Vol.2018-MUS-119 No.53
Vol.2018-SLP-122 No.53
2018/6/17

発話の好感度改善を目的とした音声加工技術の検討

堀池 梓哉[1,a)]　森勢 将雅[1]

概要：コミュニケーションにおける発話は, 他者のパーソナリティ推定に影響を及ぼすといわれているため, 相手に良い印象を与えるための 1 つの要素である. サービスの一環であるアナウンス音声も, 利用者に良い印象を与える音声であることが望ましい. 印象の良いアナウンス音声を手軽に入手するため, 入力された音声の好感度を改善する音声変換手法を提案する. 発話の好感度を改善する手法として F0 を高くする手法があるが, 本研究では, スペクトル包絡を制御することで, 音色による好感度向上を試みる. 本稿は, フォルマントや, スペクトル傾斜が好感度に影響するという知見に基づき, 各フレームのスペクトル包絡において, 振幅値を周波数軸に一定量シフトする変換によって好感度改善を図った. 変換した音声を用いて主観評価実験を行い, 実験結果では一部の音声が好感度が上がる有意傾向を示した.

1. はじめに

人は, 他者との関わりを持つためにコミュニケーションを取る. コミュニケーションを取る為の重要な手段の一つとして, 発話が挙げられる. 発話は, 言語情報を伝達するだけでなく, 他者のパーソナリティ推定に影響を及ぼす [1]. そのため, 発話の仕方によって, 相手に悪い印象を与えてしまう可能性が考えられる. それは, スーパーマーケットなどの店内アナウンス, 電車やバスといった公共交通機関の車内アナウンスに関しては, より顕著である.

アナウンスは, 目玉商品の紹介や行先案内などを連絡するものであり, 多数の人がアナウンス音声を聴く. サービスの一環として使用されるアナウンス音声の印象が良くないものであった場合, 様々なデメリットが考えられる. 例えば, 売りたいものを売ることが出来ないことや, 乗客を不快にさせてしまい, 別の公共交通機関を利用されてしまうことなどが挙げられる. 一方で, 印象が良いアナウンス音声であれば, 顧客の購買意欲を刺激して, 売り上げに貢献することや, リピーターが増加することに繋がる可能性がある. そのため, 良い印象のアナウンス音声を収録することは重要であるといえる.

多くの人に良い印象を与えるアナウンス音声の収録には, 印象の良い発話ができるプロのアナウンサや, プロの声優を招いて収録する方法がある. 一方で, プロを招くための費用や, プロの音声を収録する為の日程調整などといった手間がかかるため, 手軽に良い印象の音声を収録することは困難である. そこで, セミプロや練習生などの音声を良い印象の音声に変換できれば, より低コストで手軽に良い印象のアナウンス音声を入手することができる.

本研究では, 印象の良い発話を容易に入手するため, 発話音声の好感度を改善することを目的とした, 音声変換手法を提案する. 本手法の実現によって, アナウンスに向けた音声収録の際, プロを招かずにセミプロ等による音声を収録し, 本手法の音声加工を施すことで, より印象の良い音声を合成することが期待される. そのため, 提案した音声変換手法が好感度に与える影響と, 変換した音声の品質劣化について検証するための主観評価実験を実施する.

本稿は, 第 2 節にて, 好感度に関する定義を行い, 関連研究について述べる. 第 3 節おいて, 音声分析合成に用いるシステムである WORLD について述べ, 変換する音響特徴量を選定し, その変換方法を提案する. 第 4 節では, 本変換手法についての主観評価実験と, その結果を論ずる. 第 5 節では, 主観評価実験の考察を述べる. 第 6 節において, 今後の研究の展望について述べる.

2. 好感度に関する定義とその関連研究

2.1 本研究における発話の好感度に関する定義

音声から知覚される好感度を改善するためには, まず, 好感度に関する定義を行う必要がある. ここで, 好感度は発話者個々人により依存するため, 客観的指標としては定義しにくい. 本研究では, 好感度を相手に与える主観的な聴覚的印象と定義する. すなわち, 発話の好感度が低い発話とは, 相手に悪い聴覚的印象を与える発話を指し, 発話の好感度が高い発話とは, 相手に良い聴覚的印象を与える発話を指す. また, 発話の好感度が中程度であるものは, 相手に

[1] 山梨大学
[a)] g18tk013@yamanashi.ac.jp

© 2018 Information Processing Society of Japan

良い聴覚的印象も, 悪い聴覚的印象も与えない発話を指す. この好感度は聴取者個々人により異なる可能性があるが, その中で共通する傾向を導き出し, 音声加工によって改善することが本研究の目標となる.

2.2 関連研究

関連研究として, 発話の好感度と声の高さの関連について着目した研究と, 男性と女性の評価の差異に着目した研究について述べる. Jones ら [2] は, 女性の声の高さが評価の好みに与える影響について調査した. この研究では, 声の高さを制御した音声を合成し, それぞれの音声を逆再生したものも用意した. これらの音声を男性 30 名, 女性 30 名に聴取させた. その結果, 順再生された音声では, 評価者が男性の場合, 声を高く変換したものを好むという傾向となった. 一方で, 評価者が女性の場合, 声の高さを制御しても特に評価に有意な差が見られないことが示された. また, 逆再生された音声では制御した音声と元音声の評価に有意な差はなかった. 吉元ら [3] は, 評価者に発話の F0 を操作させることによって好感度の改善を図った. その結果, 女声発話に関して, 平均的な女性の声の高さよりも少し高めの声が好まれるという結果が得られた.

これらの研究結果は, 女性発話に関して, F0 を高くすることで好感度が高くなることを示唆している. そのため, 行った操作は F0 のみであり, スペクトル包絡や非周期性指標といった音響特徴量の変換による好感度の影響の調査ではない.

横森ら [4] は, 男性の評価と女性の評価の相関や, 好感度に影響を与える音響特徴量について調査した. この研究では, 女性発話を用いて主観評価実験を行った. これらの音声を男性 47 名, 女性 80 名に聴取させ, 好感度の評価を行った. その結果, 男性の評価と女性の評価の相関は低く, 性別によって評価の傾向が異なる可能性を示唆した. また, 評価と対応がある音響特徴量があるか調査した結果, 女性の評価と F0, フォルマント周波数, 倍音振幅差分, スペクトル重心 [5], スペクトル傾斜に相関がみられ, 男声の評価とスペクトル傾斜に相関がみられた.

この研究は, 複数の音声の好感度を評価し, 音響特徴量との相関を調査した研究であるが, 元音声と音響特徴量を制御した音声の好感度の変化を確認していない. また, Junes ら [2] と吉元ら [3] の研究では, F0 を高く変換することで好感度が高くなることが示唆されていたが, この研究では F0 と男性の評価には相関がみられなかった. これは評価者が男性の場合, 同一のスペクトル包絡で F0 のみが異なる音声の場合, より高い F0 の発話が好まれる可能性を示唆している.

2.3 本研究の位置付け

先行研究 [2][3] では, F0 を高くする変換を行うことで, 好感度が改善することを示唆した. また, 文献 [4] では, 女性発話の好感度と音色に関する音響特徴量に相関をあることを示唆した. これらの研究から, スペクトル包絡を変化させることで, 音声の好感度が変化する可能性がある.

本研究では, 音の高さとは異なる音響特徴量の制御を行うことによって, 音声の好感度の改善を図る. その際, 横森ら [4] の実験結果に基づいて, 変換する音響特徴量を選定し, 選定した音響特徴量を変化させる音声変換手法を提案する. また, 提案した音声変換手法によって変換された合成音声を用いて音声の主観評価実験を行い, 好感度の変化を分析する.

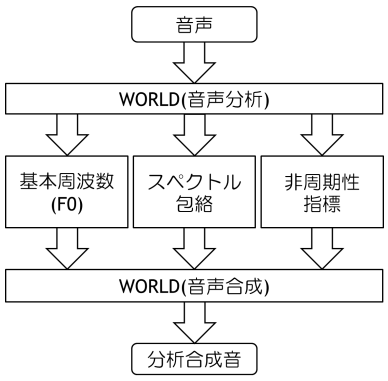


図 1 WORLD による音声処理

3. 好感度改善のための音声変換手法

3.1 音声分析合成システム WORLD

好感度の改善のため, 音声分析合成システムを用いた音声変換を行う. 音声分析合成の際に, 音声の品質劣化の影響が大きくなると, 変換手法とは別の要因で好感度が変化する可能性がある. そのため, 高い品質で音声分析合成できる WORLD[6][7] を使用する. WORLD は, 高品質な音声分析合成を目的にして提案され, 肉声に匹敵する品質の合成音を生成する音声分析合成システムである. WORLD では, 図 1 のように音声から F0, スペクトル包絡, 非周期性指標の 3 つのパラメータを高精度に推定する. ここで, 3 つのパラメータは, 独立に操作することができる. 例えば, F0 の操作を行うことでイントネーションの変換 [8] を行うことが可能となる. パラメータの操作後, 3 つのパラメータから音声を合成し, 変換音声を出力することで, 音声変換を実現する.

3.2 変化させる音響特徴量

本研究では, 2.3 項で述べたように, 横森ら [4] の評価実験より得られた, 好感度評価との相関がみられた音響特徴量を変化させる音声変換手法を提案する. この研究では, 評価者本人が発話の好感度について評価したと考えられるものを評価性因子, 評価者本人が普遍的な男性にモテる発話を想像して評価したと考えられるものを想像好感度因子と

© 2018 Information Processing Society of Japan

し, 2 つの因子と音響特徴量の相関を調査した. 本研究の目的は, 評価者本人からの評価の上昇であるため, 評価性因子と相関が見られた音響特徴量を変化させる. また, 同研究では, 性別によって好感度の評価が異なる可能性を示唆したが, 男性と女性の評価に対する影響が反対にならない限り, 全体の評価は変化すると考えられる. そのため本研究では, 性別による評価の差異を考慮しない. 男女どちらかの性別の評価性因子と相関がみられた音響特徴量は, スペクトル傾斜, フォルマント周波数, 倍音振幅差分の 3 つである.

スペクトル傾斜は, 男性の評価性因子と/u/のスペクトル傾斜, 想像好感度因子と/i/のスペクトル傾斜について正の相関が見られた. この結果より, 男性はスペクトル傾斜が急なものほど評価が上がる傾向となった. また, 女性の評価性因子と/a/, /i/, /u/のスペクトル傾斜, 想像好感度因子と全ての母音のスペクトル傾斜について負の相関が見られた. この結果より, 女性はスペクトル傾斜が緩やかなものほど評価が上がる傾向となった. 男女で逆の傾向を示しており, 評価者の性差を考慮しない評価において, 音声を変換しても好感度の改善が見られないことが予想される.

フォルマント周波数については, 舌の高さに相関があるとされている F1[9], 舌の前後方向の位置と相関があるとされている F2[9] について調査した. 男性について, 各因子とすべての母音に相関は認められなかった. 一方で, 女性のすべての因子と/u/, /o/の F1, /e/の F2 に相関が認められた. この結果は, /u/, /o/の F1 が高いほど評価は高くなり, /e/の F2 が低いほど評価が高くなることを示唆している.

文献 [10] では, 低域の調波構造に着目した音響特徴量として, 第 1 倍音の振幅から第 3 倍音の振幅を引いた差分が発話の音響印象に関連するという結果が得られている. 同研究では, この音響特徴量を倍音振幅差分と定義した. 倍音振幅差分と男性のすべての因子に相関は認められなかったが, 女性の評価性因子と/a/, /e/, /o/の倍音振幅差分について有意な相関が認められた.

本稿では, 評価性因子と相関が認められた音響特徴量である F1, F2, 倍音振幅差分に着目し, これらの音響特徴量のうち F1 を変化させることで好感度の改善を試みる.

3.3 本変換手法における変換式

F1 の変換を行うにあたり, 周波数軸を線形に伸縮して変換した場合, F2 も連動して変化するため, F1 のみを変化させるには, 周波数軸の非線形変換が必要となる. このような複雑な変換は, 出力する音声の品質劣化が起こることが指摘されている [11]. 品質劣化がある音声は, 品質劣化がない音声よりも好感度が低くなると考えられるため, 品質劣化を抑える必要がある. 品質劣化を最小限に抑えて F1 を変化させるため, F1 を含むスペクトル包絡全体をシフトさせる変換手法を提案する. スペクトル包絡の各時間フレームについて, 周波数軸の正方向, もしくは負方向に各振幅値

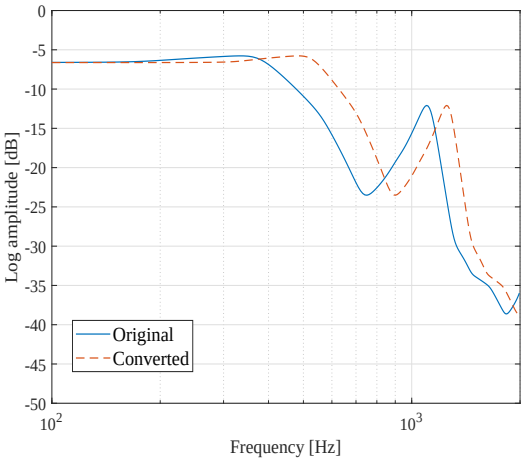


図 2 あるフレームの元音声と変化量 $v = 150$ の合成音声のスペクトル包絡

を一定量シフトする. その後, シフトしたスペクトル包絡のパラメータを用いて WORLD で合成させることで, 変換音声を出力する. この変換手法は, 単純な変換であり, スペクトル包絡の変化量も小さいため, 品質劣化が起きにくいと予想される.

本変換手法における, スペクトル包絡の変換式について説明する. 時間フレーム a の周波数 f における振幅を $H(a, f)$ とする. 変化量 v で音声変換を行うとき, 式 (1) を用いる. $f_{\max}$ はスペクトル包絡が取りうる最大周波数を表し, $f_{\min}$ は最小周波数を表す.

$$H(a,f) = \begin{cases} H(a, f-v) & (f_{\min} \leq f - v \leq f_{\max}) \\ H(a, f_{\max}) & (f - v > f_{\max}) \\ H(a, f_{\min}) & (f - v < f_{\min}) \end{cases} \quad (1)$$

図 2 に, ある時間フレームの元音声と変化量 $v = 150$ の合成音声のスペクトル包絡を示す, 縦軸は対数振幅, 横軸は周波数を表す. この図の変換では, スペクトル包絡の振幅を周波数軸の正方向にシフトしている.

4. 音声評価実験

4.1 実験に使用した発話音声

文献 [12] の音声データベースを用いて主観評価実験を行う. この音声データベースには, 男性 10 名, 女声 10 名の音声が収録されている. 収録の際, 発話者に, 「相手に良い印象を与える好感度の高い発話, 相手に特に良い印象も悪い印象も与えないであろう好感度が中程度の発話, 相手に悪い印象を与える好感度の低い発話 3 種類を自分の考えるように表現する」という教示を与え, 31 文章を発話させた. 発話文章は, テキストから得られる好感度が高いもの 10 文章, 中のもの 11 文章, 低いもの 10 文章である. 収録されている音声数は, 1 人当たり 93 音声, 合計 1860 音声である. 収録条件を表 1 に示す. 本実験では, 好感度が中程度の発話を表現した男性 3 名, 女性 3 名の音声を使用した. これ

© 2018 Information Processing Society of Japan

表 1 評価に用いた音声データーベース

収録人数	6 名 (男性 3 名, 女性 3 名)
収録環境	防音室 (A–weighted SPL : 17 dB)
発話内容	高 「桜がきれいです，あなたを愛しています，世界はすばらしいですね，会えてよかったですね.」 中 「精神は創造的高揚によって一種のメタモルフォーゼを敢行する.」 低 「見やがれ，そんなだから，お前は，大学にも，行かれないんだ，行けるもんか.」
A/D 変換	96 kHz/24 bit
収録機器	Antelope ECLIPSE 384 NEUMANN U87Ai

表 2 音声評価実験の実験条件

評価方法	CMOS 評価法
評価音声数	6 名 × 3 文章 × 6 通り = 108 音声
評価者数	7 名
評価環境	一般的な教室 (A–weighted SPL : 35 dB)
再生機材	Roland QUAD–CAPTURE SENNHEISER HD 650

は, 男女による効果の差異, 個人による効果の差異を確認するためである. また, 発話内容から得られる好感度によって本変換手法の効果の違いがあるかを確認するため, 発話した文章は, テキストから得られる好感度が高いもの, 中のもの, 低いものである 3 文章を用いた.

4.2 実験条件

評価者数などの実験条件を表 2 に示す. 実験は日常的に聴取する発話の好感度を測るため, 一般的な教室で行った. 全ての評価者は, 正常な聴力を有している. 本提案手法を用いて, スペクトル包絡全体を −50 Hz, 0 Hz, +50 Hz, +100 Hz 変化させる変換を行い, 1 つの音声から 4 つの変換音声を合成する. これらの変化量は, 予備実験により決定した. その後, 変化量 0 Hz の音声を基準音声とし, その他の音声を評価音声とした CMOS (comparison mean opinion score) 評価法で評価実験を実施する. CMOS 評価法では, 基準音声と評価音声の 2 つの音声を比較して評価を行う. 2 つの音声の順番はランダムに変化し, どちらが評価音声であるかは評価者に知らせない. 提示する 1 組の発話音声の長さは約 16 秒であり, 評価者には 1 組の発話音声を提示した後, 約 4 秒の回答時間を与えた. その試行を全 108 音声について行った. なお, 話者音声の提示順序は, ランダマイズした. 回答は質問紙形式であり, 以下の質問を用いた.

Q1 : この人の声は好きですか
Q2 : この音声の音質は良いですか

これら 2 つの質問項目について, それぞれ −3〜3 の 7 段階評価で評価させた. −3 は前音声よりも非常に悪い, −2 は前音声よりも悪い, −1 は前音声よりもやや悪い, 0 は前音声とほぼ同じ, 1 は前音声よりもやや良い, 2 は前音声よりも良い, 3 は前音声よりも非常に良い, である. Q1 の質問項目を選択した理由は, 関連研究 [4] の評価実験において, この質問項目が男女共に評価性因子の因子負荷量が大きかったためである. 評価性因子の負荷量が多いことは, 実際に聴取した発話音声の印象を評価に反映しやすい質問項目であると考えられるため, 本実験でもこの質問項目を用いた. Q2 は, 変換音声の品質の劣化度を明らかにするためである.

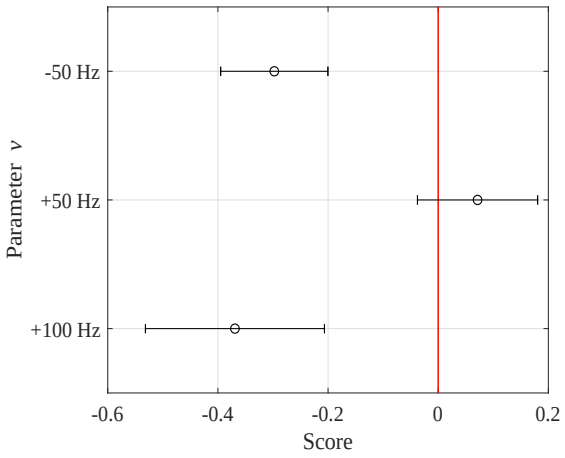


図 3 Q1 「この人の声は好きですか」 の評価結果

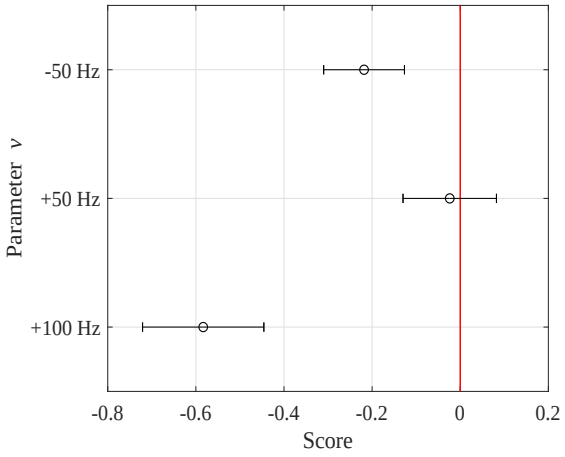


図 4 Q2 「この音声の音質は良いですか」 の評価結果

4.3 実験結果

主観評価実験の結果を図 3 に示す. 横軸は変化量, 縦軸は Q1 に対する評価値, 誤差棒は 95%信頼区間を示す. −50 Hz, +100 Hz の音声は 95%信頼区間に評価値 0 が含まれないことから, 評価が有意に低下することを示した. これは, スペクトル包絡全体を周波数軸の負方向に 50 Hz 以上シフトする音声変換と, 正方向に 100 Hz 以上シフトする音声変換を行うと評価が下がる傾向を示す. +50 Hz の評価値は 0 よりも高かったが, 有意差は見られなかった.

図 4 に, Q2 に対する評価値と 95%信頼区間を示す. −50 Hz, +100 Hz の音声は, 95%信頼区間に評価値 0 が含まれないことから, 評価が有意に低下することを示した. これ

© 2018 Information Processing Society of Japan

表 3　テキスト別の評価値

	−50 Hz	+50 Hz	+100 Hz
テキスト評価高	**−0.20**	0.05	**−0.43**
テキスト評価中	**−0.35**	0.14	**−0.35**
テキスト評価低	**−0.35**	0.05	**−0.31**

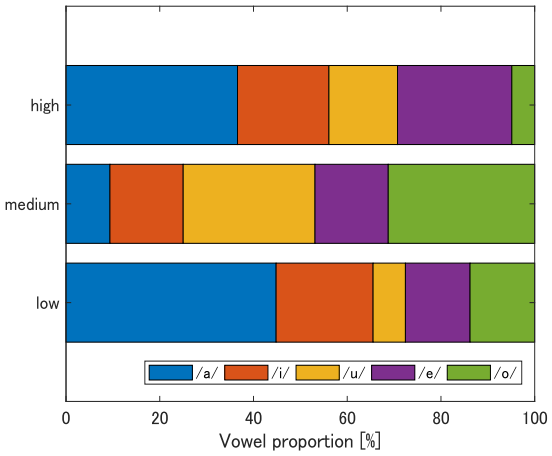


図 5　文章毎の母音割合

表 4　発話者別の評価値

	−50 Hz	+50 Hz	+100 Hz
男性 1	**−0.67**	0.29	**−0.36**
男性 2	**−0.33**	−0.19	**−0.38**
男性 3	**−0.38**	0.17	**−0.76**
女性 1	−0.10	0.04	0.12
女性 2	−0.21	0.07	−0.31
女性 3	−0.10	0.05	**−0.52**

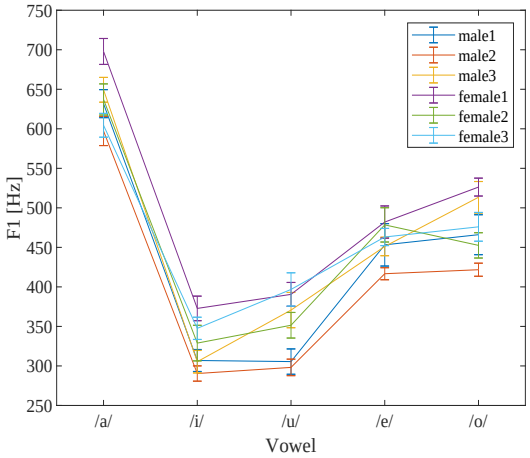


図 6　発話者別の F1

らの結果より, −50 Hz 以下, +100 Hz 以上の変換は, 品質劣化を起こすことが示されたといえる.

テキスト別の Q1 の評価値の平均を表 3 に示す. 90%信頼区間に評価値 0 が含まれないのものを下線で, 95%信頼区間に評価値 0 が含まれないのものを太字で, 99%信頼区間に評価値 0 が含まれないのものを太字に加え下線で示す. テキスト別の評価結果では, すべてのテキストで −50 Hz, +100 Hz の音声の評価が有意に低下した. また同様に, すべてのテキストで +50 Hz の評価値の有意差は見られなかったが, 評価値が 0 よりも高かった.

また, 発話者別の Q1 の評価値の平均を表 4 に示す. 発話者別の評価結果では, すべての男声の −50 Hz, +100 Hz の音声の評価が有意に低下した. 一方で, 女性 1 の +100 Hz の音声の評価は有意差が見られなかったが, 0 よりも高かった. また, 男声 1 の +50 Hz の音声の評価は, 90%信頼区間に評価値 0 が含まれないことから, 評価が上がる有意傾向であることを示した. この結果より, 提案手法によって好感度を改善する有意傾向が見られる音声の存在が示された.

5. 考察

5.1 全体評価結果に関する考察

主観評価実験の結果である図 3 から, 変化量 −50 Hz, +100 Hz の音声は, 変化量 0 Hz の音声と比べて評価値が有意に下がるという結果が得られた. これは, −50 Hz の変換は, 好感度が下がる方向への変換による可能性がある. また, +50 Hz の変換で評価値が 0 よりも高くなったことから, 周波数軸を正方向へシフトする変換は, 好感度が上がる変換である可能性を示唆する. 一方で, +100 Hz の変換は, 好感度が低下する評価傾向が見られた. これは図 4 より, +100 Hz の変換は, 変化量 0 Hz の音声と比べて品質が落ちている結果が得られており, 品質劣化により, 好感度の低下した可能性を示唆している.

以上のことから, +50 Hz と +100 Hz の間に好感度のピークがある可能性がある. 一方で, 変換を行うことで好感度が低下する可能性もあり, 最適値は話者に依存する可能性がある.

5.2 テキスト別の評価結果に関する考察

Q1 のテキスト別の評価結果である表 3 から, どのテキストも 変化量 −50 Hz, +100 Hz の音声は, 変化量 0 Hz の音声と比べて評価が有意に下がり, +50 Hz の音声は, 評価値が 0 よりも高いが有意差が認められないという結果が得られた. この結果から, 本変換手法は, テキストから得られる好感度に関係なく, すべてのテキストで同様の効果が見られることが示唆された. また, 図 5 に実験で用いた文章の各母音数を示す. この図より, 実験で用いた文章では, /a/を多く含む文章が 2 文章, /u/, /o/を多く含む文章が 1 文章であった. /i/や/e/を多く含む文章で実験を行っていないが, /i/と F1 の値が近い /u/や, /e/と F1 の値が近い /o/を多く含む文章で同様の効果が得られている. そのため, 本変換手法は, 母音ごとに変換関数を作るような処理は不要であり, 簡単な処理で効果が期待できる方法である可能性がある.

© 2018 Information Processing Society of Japan

5.3 発話者別の評価結果に関する考察

Q1 の発話者別の評価結果である表 4 から, すべての男性で変化量 −50 Hz, +100 Hz の音声の評価が有意に低下した. この結果から, 男声は, これらの変化量で好感度が下がる可能性を示唆している. 一方で, 女性 1 の +100 Hz の音声の評価で有意差が見られなかったが, 評価値が 0 よりも高かった. これは, 元音声の F1 が大きいと変化率が小さくなり, +100 Hz の変換でも大きく変化しなかったためだと仮定した. 仮定を検証するため, 音声分析ソフト Praat[*1] を用いて各発話者の F1 を分析した. 分析の際, 音声を 8000 Hz にダウンサンプリングした. Julius[13] を用いて, 各母音の区間推定を行った後, 推定された区間内の F1 の値の中央値を抽出した. 母音毎に抽出された値を平均し, 分析結果に用いた. なお, 抽出した F1 のうち, 明らかな推定ミスと判断できる結果を目視で確認し, 除外した. 求めた発話者別の F1 の平均値を図 6 に示す. この図により, 女性 1 の F1 が実験音声の中で平均的に高いことが示された. そのため, 元音声の F1 が高いと, +100Hz の変換でも大きく変化せず, 評価が有意に低下しない可能性を示した. また, 男声 1 の変化量+50 Hz の音声の評価は, 評価が上がる有意傾向であることが認められたが, 他の音声では有意傾向が認められなかった. これは, 音声によって評価が最も高くなる変化量が異なるためだと推測する.

6. おわりに

本研究では, 好感度改善を目的とした音声加工技術の開発を行った. 関連研究 [4] の実験結果より, 評価と相関が見られた音響特徴量から提案手法で変換する音響特徴量の選定を行い, F1 の変換を試みた. F1 のみを変換させる音声変換は, 出力する音声の品質劣化が起こることが予想されたため, 品質劣化を抑えた F1 を含むスペクトル包絡全体を変化させる変換手法を提案した. その後, 提案手法の有用性と自然性を確認するための音声評価実験を行った. 実験の結果, −50 Hz, +100 Hz の音声は, 0 Hz の音声と比べて好感度が有意に下がる傾向がみられた. +50 Hz は, 0 Hz の音声と比べて好感度は高かったが, 有意差は認められなかった. 一方, 一部の音声では, 有意傾向がみられることが示唆された. また, −50 Hz 以下, +100 Hz 以上の変換を行うことで音声が品質劣化を起こすことを確認した.

今後の課題としては, 提案手法によって好感度が改善される音声の傾向の調査を行うことである. 本研究では, 6 名の音声を用いたが, より多くの人の音声を用いることで, どのような特徴の音声が提案手法によって好感度が改善できるのか検証する必要がある. また, F0 の変化との相互作用がみられるか検証を行うことで, より良い印象の音声を合成することができると考えられる.

[*1] Praat : (http://www.fon.hum.uva.nl/praat/), 2018-4-27.

謝辞 本研究は, JSPS 科研費 JP16H05899, JP16H01734 の支援を受けて実施された.

参考文献

[1] 河井麻里子, 望月聡 : 対人印象形成における声的魅力とパーソナリティとの関係, 日心第 77 回大会, p. 655, (2013).
[2] B. C. Jones, D., R., Feinberg, L., M., DeBruine, A., C., Littele, and J. Vukovic,: Integrating cues of social interest and voice pitch in men's preferences for women's voices, Biology letters, Vol. 4, No. 2, pp. 192-194, (2008).
[3] 吉元照貴, 西村竜一, 入野俊夫, 河原英紀 : 基本周波数操作による音声の好感度改善に関連する物理的特徴の検討, 日本音響学会 2013 年秋季研究発表会講演論文集, pp. 335-336, (2013).
[4] 横森文哉, 二宮大和, 森勢将雅, 田中章浩, 小澤賢司 : 好感度評価の性差に着目した女性発話の音響特徴量分析, 日本感性工学会論文誌, Vol. 15, No. 7, pp. 721-729, (2016).
[5] E. Schubert, and J. Wolfe,: Does timbral brightness scale with frequency and spectral centroid?, Acta Acustica United with Acustica, Vol. 92, pp. 820-825, (2006).
[6] M. Morise, F. Yokomori, and K. Ozawa,: WORLD: a vocoder-based high-quality speech synthesis system for real-time applications, IEICE transactions on information and systems, Vol. E99-D, No. 7, pp. 1877-1884, (2016).
[7] M. Morise,: D4C, a band-aperiodicity estimator for high-quality speech synthesis, Speech Communication, Vol. 84, pp. 57-65, (2016).
[8] 内田照久 : 未知のイントネーションから想起される話者の性格印象と方言地域の特徴, 音声研究, Vol. 10, No. 3, pp. 29-42, (2006).
[9] 廣谷定男 : 母音のフォルマント分析-過程と仮定を知る-, 日本音響学会誌, Vol. 20, No. 1, pp. 41-49, (2005).
[10] R. Jurgen, K. Hammerschmidt, and J. Fischer,: Authentic and play-acted vocal emotion expressions reveal acoustic differences, Frontiers in Psychology, Vol. 2, pp. 1-11, (2011).
[11] 平井俊男, 河井恒, 津崎実, 戸田智基 : 音声合成システム XIMERA における日本語合成音の自然性劣化要因の分析, 日本音響学会誌, Vol. 62, No. 11, pp. 767-773, (2006).
[12] 横森文哉 : 好感度を対象とした音声波形の音響特徴量および言語情報の分析, 山梨大学大学院コンピュータ・メディア工学専攻修士論文発表会, (2016).
[13] 河原達也, 李晃伸 : 連続音声認識ソフトウェア julius, 人工知能学会誌, Vol. 20, No. 1, pp. 41-49, (2005).

© 2018 Information Processing Society of Japan