

合成音声の作成

システム科学技術学部 情報工学科

1年 根本 桐李

1年 川口 瞬椰

1年 今村 彩人

1年 佐々木 友愷

指導教員 システム科学技術学部 情報工学科

准教授 渡邊 貫治

1 背景

近年、音声ガイダンスや案内放送など、合成音声技術は日常生活に広く浸透している。その一因として、合成音声を利用することは声優を雇用する場合と比較して、手間や費用を大幅に削減できることが考えられる。しかし、現在一般的に用いられている合成音声はいずれも似通った声質であり、区別が付きにくいという問題点が存在している。実用性の観点からは、合成音声は聞き取りやすさを損なうべきではないが、その人間らしさの範疇の中で「個性」を与えることが出来れば、合成音声を利用したサービスの多様化が見込まれると思われる。そこで本研究では、人間が魅力的に感じやすい「楽器音」に着目し、その要素を取り入れることで個性を持った合成音声の実現を目指した。

2 音声の合成方法及び評価実験方法

2.1 概要

本研究では、MATLAB上で動作する音声分析合成システム「WORLD」[1]を利用して、人間の声に楽器音の要素を付与した。まず、人間の声及び楽器音のデータを用意し、それぞれの音声から基本周波数、スペクトル包絡といった特徴量を抽出する。次に、抽出した楽器音のパラメータを人間の声に適用し再合成を行うことで、音質や個性にどのような変化が生じるかを検証する。

2.2 合成方法

- ① まず、人間の声のデータとして、「WORLD」に付属する「あいうえお」と発話した男声のサンプルデータおよび協力者に依頼して録音した女声の2種類を使用した。また、楽器音データとしてはRWC研究用音楽データベース[2]のものを利用した。本研究で使用した人間の音声は短いため、合成における楽器音の時間をそれに合わせて、適宜10秒程度の単音(打楽器等、例外を除く)に編集して使用した。
- ② 「WORLD」を用いて、人間の声及び楽器音から特徴量を抽出し、それぞれのデータを保存した。
- ③ 抽出した楽器音のスペクトル包絡を人間の声の特徴量に適用し、「WORLD」を用いて再合成を行った。これにより、楽器音の要素を持った音声を生成した。

2.3 楽器音一覧

楽器音一覧を表1に示す。

表1 楽器音の一覧

楽器種別	総数	楽器名
弦楽器	12	アコースティックギター・ヴァイオリン・ヴィオラ・ウクレレ エレキギター・エレキベース・クラシックギター・コントラバス・チェロ・ハープ・琴・笙
管楽器	6	アルトサックス・ソプラノサックス・ホルン・チューバ・トランペット・トロンボーン
打楽器	4	ティンパニ・ドラムロール・ビブラフォン・マリンバ
気鳴楽器	4	アコーディオン・ハーモニカ・パイプオルガン・ホイッスル
鍵盤楽器	3	エレクトリックピアノ・チェンバロ・ピアノ

2.4 合成音声の特徴と変化の傾向

作成した合成音声を試聴したところ、明らかに元の音声とは異なる音声が生産されることを確認できた。しかし、本研究の合成方法では元の人間の声の要素が程度の差はあれ残っており、完全に楽器音の特徴に置き換わることはなかった。一方で、弦楽器の揺らぎなど楽器音内で音程が変化している場合には、その変化が合成後の音声にも反映されているように見受けられた。しかし、女声ベースの音声では元の発話の音程が残っている傾向が高かった。

2.5 評価方法

完成した音声を、楽器名を伏せた状態でランダムに再生し、「聞き取りやすさ（1：聞き取りにくい～5：聞き取りやすい）」「抑揚（1：抑揚がない～5：抑揚がある）」「（元の音声からの）変化度合い（1：変化がない～5：変化がある）」の3つの項目をそれぞれ5段階評価で主観評価を実施した。なお、評価対象者には元の音声を最初に提示し、合成後の音声との比較が可能な状態とした。

評価は、秋田県立大学システム科学技術学部の1年生8名（共同研究者を除く）を対象に、紙媒体のアンケート形式で実施した。音声の再生にはノートPCに搭載された内臓スピーカーを使用した。評価結果をもとに、楽器音の要素を取り入れた音声合成の効果や実用性、音声合成に適した楽器の傾向などについて考察する。

3 結果

3.1 主観評価結果

評価結果に基づき、各評価軸間の傾向や関連性を視覚的に把握するため、各合成音声のスコアを3つの観点で比較した。図1は「聞き取りやすさ」と「変化度合い」、図2は「変化度合い」と「抑揚」、図3は「抑揚」と「聞き取りやすさ」の関係を2次元グラフで可視化したものである。

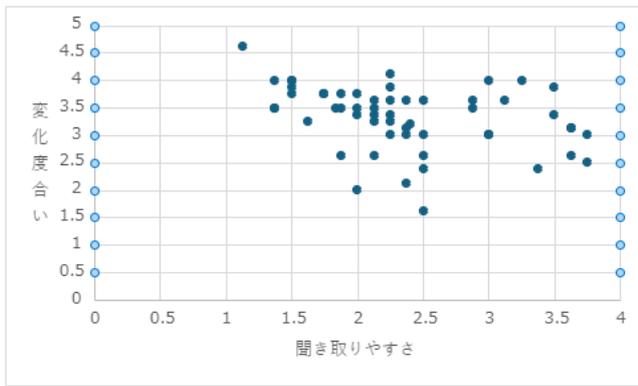


図1 聞き取りやすさと変化度合い

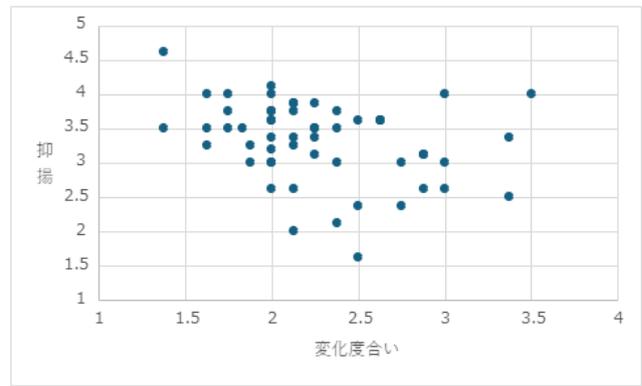


図2 変化度合いと抑揚

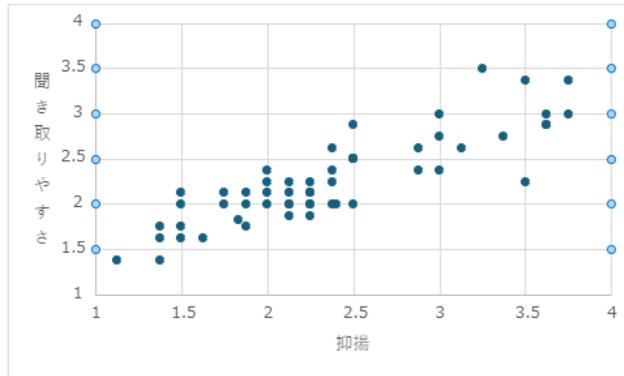


図3 抑揚と聞き取りやすさ

図3から、抑揚のスコアと聞き取りやすさのスコアの間にはおおまかな傾向として抑揚が大きくなるほど聞き取りやすさも大きくなるのが分かる。

楽器音の種別（弦楽器、管楽器、打楽器、気鳴楽器、鍵盤楽器の5種）に関しては、主観評価のスコアに明確な傾向はみられず、どの種類の楽器にも合計スコアで高スコア・低スコアのものが存在した。（表2および表3）一方で、性別による違いを分析したところ、男声よりも女声の方が総じて低いスコアを示す傾向が見られた。最も高いスコアを示した組み合わせは「男声ベース／トランペット」の10.75点であり、最も低かったのは「女声ベース／ホルン」の6.125点であった。

表2 高スコアの楽器名と各値

	楽器名	性別	聞き取りやすさ	抑揚	変化度合い	合計
1	トランペット	男	3.25	3.5	4	10.75
2	ヴィオラ	男	3.5	3.38	3.38	10.25
3	トロンボーン	男	3	3	4	10
4	ビブラフォン	男	3.75	3	3	9.75
5	エレキギター	男	3.5	2.25	3.88	9.625

表3 低スコアの楽器名と各値

	楽器名	性別	聞き取りやすさ	抑揚	変化度合い	合計
58	ホルン	女	2	2.13	2	6.13
57	アコギ	女	1.38	1.38	3.5	6.25
56	ソプラノサックス	女	1.63	1.63	3.25	6.5
55	ティンパニ	女	1.38	1.63	3.5	6.5
54	エレクトリックピアノ	女	1.88	2.13	2.63	6.63

4 考察

4.1 楽器の種類による影響

楽器の種類によって、合成後の音声の明瞭さや評価に違いが見られた。特に打楽器は女声の場合、

ティンパニやビブラフォンでは爆発音（極端な音割れやノイズ）が発生したが、男声では元の発話を損なわない合成が出来ていた（表4）。

表4 男女における打楽器でのスコア差

楽器名	性別	聞き取りやすさ	抑揚	変化度合い	合計
ティンパニ	男	3.13	2.63	3.63	9.38
ティンパニ	女	1.38	1.63	3.50	6.50
ビブラフォン	男	3.75	3.00	3.00	9.75
ビブラフォン	女	1.38	1.75	4.00	7.13

打楽器の特徴として、弦楽器や人間の声などから発生する倍音とは異なるメカニズムで、打撃時に発生する低い音と、その後の共鳴によって持続する高い音が同時に含まれている。このような構造を持つ音を合成に用いた際、性別によって反映される成分に違いが生じた可能性がある。その理由として、合成の際に利用されている特徴量である基本周波数 (F0) が 女声の方が高いため打楽器の高音成分と干渉しやすく、共鳴成分を強く拾ってしまうことにより引き起こされていると考えられる。以上のことから、合成元の性別により、合成に向いている楽器とそうでない楽器が存在すると考えられ、男声では打撃時の低い音（一番大きく鳴る音）が主に合成に影響を与え、比較的安定した合成が行われているのに対し、女声では打撃後の共鳴音、特に高音成分が強く影響し、それが不安定な変換や爆発音を引き起こしている可能性がある。

また、打楽器に限らず男声ベースの合成音声よりも女声ベースの方のスコアが低い傾向に関しては、男声のデータがWORLDに付属の実験用に録音されたものであったのに対し、女声の録音環境が整っておらず、「音量が小さい」「ノイズが入っている」といった問題が影響した可能性があり、正確な合成・比較が出来ているとは考えづらい。

4.2 音の高さによる影響

試聴した結果として、合成元の楽器音声を中・高音域から抽出した楽器（エレキギター、チェロ）は比較的自然的な合成がされやすく、抑揚がつきやすいのに対して、低音は合成が難しく、極端に低い音（エレクトリックピアノ、エレキベース）では爆発音が発生する傾向が見られた。これは、特に低音域の楽器音では基本周波数 (F0) が低いため1周期あたりの波形の長さが長くなり、合成に利用されるスペクトル包絡に含まれる周波数成分が相対的に少なくなった結果、聴覚上粗く聞こえる音質になると考えられる。一方で高音では1周期あたりの振動数が多くサンプル数を十分に確保でき、合成後の音声の違和感が少なくなるのではないかと考えられる。楽器音による音声合成を実用化する場合は、合成元の声と楽器音の基本周波数の適合性を考慮し、適用すべき音域の最適化を行う必要がある。

5 まとめ

楽器音のスペクトル包絡を人間の声に適用することで、合成音声に個性を与える手法を検討した結果、楽器音の種類や音の高さによって、合成のしやすさや音質に大きな違いがあることが確認された。従来の合成音声が抱える「声質が似通い区別がつきにくい」という課題に対して、一定の改善効果が得られたと考えられる。一方で、聞き取りやすさや抑揚といった実用面で課題の残る音声も多く、実際のアプリケーション等での利用を想定した場合、合成精度や品質の面で改善の余地がある。多様な個性を持つ合成音声を作成するためには、本研究で上手く合成できなかった楽器音に対しても適用可能な手法を探る必要があるという結果となった。一方で、合成に適した楽器音の特徴や、合成後の音声の楽器音の特性から受ける影響についての理解が深まったため、今後はこれらの知見を活用することで、より効果的な音声合成の手法を実現できると考えている。

6 出典

- [1] <https://www.isc.meiji.ac.jp/~mmorise/world/introductions.html>
- [2] <https://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-i-j.html>