information Engineering, Kinki University, Wakayama 649-6493, Japan
5. Department of Biomechanical and Human Factors Engineering, Kinki University, Wakayama 649-6493, Japan
6. Biomedical Engineering Department, Tokyo City University, Tokyo 158-8557, Japan
7. Department of Information Systems Design, Doshisha University, Kyoto 610-0321, Japan