

音声の好感度改善補助ツールの開発を 目指した好感度改善方法の検討

吉元 照貴^{1,a)} 西村 竜一^{1,b)} 入野 俊夫^{1,c)} 河原 英紀^{1,d)}

概要: 本稿では、音声の好感度改善トレーニング補助を目的とした簡易ツールの紹介と、音声の好感度改善の具体的な方法について述べる。これまで、音声から受ける印象の一つである好感度と音声の物理属性との関係を、音声モーフィングの技術を用いて調べてきた。それらの検討から、(1) 音声の物理属性の中では、基本周波数やスペクトル形状が好感度へ与える影響が大きいこと、(2) 話者の音声から聞き手が受ける好感度は、聞き手によって大きく異なっていること、(3) 聞き手が好感度を改善するように音声を操作する際に、操作された音声の話者性が変化して感じられると好感度の判断が大きく影響されることが示唆された。今回報告する好感度改善手法では、これらの知見を考慮し、音声パラメタの基本周波数とスペクトル形状を操作することで好感度の改善を図る。また、好感度を改善する操作の妨害要因となっていた話者性が変化する問題を解決するために、演劇部の学生が好感度の異なる話し方で演技した音声から求められる音声の物理属性の変化量を求めた。この変化量を好感度が低い他者の音声の物理属性の操作に用いた。ここでは、変化量を抽出した話者と聞き手および操作対象となった話者の組み合わせについて好感度の改善の効果を調べた。これらの結果に基づいて、提案する簡易ツールの概要を説明する。

キーワード: 主観評価, 音声モーフィング, 声道断面積関数, 声道長比, 基本周波数

Improving voice attractiveness by speech parameter modification for interactive voice training applications

YOSHIMOTO SHOKI^{1,a)} RYUICHI NISIMURA^{1,b)} TOSHIO IRINO^{1,c)} HIDEKI KAWAHARA^{1,d)}

Abstract: A simple voice training system for improving attractiveness is introduced with descriptions on a set of procedures which consist of the system. Those procedures are based on findings drawn from our investigations on voice attractiveness using a new voice morphing method. They are summarized as follows. a) Most contributing physical factors on attractiveness are fundamental frequency and spectral information. b) Attractiveness judgement differs among listeners. c) Change in perceived talker of the modified voice caused by physical parameter manipulation for improving voice attractiveness is disturbing for listener's judgement and adjustment. To overcome the last disturbing factor, physical parameters change within each talker for improving attractiveness is acquired recruiting student actors in our university. Several sets of physical parameters change are applied to improve attractiveness of voices with lower attractiveness score. Attractiveness of the modified voices using these sets of physical parameters change were tested for all possible combinations of the source actor, talkers of manipulated voices and the listeners. The proposed voice training system is introduced based on the results of tests.

Keywords: subjective test, voice morphing, vocal tract area function, vocal tract length ratio, fundamental frequency

¹ 和歌山大学
Wakayama University, Wakayama 640-8510, Japan
a) s140137@center.wakayama-u.ac.jp
b) nisimura@sys.wakayama-u.ac.jp

c) irino@sys.wakayama-u.ac.jp
d) kawahara@sys.wakayama-u.ac.jp

1. はじめに

聞き手の好みに合うように、声の好感度を改善させるトレーニングを補助するツールの開発を検討している。これまでの検討では、話者の好感度は聞き手の好みに大きく依存することと [1]、好感度が改善されるように変換された音声の聞き手による好感度の評価は、話者性の判断に大きく影響されることが示唆されていた [2]。また、単純な基本周波数操作などの音声パラメタの操作では、好感度改善が困難であることも示唆されていた [3]。これらの検討を背景に、時変多属性任意事例数モーフィング [4], [5], [6] を用いた実験により、音声の好感度は、基本周波数やスペクトル形状に大きく影響されることが示唆された [2]。そこで、これらの音声パラメタを操作し、好感度の改善を図り、聞き手の好み合うような音声を作成するシステムを提案する。しかし、スペクトル形状そのものは操作の自由度が高く、好感度が向上するような操作方法を直接探索することは難しい。

今回紹介する簡易ツールでは、声道断面積関数 [7] と相対的な声道長 [8] を介してスペクトル形状を操作する。このパラメタの操作量の目安を得るための準備として、話者が好感度を向上させるために採っている戦略を、演劇部の学生の協力を得て好感度改善用テンプレートとして抽出した。実際の音声の変換では、好感度を向上させる対象となる音声と聞き手の組合せに応じて、テンプレートを選択して適用する。予備的に、録音した音声の音高-声道長パターンが類似しているテンプレートを選択し、適用するシステムを実装した。本稿では、著者らが提案する音声パラメタ操作方法によって好感度を改善させて、ボイストレーニングを補助する簡易システムについて紹介する。

2. システムの概要

本ツールは、利用者の声の好感度を改善させるトレーニングを補助することを目的としている。実際に、利用者が発声した音声を録音し、音声から求めることができる音高-声道長パターンを用いて好感度を改善させるための音声パラメタ操作方法を選択し、元の音声に適用後、利用者にフィードバックすることでボイストレーニングを補助する。また、本ツールでは、声道断面積関数を利用して、声道形状の近似的な等価音響管の 3D モデルを GUI 上に表示し、可視的なフィードバックも与えることができるように実装した。

2.1 想定する利用状況と GUI の振る舞い

本ツールは、自身の声の好感度を向上させたい利用者を想定している。そのため、何度も繰り返してトレーニングができるようなシステムの構成を考えた。ツールの概観を

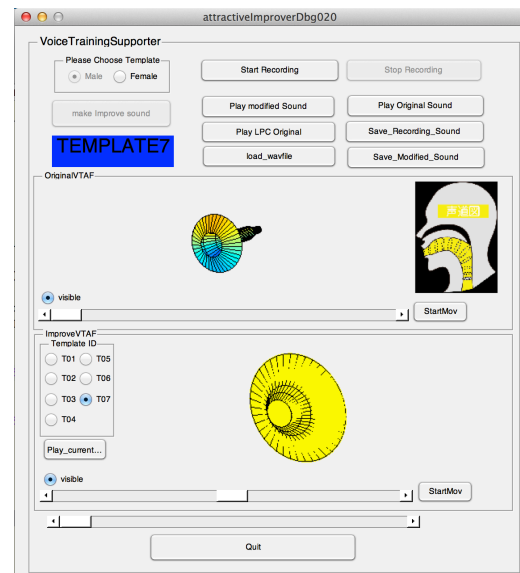


図 1 GUI of the proposed interactive voice training system for improving voice attractiveness. Upper 3D model represents the estimated user's (trainee's) vocal tract shape. The lower 3D model represents the original shape (wire frame) and the modified shape (colour surface) by applying one conversion template by selecting left panel with radio buttons.

図 1 に示す。利用者は自身の普段の声を [Start Recording] ボタンを押下した後に収録し、計算機内に音声を取り込む。発声が終了したら、[Stop Recording] ボタンを押下し、録音を終了する。録音終了後、[make improve sound] ボタンを押下して、録音した音声に好感度を改善させる処理を施す。処理が終了すると、図 1 中央付近に 2 つの等価音響管の 3D モデルが表示される。この 3D モデルは音声から算出される声道断面積関数をもとに作成した。[Play modified sound] ボタンを押下することで、好感度改善処理が施された音声再生され、音声のフレームと同期して等価音響管の 3D モデルが変形する。

2.2 システムの構成

システムは GUI に加えて、以下のサブシステムで構成される。

- TANDEM-STRAIGHT [9] による音声パラメタの分析 (基本周波数抽出)
- 音声波形に基づく声道長推定 [8], 声道断面積関数の抽出
- 音高-声道長パターンの類似性に基づく好感度改善用テンプレートの選択と適用

ここでは、利用者の音高-声道長パターンと好感度改善用テンプレート作成の元となる音声試料の音高-声道長パターンのユークリッド距離にもとづいて、距離の小さい方が類似性が高くなるように定義している。その後、後述する好感度改善方法として用意した好感度改善用テンプレートを音

表 1 演劇部学生による異なる好感度の話し方での音声収録条件
Table 1 Recording conditions to acquire student actors' skills for making their voices attractive

Number of speakers	7 male and 3 female student actors
Sentence	40 sentences
Instructions	Speak attractively Speak, not attractively
Digitization	48kHz, 24bit
Microphone	DPA 4006A
Audio interface	ROLAND UA-101
Distance to the microphone	30cm
Environment	Sound proof room (YAMAHA AVITEX)

声に適用し、好感度改善処理後に TANDEM-STRAIGHT を用いて音声合成し、合成音声と音声に対応した声道断面積関数の 3D モデルを提示する。利用者は改善処理前の音声と処理後の音声を何度でも聴き比べることができ、フィードバックを参考にボイストレーニングが行える。

3. 好感度改善用テンプレート

これまでの検討では、違う好感度の異なる話者の音声の違いを、分析結果として得られる物理属性の変化量として求め、任意の音声の好感度改善に用いていた。これらの処理では、話者の個人性の変化により、狙った好感度の改善効果が得られないという問題が生じていた。ここでは、同一の話者が異なった好感度を与えることを意図して発声した音声を用いることにより、話者性の変化を引き起こさない処理の実現を狙う。また、スペクトル包絡をそのまま用いるのではなく、声道長比と対数断面積関数の差として操作を表現することとした。この操作内容をまとめたものを、好感度改善用テンプレートと呼ぶこととする。本節では、この好感度改善用テンプレートの概要について説明する。

3.1 好感度改善用テンプレートの作成

同一の話者が発声した好感度の異なる音声を収録するために、ここでは演劇部に所属する学生男女 10 名 (男性 7 名、女性 3 名、年齢は 18~21 歳に分布) を話者として選択した。こうして選択した話者に、日常会話での挨拶などを含む 40 個の文章を、話者それぞれの解釈に基づいて、低い好感度の話し方と、高い好感度の話し方で話してもらい収録した。音声は防音室 (YAMAHA AVITEX) で圧力型コンデンサマイク (DPA 4006A) を用い、オーディオインタフェース (ROLAND UA-101) を介して MacBookPro 15" のプログラム (Audacity) により、24 bit 48000 Hz の WAVE 形式の音声ファイルとして記録した。詳細を表 1 に示す。今回は、これらの音声の中から、キャリアセンテンスに埋め込まれた CV 単音節 (具体的には子音 C に /h/、母音 V に、日本語五母音) の母音部を切り出し、TANDEM-STRAIGHT[9]

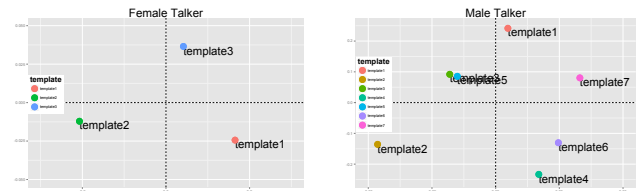


図 2 Two dimensional template allocation by MDS based on Euclidean distance between template vectors. (Left plot: templates derived from female actors, Right plot: from male actors)

により分析し、好感度の高い音声と好感度の低い音声について、LPC を介して求めた断面積関数の比と声道長比、基本周波数の比をまとめたものをテンプレートとして作成した。図 2 に、テンプレートそれぞれに含まれる操作量から距離行列を作成し、多次元尺度構成法を利用して二次元の付置を求めた結果を示す。

左が女性話者、右が男性話者によるテンプレートから求められた付置である。また、平均値を除いたテンプレート間の相関の絶対値について求めた。男女別の平均値を除いたテンプレート間の相関の絶対値を付録 A.1 節の表 A.1 に示す。テンプレート間の相関の絶対値の平均値は男性話者が 0.62、女性話者が 0.70 であり、それぞれの話者が、好感度の改善に異なる戦略を用いていることが示唆された。

4. テンプレートによる好感度改善の評価

好感度改善用テンプレートの適用による好感度改善効果を、音質評価方法である MUSHRA[10] を応用した主観評価実験により評価した。

4.1 評価用音声試料の選定

好感度改善用テンプレートを適用する音声を、男女幅広い年齢層の東京方言話者の収録音声で「はー、ひー、ふー、へー、ほー」と編集された音声を収録している母音データベース [11] から選定した。好感度の改善の程度を評価する被験者によって、18 歳以上の男性 20 名、女性 22 名分の音声の好ましさを性別ごとに 100 点満点で点数をつけてもらい順位を決定した。被験者毎に集計した各話者に対する好ましさの点数の平均値で昇順に左から並び替えたものを図 3 に示す。また、これらの平均値間での Kruscal-Willis の検定では、男性話者に対する要因は有意であった ($p = 2.1 \times 10^{-5}$) が、女性話者に対する要因は有意でなかった ($p = 7.6 \times 10^{-1}$)。

この点数の低い男女各 3 名の音声にテンプレートを適用した。また、話者の好感度は聞き手に大きく依存することを、評価用音声試料選定の際に用いた各被験者の点数によって、好みの非類似度を求め、多次元尺度構成法により二次元の付置により再度確認した。付置の結果を図 4 に示す。

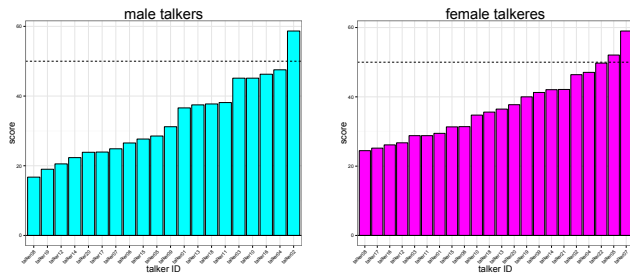


図 3 Attractiveness of each talker in a speech database. Attractiveness scores were evaluated by 5 subjects (3 male and 2 female) and averaged. Left plot shows male talkers and right plot shows female talkers.

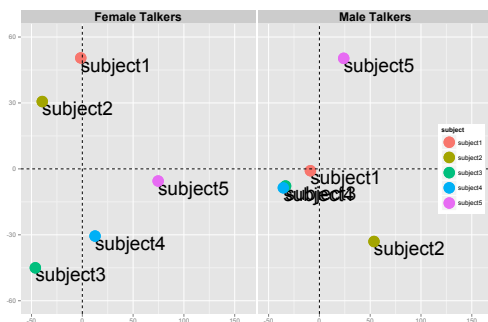


図 4 Two dimensional listener (subject) allocation by MDS based on Euclidean distance between evaluation score records of subjects. (Left plot: female voice modification, Right plot: male voice modification)

男声を評価するにあたり、被験者 3 と被験者 4 が同じような好みを持っている可能性が示唆されたが、被験者毎に多様な好みを持っている可能性が高く、以前の報告と内容が整合していることを確認できた。

4.2 実験結果

好感度の悪い男声 1 名と 7 つのテンプレートを適用して作成した音声を被験者 1 名が評価した結果を図 5 に示す。本評価では、従来、音質評価方法として用いられている MUSHRA 法 [10] を参考に、音質を評価する代わりに好ましさについて 0~100 で点数をつけることで評価した。ここでは、1 名の被験者が 1 名分の話者の音声の好感度改善効果を 3 回評価している。そのため、図 5 の縦軸は、評価時のスコアの平均値を表している。横軸は、適用したテンプレートの種類を表しており、“ORIGINAL”は、テンプレートを適用していないことを表す。他の好感度改善対象の音声、また、他の被験者が評価する場合にも同様に 3 回評価を実施した。付録 A.2 に残りの被験者の結果を示す。具体例として、1 名の被験者の全話者に対する 3 回分の点数の平均値のスコアを図 6 に示す。図中、エラーバーは 95%信頼区間を表す。図 6 から男声 1 名に対する評価を抜き出した図 5 の “ORIGINAL” のラベルに対応するバーは、元の音声自身と好ましさを比較したことを表して

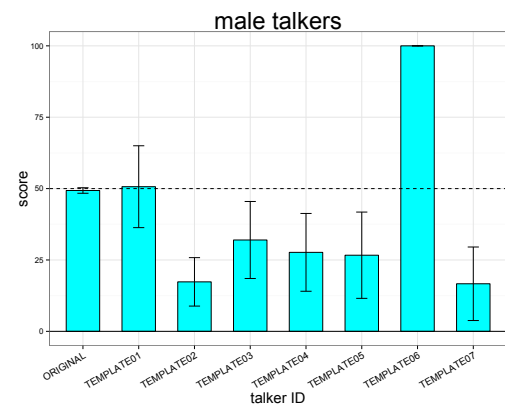


図 5 Subjective evaluation results of attractiveness. Horizontal axis represents types of modification and vertical axis represents attractiveness obtained by MUSHRA [10] procedure. Attached vertical lines represent 95% confidence intervals. (This is an excerpt from Fig. 6)

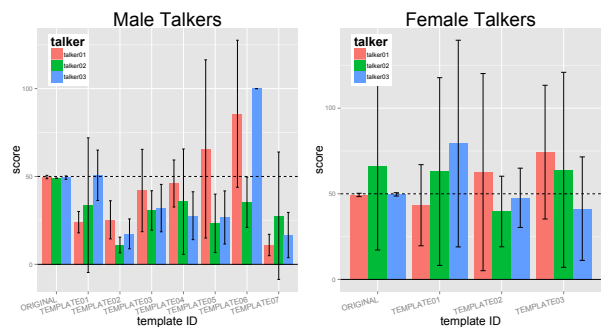


図 6 Whole subjective evaluation results of attractiveness by one listener. Each bar for each modification represents talker (voice). In this case, three least attractive voices are used. Left plot shows male voices and right plot shows female voices.

いる。この場合、平均値が 50 点で、好感度が変化していないことを示している。一方、6 番目のテンプレートで作成した音声を表す TEMPLATE6 のバーは平均値が 100 点となり好感度の改善が確認された。

5. まとめ

今回の報告では、好感度改善のためのボイストレーニングを補助する簡易的なツールについて紹介した。ツールでは、同一話者の好感度の違いを利用した音声好感度改善方法である好感度改善用テンプレートの応用への可能性を示した。好感度改善用テンプレートの好感度改善効果を確認する主観評価実験の結果、好感度を改善できる可能性が示唆された。また、好感度を向上させる対象となる音声と音声の聞き手の組合せに適したテンプレートは複数存在し、音声好感度改善方法も幾通りが存在することが示唆された。今回紹介した簡易ツールの実用性の調査と、今回利用した好感度改善方法の自動化に向けた指標の検討が今後の課題となる。

謝辞 本研究の一部は、科学研究費基盤 (B)24300073 および萌芽 26540087 による。

参考文献

- [1] 吉元照貴, 西村竜一, 入野俊夫, 河原英紀: 外挿が可能な時変多属性任意事例数モーフィングを用いた文章音声好感度の改善について, 日本音響学会研究発表会講演論文集日本音響学会 編, Vol. 2014, pp. 429–430 (2014).
- [2] 吉元照貴, 伊佐衣代, 溝淵翔平, 西村竜一, 入野俊夫, 河原英紀: 声道形状と音源情報に着目した音声の好感度改善システムの検討について, 日本音響学会研究発表会講演論文集日本音響学会 編, Vol. 2014, pp. 373–374 (2014).
- [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, 日本音響学会研究発表会講演論文集日本音響学会 編, Vol. 2013, pp. 335–336 (2013).
- [4] Kawahara, H., Morise, M., Banno, H. and Skuk, V. G.: Temporally variable multi-aspect N-way morphing based on interference-free speech representations, *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, IEEE, pp. 1–10 (2013).
- [5] 河原英紀, 森勢将雅, 坂野秀樹: 異なった時間軸を持つ複数の歌声の実時間モーフィングについて, 情報処理学会研究報告.[音楽情報科学], Vol. 2013, No. 25, pp. 1–5 (2013).
- [6] 河原英紀, 森勢将雅, 坂野秀樹: 外挿可能な時変多属性任意事例モーフィングにおける誇張の可能性について, 日本音響学会研究発表会講演論文集日本音響学会 編, pp. 555–558 (2014).
- [7] Hisashi, W.: Direct estimation of the vocaltract shape by inverse filtering of acoustic speechwaveforms, *IEEE Trans. Audio, Electroacoust.*, Vol. AU-21, No. 5, pp. 417–427 (1973).
- [8] 小林真優子, 坂口諒, 西村竜一, 入野俊夫, 河原英紀: 日本語母音データベースを用いた声道長推定法の校正について, 日本音響学会研究発表会講演論文集日本音響学会 編, Vol. 2014, pp. 305–306 (2014).
- [9] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: TANDEM-STRAIGHT: A Temporally Stable Power Spectral Representation for Periodic Signals and Applications to Interference-free Spectrum, F0, and Aperiodicity Estimation, *ICASSP2008*, pp. 3933–3936 (2008).
- [10] Recommendation, I.: BS. 1534-1. Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA), *International Telecommunication Union* (2003).
- [11] Ohyama, G., Deguchi, T. and Kasuya, H.: Construction of Japanese vowel database uttered by native speakers over a wide range of age, *Proc. Spring Meeting of ASJ* (2011).

付 録

A.1 音声パラメタ操作量のユークリッド距離による好感度改善テンプレート間の類似性

平均値を除いたテンプレート間の相関の絶対値を、男女別に表 A.1 に示す。各テンプレートの識別にテンプレートの番号と頭に “T” を付加して対応させた。

表 A.1 平均値を除いたテンプレート間の相関の絶対値
 Table A.1 Absolute values of correlation between templates

男声テンプレート						
	T02	T03	T04	T05	T06	T07
T01	0.36	0.70	0.36	0.69	0.29	0.98
T02		0.92	0.73	0.93	0.78	0.56
T03			0.40	0.99	0.47	0.85
T04				0.43	0.99	0.15
T05					0.50	0.83
T06						0.07

女声テンプレート		
	T02	T03
T01	0.41	0.81
T02		0.87

A.2 好感度改善用テンプレート適用音声の好感度改善評価

本文掲載以外の被験者の実験結果を以下に示す。

被験者 2

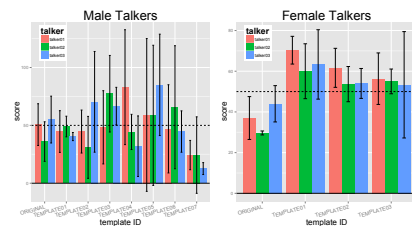


図 A.1 Whole subjective evaluation results of attractiveness by one listener.

被験者 3

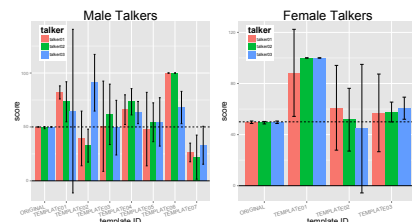


図 A.2 Whole subjective evaluation results of attractiveness by one listener.

被験者 4

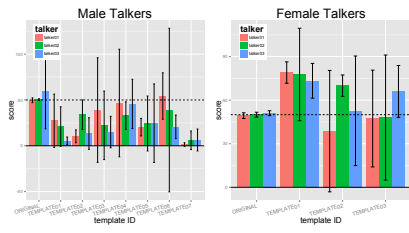


図 A.3 Whole subjective evaluation results of attractiveness by one listener.

被験者 5

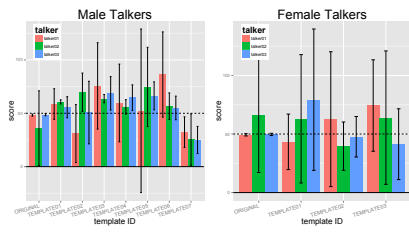


図 A.4 Whole subjective evaluation results of attractiveness by one listener.