様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　　９　日現在

機関番号：　１７２０１
研究種目：　挑戦的萌芽研究
研究期間：　2014～2016
課題番号：　２６５４００６８

研究課題名（和文）感情的プロソディを変更した音声のフィードバックによる気分誘導方法の研究

研究課題名（英文）A method of mood induction according to feedback of voice sound modified emotional prosody

研究代表者

大島　千佳（Oshima, Chika）

佐賀大学・工学（系）研究科（研究院）・客員研究員

研究者番号：１０３９５１４７

交付決定額（研究期間全体）：（直接経費）　　2,900,000円

研究成果の概要（和文）：本研究は，生理的な変化が起きていると誤認識することで，感情が誘導されるという考え方をもとに，話者の音声をリアルタイムに変更して本人にフィードバックするシステムの開発を目指した．まず，どのような音響情報により，聞き手が感じ取る感情が操作されるかを示した先行研究をサーベイした．利用された音響情報は数十種類にのぼった．しかし，本研究のようにリアルタイムに変更する場合には，扱える音響情報は限られる．そこで音高と音量を24のアルゴリズムで変換する方法を提案した．実験の結果，「平静な」「のんびりした」「ゆったりした」感情に誘導するには，周波数ごとに音量を変換する方法が有用であることがわかった．

研究成果の概要（英文）：Our research aimed to construct a system that helps people to improve their mood naturally using a theory that emotions can be caused by the false recognition of physiological reactions. The system needs to modify a person's voice in real time. The first, we surveyed what kind of acoustic features are modified for emotion recognition in speech from related researches. There were dozens of kinds of acoustic features used for these researches. However, most of the acoustic features are not adequate to modify the voice in real time. Therefore, we developed the system that modifies pitches and volumes of real-time speech. We described 24 algorithms for converting sounds. Then, we conducted an experiment in which subjects read an essay while listening to their individual utterance converted by the system in real time. The results for the items calm, lazy, and comfortable show that the sounds converted like an equalizer were more agreeable than those converted the whole pitch.

研究分野：　ヒューマンコンピュータインタラクション

キーワード：　感情の誤認識　リアルタイム音声変換　音量　音高　認知症

１．研究開始当初の背景

認知症には，認知機能の低下という中核症状と，それに伴って発生する行動・心理症状（BPSD）がある．一般的に介護者の負担となるのは，奇声や徘徊，暴力や妄想などのBPSD である．BPSD を緩和する薬物以外の療法として，心理療法が期待されている．

研究代表者は，科学研究費補助金（基盤研究Ｂ)，及び学振特別研究員(RPD)奨励費により，認知症患者の BPSD を緩和する音楽システムの研究に取り組んできた．その中で，患者が BPSD により発する声（音高）に応じた音楽を提示する“MusiCuddle[1]”を開発した．これに“ボコーダ”というリアルタイム音声変換機能を追加し，被験者の発声を，事前に準備した音高に変換してフィードバックした．その結果，音楽の調子（長調／短調）によって，被験者の気分が，陽気な／悲しい，にそれぞれ誘導された[2]．ここから，認知症患者の声に含まれる「感情的プロソディ（感情によって変化する韻律）」をリアルタイムに変更してフィードバックすることで，患者を適切な気分へ誘導し，BPSD が緩和できるとの着想に至った．

これまで，感情と，生理的な反応との関係が議論されてきた[3][4][5]が，我々の実験の結果[2]は，生理的な反応の誤認識により，感情が引き起こされることを示唆している．「悲しいから泣く，楽しいから笑う」という，感情から生理的な変化が引き起こされると主張したキャノン＝バード説[3]と，「泣くから悲しい，笑うから楽しい」という，生理的な変化から感情が引き起こされると主張したジェームズ＝ランゲ説[4]がある．一方で，シャクター・シンガーの情動二要因説では，感情が引き起こされるには生理的な変化とその原因を認知することの両方であるとする[5]．我々は，この情動二要因説を拡張し，実際には生理的な変化は起きていないが，起きていると誤認識することで，感情が誘導されるという説を提唱している[6]．自分の発声が実際とは異なり，明るく，弾んで，陽気であると誤認識することで，幸せな感情が引き起こされるというものである．

２．研究の目的

話者の発話に含まれるプロソディをリアルタイムに変換するシステムを構築し，目的とする感情に誘導できるようにする．

３．研究の方法

(1)ボコーダを接続した MusiCuddle から提示される音は，音楽フレーズを伴った機械的な声になってしまい，日常で使うには不自然である．そこで我々は，リアルタイムに発話の音量や音高を変換し，話者にフィードバックできるシステムを構築する．

(2)システムを使って音量・音高を変換するアルゴリズムを提案し，各アルゴリズムにより変換された音声と感情との関係について分析する．

４．研究成果

(1)ハードウェア構成

図 1 は，発話の音高／音量を変換するシステムのハードウェア構成を示す．実験では，スタンドマイクとヘッドフォンを使用する．しかし，指向性マイクやスピーカにより，話者に負担をかけない方法も可能である．

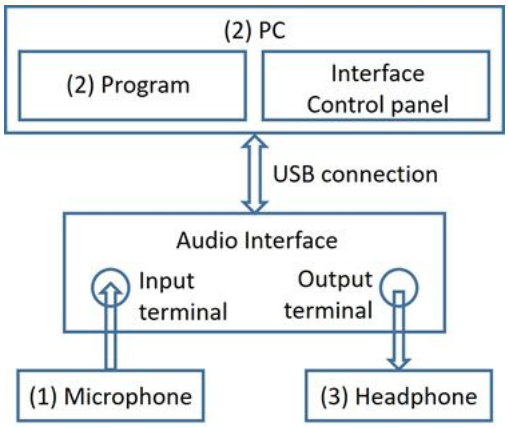

図 1:システムのハードウェア構成

(2)音高・音量変換アルゴリズム

図 2 は，変換の過程を示す．まず，オーディオデータから音高の値（InP(t)）と音量の値（InV(t)）を抽出する．システムはアルゴリズムにより， InP(t)または InV(t)のいずれかの値を使用する． InP(t)または InV(t)を使って，発話の音高または音量を変換するための変数が計算される．システムは，アルゴリズムの種類により，計算された変数(Var_Xx(t): Var_Pt(t)， Var_Pc(t)，Var_Vt(t)，Var_Vc(t))のいずれかを使って，音高または音量を変換する．音高全体を変換する方法，音量全体を変換する方法，そして周波数ごとに音量を変換する方法の 3 種類の方法がある．

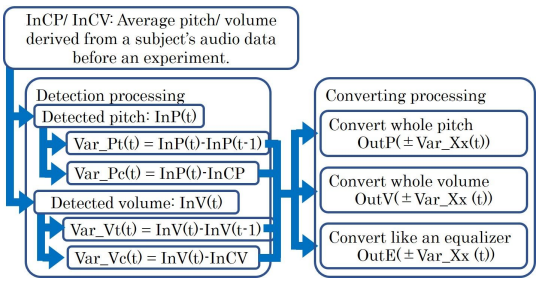

図 2:音声の変換過程

発話の音高と音量は，検出処理において，64 ms ごとに検出される．本稿では，64 ms の時間を，“t”という単位で示す．t における音高は “InP(t) (Hz)”と示し，音量は“InV(t) (dB)”と示す．それゆえに，InP(t)または InV(t) は 64 ms ごとに検出される．

音高または音量を変換するために， 以下の式で計算される，Var_Pt(t)，Var_Pc(t)，Var_Vt(t)，または Var_Vc(t)のいずれかの変数を使用する．

Var_Pt(t) = InP(t) - InP(t-1),eq. (1)

Var_Pt(t)は，現在(t)の音高と 1 つ前の時間単位(t-1)での音高との差異を示す．

Var_Pc(t) = InP(t) - InCP,eq. (2)
Var_Pc(t)は，現在(t)の音高と各実験協力者の実験前の発話データから抽出した音高の平均値(InCP)との差異を示す．
Var_Vt(t)= InV(t) - InV(t-1),eq. (3)
Var_Vt(t)は，現在(t)の音量と１つ前の時間単位(t-1)での音量との差異を示す．
Var_Vc(t) = InV(t) - InCV,eq. (4)
Var_Vc(t)は，現在(t)の音量と各実験協力者の実験前の発話データから抽出した音量の平均値(InCV)との差異を示す．
システムは，入力された音声をフーリエ変換（FFT）したあと，音高（周波数）と音量を変換する．その後，逆フーリエ変換（IFFT）により出力用音声データを作成する．話者が，変換された自分の声を聴くまでの遅延時間は，実測で 75ms であった．

全体の音高の変換

全体の音高は変換率 OutP(t))により変換される．たとえば，“OutP(t)=1”ならば音高は変換されず，“OutP(t)=2” ならば 1 オクターブ上の音高に，“OutP(t)=0.5”ならば 1 オクターブ下の音高に，それぞれ変換される．
OutP(t)は次のアルゴリズムにより計算される (PtP+, PtP-, PcP+, PcP-, VtP+, VtP-, and VcP+, VcP-).
PtP+: OutP(t) = 1+ Var_Pt (t) /500, eq. (5)
PcP+: OutP(t) = 1+ Var_Pc (t) /500, eq. (6)
VtP+: OutP(t) = 1+ Var_Vt (t) /250, eq. (7)
VcP+: OutP(t) = 1+ Var_Vc (t) /100, eq. (8)
アルゴリズム，PtP+, PcP+, VtP+, VcP+の変換率 OutP(t)は，Var_Pt (t)，Var_Pc (t)，Var_Vt (t)，または Var_Vc (t)が正（負）のとき，正（負）になる．
PtP-: OutP(t) = 1- Var_Pt (t) /500, eq. (9)
PcP-: OutP(t) = 1- Var_Pc (t) /500, eq. (10)
VtP-: OutP(t) = 1- Var_Vt (t) /250, eq. (11)
VcP-: OutP(t) = 1- Var_Vc (t) /100, eq. (12)
アルゴリズム，PtP-, PcP-, VtP-, VcP-の変換率 OutP(t)は，Var_Pt (t)，Var_Pc (t)，Var_Vt (t)，または Var_Vc (t)が正（負）のとき，負（正）になる．
$0.93 \leqq OutP(t) \leqq 1.15$, eq. (13)
eq.(13)に示す OutP(t)の範囲は，我々の予備実験の結果をもとに決定された．この範囲の変換率であれば，変換された音声を聞いてもほとんど不快ではないと期待する．

全体の音量の変換

全体の音量は変換率 OutV(t))により変換される．たとえば，“OutP(t)=1”ならば，音量は変換されず，“OutP(t)=2” ならば元の音量の 2 倍に，“OutV(t)=0.5”ならば元の音量の半分に，それぞれ変換される．OutV(t)は次のアルゴリズムにより計算される (PtV+, PtV-, PcV+, PcV-, VtV+, VtV-, and VcV+, VcV-).
PtV+: OutV(t) = 20 × log10{1+ Var_Pt (t)/100}, eq. (14)
PcV+: OutV(t) = 20 × log10{1+ Var_Pc(t)/100}, eq. (15)
VtV+: OutV(t) = 10^{Var _Vt (t)/20}, eq. (16)
VcV+: OutV(t) = 10^{Var _Vc (t)/20}, eq. (17)
アルゴリズム，PtV+, PcV+, VtV+, VcV+の変換率 OutV(t)は，Var_Pt (t)，Var_Pc (t)，Var_Vt (t)，または Var_Vc (t)が正（負）のとき，正（負）になる．
PtV-: OutV(t) = 20 × log10{1- Var _Pt (t)/100}, eq. (18)
PcV-: OutV(t) = 20 × log10{1- Var _Pc(t)/100}, eq. (19)
VtV-: OutV(t) = 10^{- Var _Vt (t)/20}, eq. (20)
VcV-: OutV(t) = 10^{- Var _Vc (t)/20}, eq. (21)
アルゴリズム，PtV+, PcV+, VtV+, VcV+の変換率 OutV(t)は，Var_Pt (t)，Var_Pc (t)，Var_Vt (t)，または Var_Vc (t)が正（負）のとき，負（正）になる．
音量の上限は，システムが出力できる最大値である．

周波数ごとの音量の変換

周波数ごとに音量を変換する方法は，イコライザと同様と考えてよい．周波数帯(0Hz-12000Hz）を，512 のパートに分ける．システムは各パートの音量を変換する．変換パラメータの“OutE (t)”は，1-512 の幅で決定される．

$$Ev(n) = 0.7 \times \left[\left\{\frac{1}{\left(\frac{OutE-n}{10}\right)^2+1}\right\}\right] + 0.3, \quad n=\{1,2,3\cdots 512\},$$ eq. (22)

Ev(n)は各周波数の音量を示し，変換パラメータ“OutE(t)”により，変換される．OutE(t)は，次のいずれかのアルゴリズムにより 0～100%の範囲で計算される（PtE+, PtE-, PcE+, PcE-, VtE+, VtE-, VcE+, VcE-）図 3 は，周波数ごとの音量の例である．

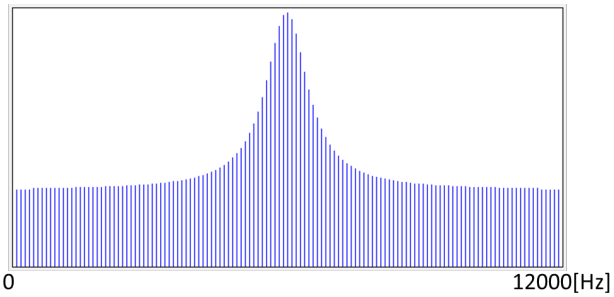

図 3:周波数ごとの音量変換の例

PtE+: OutE(t) = Fnc[OutE(t-1)+ Var_Pt (t)], eq. (23)
PcE+: OutE(t) = Fnc [255+ Var _Pc(t)], eq. (24)
VtE+: OutE(t) = Fnc [OutE(t-1)+ 8.2×Var _Vt (t)], eq. (25)
VcE+: OutE(t) = Fnc [255+ 8.2×Var _Vc (t) , eq. (26)
Fnc [ ]は，小数点以下を切り捨て，0 以下を 1 に，513 以上を 512 にする関数である．アルゴリズム，PtE+，PcE+，VtE+，VcE+の変換率 OutE(t)は， Var_Pt (t)， Var_Pc (t)， Var_Vt (t)，または Var_Vc (t)が正（負）のとき，正（負）になる．
PtE-: OutE(t) = Fnc [OutE(t-1)- Var _Pt (t)], eq. (27)
PcE-: OutE(t) = Fnc [255- Var _Pc(t)], eq. (28)
VtE-: OutE(t) = Fnc [OutE(t-1)- 8.2×Var_Vt (t)], eq. (29)
VcE-: OutE(t) = Fnc [255- 8.2×Var _Vc (t)], eq. (30)
アルゴリズム，PtE-，PcE-，VtE-，VcE-の変換率 OutE(t)は， Var_Pt (t)， Var_Pc (t)， Var_Vt (t)，または Var_Vc (t)が正（負）のとき，負（正）になる．

(3)実験

実験協力者は 12 名の男性の大学生であった．各協力者は，用意されたエッセイを 10 回読んだ．1 回目はマイクのテストや協力者の練習のためである．2 回目は変換しない条件であり，協力者は変換されない状態の自分の音声を，ヘッドフォンを通して聞いた．3～10 回目は，24 の変換アルゴリズムのうちの 8 種類の方法により，発話の音高または音量をリアルタイムに変換した．

音読に使用するエッセイは，北大路魯山人のエッセイ「美味い豆腐の話[7]」であり，5 分以内には読み終わる．本実験では，発話の音声の変換による，感情の変化を調べたいため，エッセイの内容で大きく感情が変化することは避けたい．エッセイが引き金となる感情の大きな変化は起きにくいと考えて，この作品を選択した．

協力者は，エッセイを 1 回音読するごとに，質問紙に書かれた 24 の形容詞について，自分の感情がどの程度あてはまるか，1（全くあてはまらない）～4（とてもあてはまる）の 4 段階で示す．24 の形容詞は，8 つの尺度とその下位が 10 項目から成る，多面的感情状態尺度[8]から選択した．尺度「抑鬱・不安」から，ふさぎこんだ，悲観した，くよくよした，不安な，の 4 項目，尺度「倦怠」から，無気力な，だるい，疲れた，不機嫌な，の 4 項目，尺度「活動的快」から，陽気な，気持ちの良い，はつらつとした，気力に満ちた，の 4 項目，尺度「非活動的快」から，ゆったりした，平静な，やわらいだ，のんびりした，の 4 項目，尺度「敵意」から，敵意のある，憎らしい，挑戦的な，怒った，の 4 項目，そして尺度「親和」から，恋しい，うっとりした，すてきな，情け深い，の 4 項目である．

(4)分析

まず，各形容詞において「24 の条件（アルゴリズム）での評価の中央値は等しい」という帰無仮説で，多重比較（Steel-Dwass 法）で調べた．しかし，各条件の協力者が 4 名ずつで 10 名に達していないため，有意差が出にくく，検定には適していない．そこで次に，4 つの条件（変換なし，全体の音高を変換，全体の音量を変換，周波数ごとに音量を変換）に結果をまとめて，Steel-Dwass 法で検定した．さらに，24 の条件をクラスタ分析（ward 法）した．

(5)結果

24 の条件（アルゴリズム）では，帰無仮説が棄却されなかった．しかし，図 4-11 に示すように，4 つの条件にまとめて検定すると，8 つの形容詞（悲観的な，陽気な，気分のよい，はつらつとした，怒った，平静な，のんびりした，ゆったりした）について有意な差異がみられた．図中の**は $p<0.01$ を示し，*は $p<0.05$ を示す．

図 4-8 の結果から，発話を変換しない条件の結果と，3 つのいずれかの方法で変換した条件の結果に差異があることがわかる．声が変換されない場合には，システムにより変換された声を聞きながら音読するよりも，悲観的ではなく，陽気で，気分がよく，はつらつとし，怒りはなかった．

一方で，図 9-11 の結果から，音高を変換した条件と周波数ごとに音量を変換した条件の間で，平静な，のんびりした，ゆったりした（3 つとも，尺度「非活動的快」に属する）で差異があった．

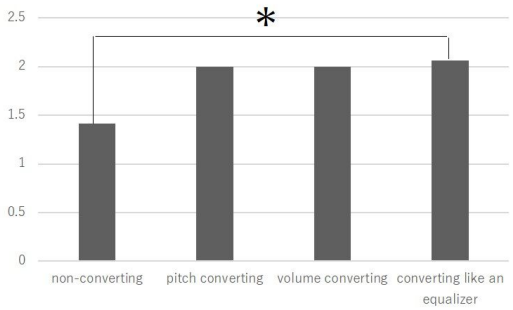

図 4　「悲観した」の評価の平均値

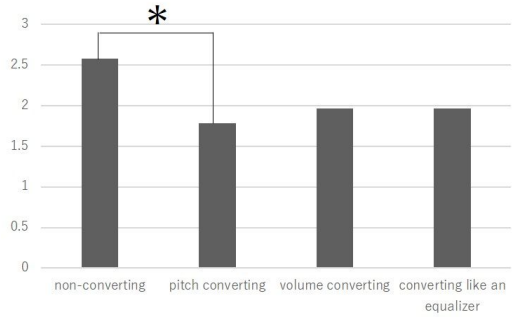

図 5　「陽気な」の評価の平均値

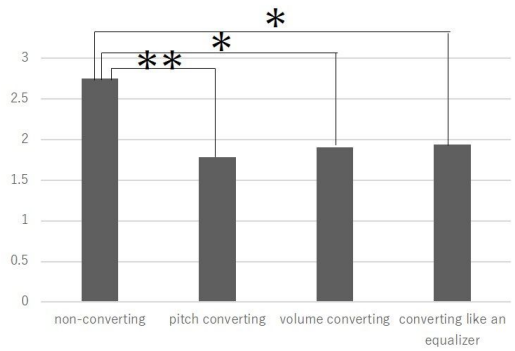

図 6　「気持ちのよい」の評価の平均値

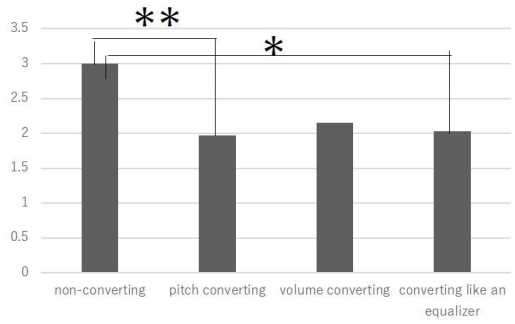

図 7　「はつらつとした」の評価の平均値

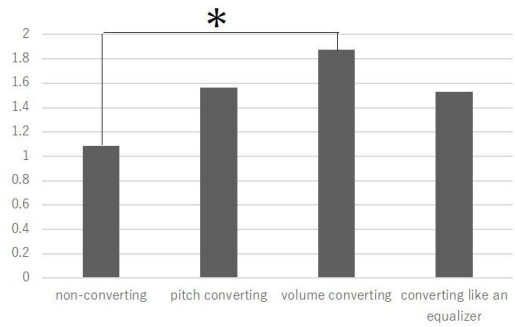

図 8　「怒った」の評価の平均値

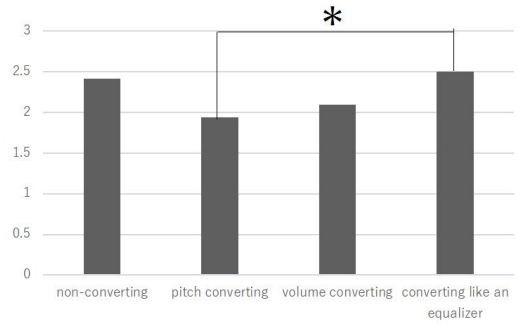

図 9　「平静な」の評価の平均値

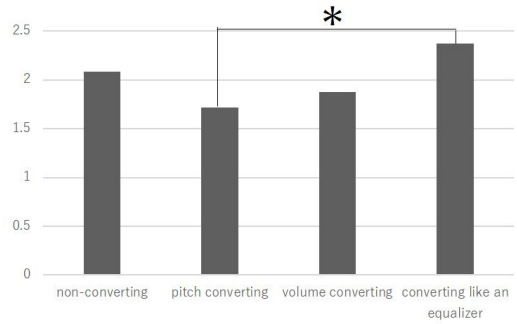

図 10　「のんびりした」の評価の平均値

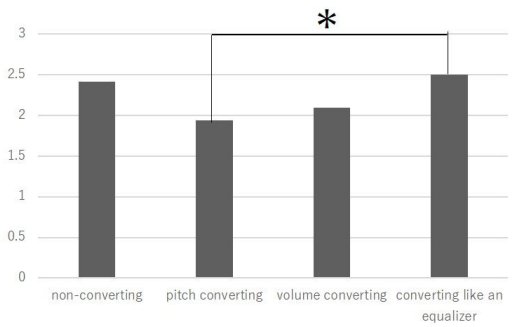

図 11　「ゆったりした」の評価の平均値

クラスタ分析により，距離 2.5 で，24 種類のアルゴリズムと変換なしの条件は，9 つのグループに分けることができた．同じグループ内は，たとえば“VcV-，”“PtE-，”“VcE-，”“PcE-“のように，正負のマークが同じになりやすいことがわかった．

距離 5 の分割では，“VtE-，”“PtE+，”“VcE+，”“PcE+，”“VtE+”や，“VtP-，”“VcP-，”“PtP-”のように，変換するもの（全体の音高，全体の音量，周波数ごとの音量）が同じもの同士が，同じグループになりやすいことがわかった．

(6)考察

実験協力者が音読後に示した感情の結果から，まず，変換された音声を聞いた後はあまり快くないことがわかった（図 4，6，8）．ここから，システムにより声を変換されたことが，話者に明らかにわかりすぎるといえる．我々は，話者が自分の声を変換されたことに気が付かなくても，自分に生理的反応が起き

ていると誤認識して，感情が起きると考える．よって，話者が変換されたことに気づきにくい手法[9]への改良が必要である．

また，声を変換しない方が，陽気で，はつらつとする(図 5，7)ことから，変換すると音がくすむように聞こえると考えられる．

一方で，平静な，のんびりした，ゆったりした感情に誘導するには，周波数ごとに音量を変換する方法の方が，全体の音高を変換する方法よりも適切であるといえる（図 9-11）．

クラスタ分析の結果からは，変換に使用する変数（時間単位 1 つ前の音高／音量や，実験前の音高／音量の平均値との差異）の正負と，変換率の正負を一致させるか否かによって，協力者の感情が異なる傾向になることがわかった．

さらにクラスタ分析の結果から，特に，周波数ごとに音量を変換するアルゴリズム同士で，協力者の感情が似た傾向になることがわかった．

＜引用文献＞

① C. Oshima, N. Itou, K. Nishimoto, K. Yasuda, N. Hosoi, H. Yamashita, K. Nakayama, and E. Horikawa: A Music Therapy System for Patients with Dementia Who Repeat Stereotypical Utterances, Journal of Information Processing, 21, 2, 283/294 (2013)

② C.Oshima, K. Nakayama, N. Itou, K. Nishimoto, K. Yasuda, N. Hosoi, H. Okumura, and E. Horikawa: Towards a System that Relieves Psychological Symptoms of Dementia by Music, Interna-tional Journal on Advances in Life Sciences, 5, 3&4, 126/136 (2013)

③ W.B. Cannon: The James-Lange theory of emotions: a critical examination and an alternative theory. Am J Psy-chol., 39, 106/124 (1927)

④ W. James: What is an emotion? Mind,9,188/205 (1884)

⑤ S. Schachter, J.E. Singer: Cognitive, Social and Physiological Determinants of Emotional State, Psychological Review, 69, 5, 379/399 (1962)

⑥ K. Nakayama, C. Oshima, R. Higashihara, K. Machishima: Mood Induction by Emotional Prosody Modification -Experiments that students read scenario of a folk story-, Proc. of the SICE Annual Conference 2015, 500/505 (2015)

⑦ 北大路：美味い豆腐の話，星岡（初出）(1933)http://www.aozora.gr.jp/cards/001403/files/49961_37670.html

⑧ 寺崎，岸本，古賀：多面的感情状態尺度の作成，心理学研究，62，6，350/356 (1992)

⑨ J.J. Aucouturier, P. Johansson, L. Hallb, R. Segnini, L. Mercadié, and K. Watanabe: Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction, Proceedings of the National Academy of Sciences of the United States of America, 113, 4, 948/953 (2015)

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

① 大島千佳，中山功一：音楽で継続する発話／発声を鎮める，音楽情報処理による障害者支援：情報処理，57(3)，pp.266-269（2016-02-15）.（査読なし）
http://id.nii.ac.jp/1001/00147617/

〔学会発表〕（計 5 件）

① 中山功一，谷口聖人，町島希美絵，大島千佳：患者の聴覚特性に応じた話し方の訓練システムの研究開発，情報処理学会アクセシビリティ研究会，2016-AAC-1(16),2016 年 7 月 29,30 日(NII，東京）2016.

② 中山功一，志田玲人，大島千佳：音声変換フィードバックによる気分誘導システムの実装，計測自動制御学会，システム・情報部門学術講演会 2015 講演論文集，発表日：2015 年 11 月 19 日（函館アリーナ，函館）2015.

③ 町島希美絵，石井弓子，大島千佳，細井尚人，中山功一：デイケア施設を利用する認知症者のための作業療法の個人化手法，第 42 回知能システムシンポジウム，DVD 論文集，F-02，2015．発表日：2015 年 3 月 18 日（北野プラザ六甲荘，神戸）

④ 大島千佳：プロソディや音楽による気分誘導，関係論的システムデザイン研究センター，第 8 回シンポジウム・プログラム．発表日：2015 年 3 月 2 日（同志社大学，京都）

⑤ 大島千佳，町島希美絵，中山功一：アクティビティ・ケアとしてのピアノ・レッスン　～達成感から得られる認知症者の QOL の向上に向けて～，計測自動制御学会　システム・情報部門学術講演会 2014 講演論文集，pp.788-793，2014．発表日：2014 年 11 月 21 日（岡山大学，岡山）

６．研究組織

(1)研究代表者

大島 千佳（Chika Oshima）
佐賀大学・大学院工学系研究科・客員研究員
研究者番号： 10395147

(2)研究分担者

中山 功一（Nakayama Koichi）
佐賀大学・大学院工学系研究科・准教授
研究者番号： 50418498