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

令和　２ 年　 ５ 月 ２４ 日現在

機関番号：１７２０１
研究種目：挑戦的研究（萌芽）
研究期間：2017～2019
課題番号：１７Ｋ２００１１

研究課題名（和文）音声による認知フィードバックを利用した気分誘導技術の研究開発

研究課題名（英文）Toward construction of a mood induction technology according to cognitive feedback

研究代表者
大島　千佳（Oshima, Chika）

佐賀大学・理工学部・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）　　4,900,000円

研究成果の概要（和文）：話者の音声をアレンジして話者自身にリアルタイムで聞かせることで，話者に心的状態を誤認識させ，心的状態を誘導する技術を明らかにすることを研究の目的とした．
本研究では音量・音高をリアルタイムに変換するEPROCs(Emotional PROsody Conversion system)と，既存の，音声に周波数を変換するフィルタをかけることで任意の感情への誘導するシステムによる，機械学習を用いた感情分類の実験を行った．その結果，後者のフィルタをかけることで音質データを特徴として分類器を作成した方が，前者の音高・音量を特徴として分類器を作成するよりも，高い正解率で感情を推定した．

研究成果の学術的意義や社会的意義
発話音声の音高と音量が変換された自分の声を聞く話者の気分が誘導されることを実験で明らかにしたいと考えた．しかし，話者の生の音声と変換した音声が二重に聞こえることと，変換した音声のわずかな遅延により，音声変換のみを気分誘導の要因とすることができなかった．そこで本研究では既存の周波数を変換するフィルタをかける手法と，本研究の音高と音量を変換する手法を比較する実験を行った．音高と音量を変換する手法は感情分類の正解率が低くなることがわかった．

研究成果の概要（英文）：The purpose of the research was to construct the technology that induces the speaker to misunderstand the mental state by arranging the speaker's voice and letting the speaker to listen to it in real time.
In this study, we compared our system, EPROCs (Emotional PROsody Conversion system) that converts volume and pitch in real time with a system that guides to any emotion by applying a filter that converts frequency to voice. We conducted an emotion classification experiment according to machine learning. As a result, a classifier that features sound quality data by applying the latter system could estimate the emotion with a higher accuracy rate than another classifier that features the pitch and volume data by applying the former system.

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：音声変換　音高と音量　機械学習　分類器

様　式　C－１９、F－１９－１、Z－１９、CK－１９（共通）

１．研究開始当初の背景

古くから「悲しいから泣く，楽しいから笑う」という「感情→生理学的変化（行為）」であるとするキャノン＝バード説と，「泣くから悲しい，笑うから楽しい」という「生理学的変化→感情」であるとするジェームズ＝ランゲ説がある．これに対し，生理学的変化の原因を類推することで感情が決定されるという「原因帰属の認知→感情」であるとするシャクターの情動二要因説[Schachter 62]がある．申請者らは，この情動二要因説を拡張し，実際には生理学的に変化していない（実際には笑っていない）が，変化している（自分が笑っている）と誤認識することで，実際に心的状態が誘導される（楽しくなる）という「生理学的変化の誤認識→感情」であるとする仮説を提唱している[中山 15]．従来，生理学的な身体変化が情動に与える影響（バイオフィードバック）は数多く検証されている．一方，申請者らの提唱する，音声による生理学的変化の誤認識のみで心的状態が誘導される（認知フィードバック）という研究報告はない．誤認識で心的状態が誘導されれば，世界的に新しい学術的成果となる．申請者らは，予備実験[Nakayama15][Oshima15]から，この仮説が支持されると考えた．

【引用文献】

[Schachter 62] Schachter, S, Singer, J.E.: Cognitive, social and physiological determinants of emotional state, Psychological Review, Vol.69, No.5, pp.379-99,1962.

[中山 15] 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

[Nakayama15] K. Nakayama, C. Oshima, R. Higashihara and K. Machishima,“Mood Induction by Emotional Prosody Modification -Experiments that students read scenario of a folk story-,” the SICE Annual Conference 2015, pp.500-505.

[Oshima15] 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,” International Journal on Advances in Life Sciences, Vol.5, No. 3&4, pp.126-136.

２．研究の目的

話者の音声をアレンジして話者自身にリアルタイムで聞かせることで，話者に心的状態を誤認識させ，心的状態を誘導する技術を明らかにすることを研究の目的とした．そのために，心的状態を誘導する発話要因を機械学習手法により明らかにし，その発話要因をアレンジしてフィードバックする技術を確立することを目指した．

本研究では音量・音高をリアルタイムに変換する手法により任意の感情へ誘導することを目指している．既存の，音声に周波数を変換するフィルタをかけることで任意の感情への誘導するシステムと比較実験を行うことで，目的とした感情の付与が可能であるか検討した．

３．研究の方法

３．１．リアルタイム変換システム EPROCs

開発した EPROCs(Emotional PROsody Conversion system)は，音声に含まれる音量と音高をリアルタイムに変換することのできるシステムである．64ms ごとに，ファイルから入力された過去 64ms 間の音声の音高と音量を検出する．最後に，変換した音声をファイルに出力する．

３．１．１．　音量・音高検出

音声検出・変換処理を行う関数では，順に，32bit から 16bit への音声データの変換，リサンプリング，音量・音高の検出，音量・音高の変化量の指定，音量・音高の変換，16bit から 32bit への音声データの再変換が行われる．音量・音高の検出を行う関数は，16bit，8000Hz の条件下で利用する．関数内に 512samp のキューバッファを持っており，呼び出される度に 43samp(8000Hz)(=256samp(48000Hz)　5.3ms)の入力バッファが格納される．キューバッファが埋まると音量・音高の検出を行い，それぞれの値を出力する．すなわち，512samp/8000Hz=64ms 前までのデータを用いて検出している．また，音高推定の信頼度も出力しており，この値と音量の積と，閾値を比較するなどして，音声区間を特定することができる．

３．１．２．　音量・音高変換

音高変換を行う関数は，16bit，48000Hz 条件下で利用する．関数内に 2048samp のキューバッファを持っており，ここに約 42.7ms(　2048samp/48000Hz)間の音声データが保存できる．関数が呼び出される度に 256samp(約 5.3ms　256samp/48000Hz)の入力バッファが格納される．8 回呼び出されてキューバッファがすべて埋まると，全体音高変化量に従って音高変換を行う．変換された約 42.7ms(2048samp)の音声は，出力バッファに格納され，出力される．すなわち，全体音高量に従って，42.7ms 間の音声の音高を変換する．全体音高変換後，同様に決定された全体音量変化率に従って音量変換を行う．全体音量変化量は，0 以上で上限はないが，16bit の範囲(-32768~32767)を超えないように調整される．

３．２．実験

DAVID(Da Amazing Voice Inflection Device) [Aucouturier 2016]は，EPROCs と異なり，音声に周波数を変換するフィルタをかけることで任意の感情への誘導を実現している．そこで本実験では，DAVID と音量・音高を変換する EPROCs を比較して，音響的特徴が機械学習における感情の認識にどのような影響を与えるか検証する．

３．２．１．感情ラベル付き音声データ

本研究で用いた感情ラベル付き音声データは，音声資源コンソーシアムで公開されている，感

情評定値付きオンラインゲーム音声チャットコーパスである[OGVC]．音声データは，オンラインゲーム中のプレーヤーに音声チャットを利用させ，自然に感情が表出した音声を録音した自然対話音声と，自然対話音声の転記テキストを元に，プロの俳優 4 名に対話形式で発声させた演技音声の 2 種類が用意されている．本研究では，自然音声に比べて感情の推定の容易だった演技音声を用いた．

演技音声は，4 名のプロの俳優(男性 2 名，女性 2 名)により発話されている．音声データは WAVE 形式(44.1kHz・16bit・モノラル)で録音されており，発話の前後には 400ms の無音区間が挿入されている．それぞれの話者で発話数は同一となっている．演技音声における発話時間は前述の無音区間を除き，最短で約 1 秒，最長で約 4 秒となっており，平均発話時間は約 2 秒である．本実験では，怒り(ANG)，喜び(JOY)，悲しみ(SAD)，驚き(SUR)の 4 種類の感情の音声ファイルを用いた．それぞれ，320 個，336 個，336 個，384 個で，合計 1376 個であった．

３．２．２．機械学習を用いた感情分類の検証

機械学習を用いた感情分類の実験を行った．感情ラベル 4 種類 1376 個の音声データのうち，960 個の音声データをトレーニングデータとして用いて分類器を構築し，416 個の音声データをテストデータとした．

音質による分類と，音量・音高による分類の 2 種類の分類を行った．音質による分類では，DAVID [Aucouturier 2016]を用いて各感情を目標とした音声変換を加えた後，そのデータの MFCC(メル周波数ケプストラム係数)と呼ばれる特徴量を用いて実験を行った．MFCC とは，人間の聴覚特性に合わせ，個人差の大きいピッチ成分を除去し，韻律の特定に重要な声道の音響特性やフォルマント成分などの，低周波部分を細かく抽出できる手法である．

音量･音高による分類では，EPROCs を用いて変換後の音量･音高を特徴量として用いた．EPROCs を用いることで容易に音量・音高の操作ができる．音量は 0 から 75 の範囲で表され，音高は 0 から 400 の範囲で表される．

３．２．３．変換のパターン

（１）音質データを特徴量として分類器を構築し，テストデータがどれだけ正確に感情を分類できるか検証した．

(a)コントロール条件：音声変換システムを使用しない．

(b)JOY 変換：DAVID を用いて，トレーニング・テストデータの両方に JOY を目標とした変換を加えた．

(c)JOY ラベル以外の音声を JOY に変換：DAVID を用いて，テストデータの JOY ラベル以外の音声に JOY を目標とした変換を加えた．トレーニングデータには変換を加えていないため，(a)と比較し，JOY の予測が増加するか検証する．

(d)SAD 変換：DAVID を用いて，トレーニング・テストデータの両方に SAD を目標とした変換を加えた．

(e)SAD ラベル以外の音声を SAD に変換：DAVID を用いて，テストデータの SAD ラベル以外の音声に SAD を目標とした変換を加えた．トレーニングデータには変換を加えていないため，(a)と比較し，SAD の予測が増加するか検証する．

(f)SAD ラベルの音声を JOY に変換，JOY ラベルの音声を SAD に変換：DAVID を用いて，テストデータの JOY ラベルの音声に SAD，テストデータの SAD ラベルの音声に JOY を目標とした変換を加えた．(a)と比較し，JOY と SAD の予測の増減の変化を検証する．

(g)ANG ラベルの音声を JOY に変換：DAVID を用いて，テストデータの ANG ラベルの音声に JOY を目標とした変換を加えた．(a)と比較し，ANG の予測の減少と JOY の予測の増加について検証する．

（２）音量・音高データを特徴量として分類器を構築し，テストデータがどれだけ正確に感情を分類できるか検証した．

(h)コントロール条件：音声変換システムを使用しない．

(i)JOY 変換：EPROCs を用いて，テストデータ全体に JOY を強調させる変換を加えた．(h)と比較し，JOY の予測の増加について検証する．

(j)JOY 抑制変換：EPROCs を用いて，テストデータ全体に JOY を抑制させる変換を加えた．(h)と比較し，JOY の予測の減少について検証する．

(k)ANG 抑制変換：EPROCs を用いて，テストデータ全体に ANG を抑制させる変換を加えた．(h)と比較し，ANG の予測の減少について検証する．

(l)SAD 変換：EPROCs を用いて，テストデータ全体に SAD を強調させる変換を加えた．(h)と比較し，SAD の予測の増加について検証する．

３．２．４．EPROCs による音声変換方法

a．直前音量との差分から微分を取り，音高を変化させる方法【Vd:P±】

入力される音声の 128ms 前の 128ms 間の平均音量を取り，直前音量として扱う．その直前音量とマイクから入力される音声の音量との差分から，出力される『音高』を差分に比例して，差分値が正(負)である時に正(負)方向に変化率を変動するアルゴリズムと，差分値が正(負)である時に負(正)方向に変化率を変動するアルゴリズム，2 つの変換アルゴリズム(Vd;P±)による変換を行った．SAD 変換では Vd:P+の変換アルゴリズムを用いた．

b．基準音量との差分から，音量を変化させる方法【Vt:V±】

あらかじめ設定されている基準音量とマイクから入力される音声の音量との差分から出力さ

れる『音量』を差分に比例して，差分値が正(負)である時に正(負)方向に変化率を変動するアルゴリズムと，差分値が正(負)である時に負(正)方向に変化率を変動するアルゴリズム，2 つの変換アルゴリズム(Vt:V±)による変換を行った. JOY 変換では Vt:V+の変換アルゴリズムを用いた.

c．直前音高との差分から微分を取り，音高を変化させる方法【Pd:P±】

入力される音声の 128ms 前の 128ms 間の平均音高を取り，直前音高として扱う．その直前音高とマイクから入力される音声の音高との差分から，出力される『音高』を差分に比例して，差分値が正(負)である時に正(負)方向に変化率を変動するアルゴリズムと，差分値が正(負)である時に負(正)方向に変化率を変動するアルゴリズム，2 つの変換アルゴリズム(Pd;P±)による変換を行った．非 JOY 変換では Pd:P+の変換アルゴリズムを，非 ANG 変換では Vd:P-の変換アルゴリズムを用いた.

３．２．５．分析

上記の各変換方法をコントロール条件と比較した.

$$\text{precision（適合率）} = \frac{TP\ (\textit{真陽性})}{FP\ (\textit{偽陽性}) + TP\ (\textit{真陽性})}$$

$$\text{recall（再現率）} = \frac{TP\ (\textit{真陽性})}{FN\ (\textit{偽陰性}) + TP\ (\textit{真陽性})}$$

$$\text{F 値} = 2 \times \frac{precision \times recall}{precison + recall}$$

【引用文献】

[Aucouturier 2016] Aucouturier, J. J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. Proceedings of the National Academy of Sciences, 113(4), 948-953.

[OGVC] 感情評定値付きオンラインゲーム音声チャットコーパス(OGVC). http://research.nii.ac.jp/src/OGVC.html (参照 2020-4-12).

４．研究成果

４．１　結果

表 4.1 に音質データを特徴として分類器を作成して推定した感情分類の結果を適合率と F 値で示す．どの感情も高い率で正解の感情を推定した.

表 4.1. 音質データを特徴とした機械学習による分類の結果（コントロール条件）(%)

		機械学習による推定			
		ANG	JOY	SAD	SUR
ラベル	ANG	**50**	19	8	19
	JOY	12	**67**	8	16
	SAD	12	19	**59**	11
	SUR	13	21	5	**77**
F値		0.546	0.585	0.652	0.644

表 4.2 に変換 b-e と g，表 4.3 に変換 f の結果（適合率）を示す．これらの結果は，全ての音声に DAVID を用いて音質を JOY に近づくように変換しても，JOY と予測されることが少ないことを示している．また，SAD ラベル以外の音声に DAVID を用いて音質を SAD に近づくように変換しても，SAD と予測されることが増加しないことを示している．一方で全体を SAD に変換すると SAD の適合率が上昇した．ANG を JOY に変換すると ANG ラベルの音声が JOY に分類され適合率は大きく低下した.

表 4.2. 音質データを特徴とした機械学習による分類の結果（変換 b-e,g）(%)

		推定				
		(b)JOY変換	(c)JOY以外JOY変換	(d)SAD変換	(e)SAD以外SAD変換	(g)ANGをJOYに変換
		JOY	JOY	SAD	SAD	ANG
ラベル	ANG	17	23	13	30	20
	JOY	57	51	18	21	60
	SAD	15	16	71	63	19
	SUR	22	23	16	19	17

表 4.3 音質データを特徴とした機械学習による分類の結果（変換 f）（%）

		推定	
		JOY	SAD
ラベル	JOY	57	12
	SAD	16	67

表 4.4 に音量・音高データを特徴として分類器を作成して推定した感情分類の結果を適合率と F 値で示す．どの感情も音質データを特徴とした分類器による感情分類よりも適合率，F 値ともに低い．

表 4.4. 音量・音高データを特徴とした機械学習による分類の結果（コントロール条件）

	機械学習による推定				
		ANG	JOY	SAD	SUR
ラベル	ANG	**38**	16	22	20
	JOY	19	**37**	20	27
	SAD	24	20	**41**	16
	SUR	12	24	16	**64**
F値		0.402	0.370	0.410	0.527

表 4.5 に変換 i～l の結果（適合率）を示す．どの適合率も低く，これらの変換方法の効果がないことを示している．

表 4.5. 音量・音高データを特徴とした機械学習による分類の結果（変換 i-l）（%）

		推定			
		(i)JOY変換	(j)JOY抑制変換	(k)ANG抑制変換	(l)SAD変換
		JOY	JOY	ANG	SAD
ラベル	ANG	23	21	38	19
	JOY	36	33	29	21
	SAD	31	25	15	32
	SUR	12	26	18	26

４．２　まとめ

話者の音声をアレンジして話者自身にリアルタイムで聞かせることを行うために，まずは，機械学習を用いた感情認識プログラムを開発して評価実験を行った．本研究では，音量と音高のデータを変換することで，感情を付与することを目指したが，周波数を変換するフィルタをかける方法よりも適合率は低かった．

また，話者の音声をリアルタイムに変換する実験も試みたが，話者自身の声と変換後の声の両方が聞こえてしまい，正確に評価することが困難であった．

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Haruka Yanagi, Chika Oshima, Koichi Nakayama	４．巻 11570
２．論文標題 Estimating Timing of Head Movements Based on the Volume and Pitch of Speech	５．発行年 2019年
３．雑誌名 Lecture Notes in Computer Science	６．最初と最後の頁 322-332
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-22649-7_26	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計2件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 野柳晴華，大島千佳，中山功一
２．発表標題 対話ロボットの自然な“あいづち”に向けた研究
３．学会等名 第14回コンピューテーショナル・インテリジェンス研究会
４．発表年 2018年

１．発表者名 谷口聖人，大島千佳，中山功一
２．発表標題 発話の韻律変換による感情表現の検討
３．学会等名 情報処理学会全国大会
４．発表年 2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	中山　功一 (Nakayama Koichi) (50418498)	佐賀大学・理工学部・准教授 (17201)	