

歌唱熱唱度の言語依存性の調査

郝沢宇^{†1} 能勢隆^{†1} 伊藤彰則^{†1}

歌唱熱唱度は歌声評価のための新しい評価指標である。この指標は日本語歌唱を対象として開発されたが、これが他の言語による歌唱にも有効なのか、あるいは他の言語を母語とする聴取者の熱唱度知覚と整合するのかがどうかについては検討を要する。本研究では、日本語母語話者と中国語母語話者を対象に、同一曲の日本語と中国語バージョンを使って熱唱度知覚実験を行った。その結果、日本語母語話者と中国語母語話者の熱唱度知覚はおおむね一致したが、日本語を専攻として学習している中国語母語話者はそれ以外と異なる特徴を示した。機械的な熱唱度評価は日本語歌唱・日本語母語話者の知覚との相関が高く、中国語歌唱・中国語母語話者の知覚とは相関が低かった。

Investigation of Language Dependence of Singing Enthusiasm

ZEYU HAO^{†1} TAKASHI NOSE^{†1} AKINORI ITO^{†1}

Singing enthusiasm is a new index of evaluating singing voice proposed by Daido et al. This index was developed by modeling Japanese listeners' perception of Japanese singing voice; here, there is a possibility that perception and/or calculation of singing enthusiasm depends on the language of the singing voice or the listeners' mother tongue. In this paper, we investigated dependency of perception of singing enthusiasm on the language of the singing voice and the linguistic background of the listeners. We recorded two songs having the same melody and different lyrics in Japanese and Chinese, and employed Japanese and Chinese native speakers to evaluate the enthusiasm of the singing voice. As a result, perception of the singing voices in the two languages was almost consistent by the listeners of the both native speakers, but the Chinese native speakers learning Japanese as their major showed different kind of perception. Besides, we investigated consistency between the human perception and automatic estimation of singing enthusiasm for both languages. As a result, correlation of the subjective and automatically-estimated singing enthusiasm values was high for Japanese evaluators listening Japanese songs, but the correlation was lower when evaluating Chinese singing voices or comparing to the evaluation by Chinese native speakers.

1. はじめに

歌声における様々な要素の知覚とそのモデル化は古くから行われてきている。たとえばクラシック音楽における歌唱と物理量の関係の調査 1)、歌の才能と声質の関係の調査 2)、幼児向けの歌声の特徴 3)、話声と歌声の特徴 4)、非プロフェッショナルによる歌のうまさ 5)6)、歌声の「表現」と自然さ 7)などである。我々のグループでは、歌唱のうまさではなく「一生懸命さ」をあらわす評価指標である「熱唱度」という概念を提唱し、多くの聴取者について熱唱度の知覚に一貫性があること、歌唱音声の物理量から熱唱度を推定することができることなどを示した 8)。

文献 8)では熱唱度が知覚的に実在することを示したものの、歌唱者・評価者はすべて日本語母語話者であり、評価に用いた歌もすべて日本語であった。それでは、日本語以外を母語とする話者が日本語の歌を聞いた場合、あるいは日本語母語話者が日本語以外の歌を聞いた場合はどうなるだろうか？

聴取者の文化的背景と音楽の知覚の関係に関する研究はいくつかある。Balkwill と Thompson は、西洋音楽のバックグラウンドを持つ聴取者にインドのヒンドウスタニ音楽を聴かせ、その音楽が表わす感情を答えさせた 9)。その結果、聴取者は音楽が意図した感情が知覚できたと報告して

いる。また、Morrison と Demorest は、アメリカ人とトルコ人に中国・トルコ・西洋音楽に属するある程度の長さの音楽を聴かせ、そのあとに短い音楽を聴かせて、短い音楽が長い音楽の一部かどうかを答えさせる実験を行った 10)。その結果、聴取者が親しんでいる音楽の方がよりよく判断できるという結果を得ている。音楽と言語の相互作用に関する研究も多数あり、例えば音楽と言語の構造知覚に類似性があるという指摘や 11)、歌唱練習が話声の声質やピッチに影響を与えるという知見 12)13)、言語のリズムとその文化での音楽のリズムに関連があるという指摘 14)などがある。これらのことから、歌声の熱唱度知覚においても、言語的・文化的背景が何らかの影響を及ぼす可能性がある。

そこで本研究では、日本語母語話者と中国語母語話者について日本語と中国語の歌声を聴かせ、その熱唱度知覚を比較する。また、熱唱度の自動評価とそれぞれの聴取者群の評価結果を比較し、熱唱度知覚の言語依存性について検討する。

2. 歌唱熱唱度

最初に、熱唱度について概説する。熱唱度は「歌唱の一生懸命さ」を示す指標である。これには、歌唱者がどの程度熱唱を意図して歌唱したかを示す「本人熱唱度」と、聴

^{†1} 東北大学
Tohoku University

取者がそれをどの程度熱唱に知覚したかを示す「知覚熱唱度」があり、単に熱唱度といえは知覚熱唱度を指す。Daidoらの研究 8)により、複数の聴取者の知覚熱唱度評価値には比較的高い相関がある(0.7前後)ことから、知覚熱唱度が多くの聴取者に共有された知覚であることが示された。また、本人熱唱度の違い(「熱唱」と「非熱唱」)に応じて、知覚熱唱度の分布に差があるため、本人の熱唱に関する意図は聴取者にある程度伝わるのが明らかになっている。しかし、本人熱唱度による知覚熱唱度の差よりも、歌唱者による知覚熱唱度の違いの方が大きいと、単独の歌唱音声からの本人熱唱度推定は難しいこともわかった。さらに、歌唱音声から3つの特徴量(A特性パワー、ずり下げ、ビブラート)を抽出し、これらの特徴量を重回帰分析によって組み合わせることで、人間の熱唱度知覚に近い客観評価値を得ることが示された。

前述の通り、本人熱唱度の識別は難しいが、コーラス全体における有声区間パワーを数秒のセグメントごとに求め、その平均と分散をモデル化することで、70%程度の精度で本人熱唱度識別が可能であることがわかっている 15)。

3. 知覚熱唱度の言語依存性(その1)

3.1 調査の概略

まず、今回行った知覚熱唱度の言語依存性調査の概略について述べる。

本研究の目的は、以下の2つである。

1. プロでない歌唱者の歌唱音声を聴取者が聴いたとき、知覚される熱唱度は聴取者の母語および歌唱音声の言語にどのように影響されるのかを明らかにする。
2. これまで開発した知覚熱唱度の自動評価手法が、異なる言語の歌唱に対して有効なのか、また異なる母語話者の熱唱度知覚推定にも有効なのかを検証する。

そのために、まず歌唱音声データの収録を行う。今回は言語による違いだけを見るため、同じメロディに対して複数の言語(日本語と中国語)で歌詞がついている曲を選ぶ。次に、日本語母語話者と中国語母語話者に、前記の歌唱音声(日本語および中国語)を聴かせ、熱唱度評価を行う。この結果を分析し、前述の1,2の項目の検討を行う。

3.2 歌唱音声データ

歌唱音声を収録するための曲として、「雪の華」(松本良喜作曲)を選択した。この曲は、Satomi 作詞による中島美嘉の歌であるが、中国、韓国、米国等でカバーされている。今回は、いくつかある中国語カバーのうち、蔡淳佳によるカバー「对不起我爱你」(作詞:徐世珍)を用いた。なお、日本語・中国語で歌詞の意味内容が異なる。

歌唱者は日本語母語話者6名と中国語母語話者6名で、いずれも21歳から25歳の大学生・大学院生である。歌唱者のいずれも歌唱の専門的なトレーニングを受けていない。

収録は防音室内で行われ、歌唱者はそれぞれの母語で上記の曲を2回歌唱した。うち1回は「熱唱」、もう1回は「普通」で歌唱するよう指示した。「熱唱」と「普通」の解釈は歌唱者に任されている。これを44.1kHz サンプリング・16ビット量子化・モノラル形式で収録した。

収録した歌唱音声から一部を切り出し、評価用の音声セットを作成した。評価用セットとして“Theme set”と“Variety set”の2種類作成した。“Theme set”は1つの歌唱から同じメロディの部分で2カ所抽出したものである。“Variety set”は、導入部・中間部・サビ・エンディングの4カ所から抽出したものである。長さはいずれも2小節程度である。最終的に作成された音声クリップは、Theme set で48個(2フレーズ×2言語×6歌唱者×2熱唱度)、Variety set で96個(4フレーズ×2言語×6歌唱者×2熱唱度)である。すべての音声クリップの振幅は正規化されている。

3.3 評価手順

評価実験には、6名の日本語母語話者と6名の中国語母語話者が参加した。評価者は、それぞれの音声クリップをヘッドフォンにより聴取した後、表1に示す評価語を選択した。各評価者は日中両言語による歌唱音声をすべて評価した。音声の聴取および評価はすべて防音室内で行われた。

表1 熱唱度の評価語と値

評価語	値
熱唱に聞こえる	2
どちらでもない	1
熱唱に聞こえない	0

3.4 評価値の評価者間相関

まず、評価者の評価がどの程度一致しているかについて、ピアソンの相関係数によって分析した。評価者を日本語母語話者(JP)と中国語母語話者(CN)に分け、それぞれのグループ内での2名の相関係数の平均(JP-JP, CN-CN)、および日本語母語話者と中国語母語話者の間の相関係数の平均(JP-CN)を計算した。これは次のようにして計算された。

日本語歌唱音声を s_1^j, \dots, s_N^j 、中国語歌唱音声を s_1^c, \dots, s_N^c とする。グループ $G \in \{JP, CN\}$ に含まれる評価者 $e \in G$ について、 e による歌唱音声 s の熱唱度評価値を $E(s, e)$ とする。このとき、歌唱音声の言語 $L \in \{J, C\}$ に関するグループ G_1, G_2 の話者間相関 $\rho_L(G_1, G_2)$ を次のように計算する。

$$\bar{E}_L(e) = \frac{1}{N} \sum_{i=1}^N E(s_i^L, e) \quad (1)$$

$$Q_L(e) = (E(s_1^L, e) - \bar{E}_L(e), \dots, E(s_N^L, e) - \bar{E}_L(e)) \quad (2)$$

$$\rho_L(e_1, e_2) = \frac{Q_L(e_1) \cdot Q_L(e_2)}{\|Q_L(e_1)\| \cdot \|Q_L(e_2)\|} \quad (3)$$

$$N(G_1, G_2) = \begin{cases} \frac{|G_1| \cdot (|G_1| - 1)}{2} & G_1 = G_2 \\ |G_1| \cdot |G_2| & G_1 \neq G_2 \end{cases} \quad (4)$$

$$\rho(G_1, G_2) = \frac{1}{N(G_1, G_2)} \sum_{e_1 \in G_1} \sum_{e_2 \in G_2, e_2 \neq e_1} \rho_L(e_1, e_2) \quad (5)$$

結果を図 1 に示す．ここでは、日本語歌唱と中国語歌唱の評価結果を分けて表示している．誤差棒は標準偏差である．評価者グループの組み合わせ(JP-JP, CN-CN, JP-CN)と歌唱音声の言語を要因とした 2 元配置分散分析を行ったところ、評価者グループの組み合わせに有意な差が観測された(p=0.0415)．そこで Tukey 法による多重比較検定を行ったところ、JP-JP と CN-CN の間に有意差が見られた(p=0.0354)．すなわち、中国語母語話者同士の方が有意に評価値の相関が低かった．

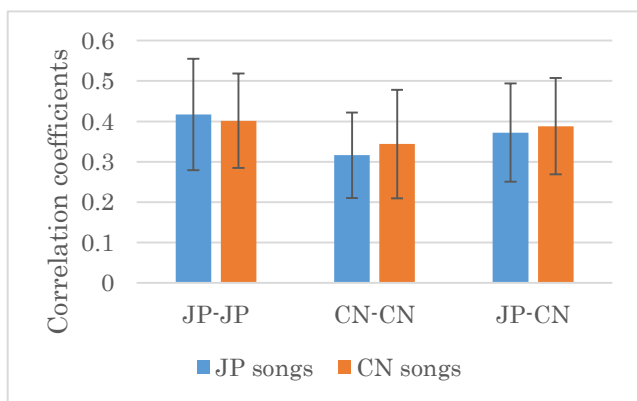
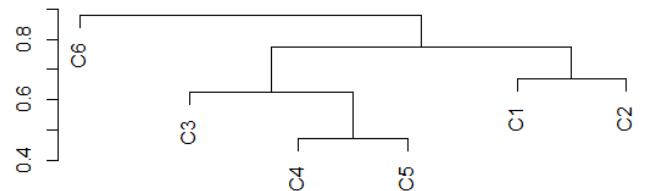


図 1 評価者間の相関係数の平均値

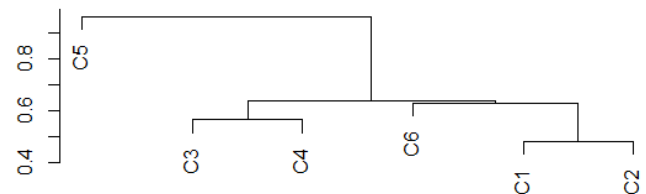
中国語母語話者による評価の相関が低いことに関して、評価者のクラスタ分析を試みた．中国語母語話者 e_1, e_2 の評価値の相関 $\rho_L(e_1, e_2)$ に対して、

$$d_L(e_1, e_2) = 1 - \rho_L(e_1, e_2) \quad (6)$$

を 2 評価者間の距離とみなし、Ward 法による階層的クラスタリングを行った．結果を図 2 に示す．(a), (b)はそれぞれ日本語歌唱および中国語歌唱の評価値の相関に基づくクラスタリング結果である．ここで我々は、(b)の結果に注目した．中国語歌唱（評価者にとっては母語による歌唱）の評価において、ほとんどの評価者が互いに高い相関（小さい距離）を示す中、評価者 C5 だけが異なる傾向を見せている．この評価者は、今回の 6 名の中国語母語評価者の中で、唯一日本語を専攻している．そこで、日本語への曝露量の違いが評価に関係している可能性があると考え、追加実験を行った．



(a) 日本語歌唱の評価結果によるクラスタ



(b) 中国語歌唱の評価結果によるクラスタ

図 2 中国語母語評価者のクラスタリング結果

4. 知覚熱唱度の言語依存性（その 2）

4.1 評価者の追加

本節の実験では、前節の実験に加えて、大学において日本語を専攻する中国人留学生 7 名を評価者に加えた．日本語母語話者、日本語を専攻しない中国語母語話者、および日本語を専攻する中国語母語話者をそれぞれ Group A, B および C とする．それぞれのグループの内容及び人数を表 2 に示す．新たに実験に参加した評価者についても、前節と同じ評価を依頼した．

表 2 各グループの人数

Group	対象	人数
A	日本語母語話者	6
B	日本語を専攻としない中国語母語話者	5
C	日本語を専攻する中国語母語話者	8

4.2 評価者間の評価値の相関

3 つのグループのそれぞれについて、グループ内での評価者間相関の平均をとったものを図 3 に示す．誤差棒は標準偏差である．この結果に対し、グループと歌唱音声言語の 2 つを要因とする 2 元配置分散分析を行った．その結果、グループ間に有意な差があり(p=0.00246)、歌唱言語には有意な差はなかった．下位検定として Tukey の多重比較検定を行ったところ、Group A と Group C の間のみに有意な差が見られた(p=0.00168)．このことから、日本語母語話者の中での評価の相関に比べ、日本語を専攻する中国語母語話者の評価値の相関が有意に低いという結果となった．原因については定かではないが、日本語の習得の度合いによって歌唱音声の評価が影響を受けた可能性が考えられる．

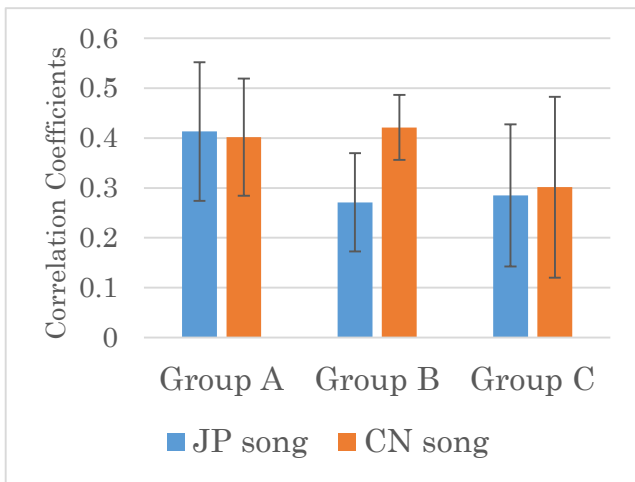


図 3 評価者間の相関係数の平均値

4.3 熱唱度の評価値と歌唱言語の関連

各グループと歌唱音声言語について、熱唱度の評価値の平均を比較した。結果を図 4 に示す。差は微妙であるが、2元配置分散分析を行ったところ、グループ間に有意な差が見られた($p=0.00871$)。歌唱音声言語間に有意な差はなかった。Tukey 法による多重比較検定を行ったところ、Group B と Group C の間に有意差が見られた($p=0.013$)。これは前節での結果と整合している。

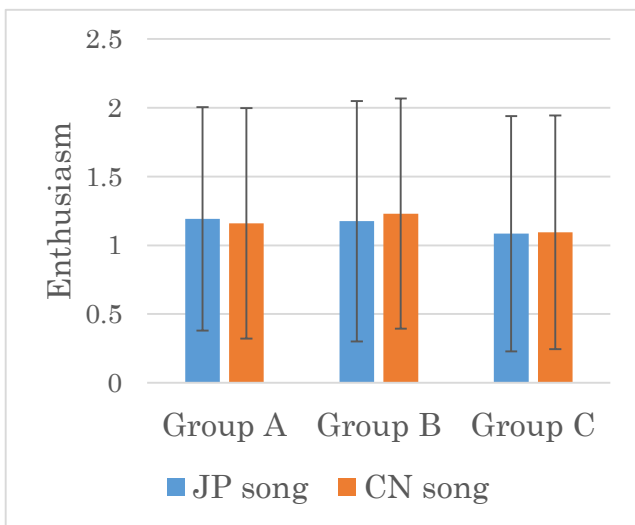


図 4 グループ・歌唱音声言語と熱唱度評価値

4.4 フレーズと熱唱度の関連

今回の評価用音声は、繰り返し部分(Theme)と、Variety set の各部 (導入部(Beginning), 中間部(Bridge), サビ(Climax), エンディング(Ending)) からなっているので、それぞれの歌唱部分と歌唱言語にどのような関係があるのか調査した。その結果を図 5 と図 6 に示す。歌唱音声言語、フレーズおよびグループの3つを要因とする分散分析を行ったところ、フレーズ間($p<0.001$)およびグループ間($p=0.0287$)に有意な差が見られた。歌唱音声言語間には有意差は見られなかった。

Tukey 法による多重比較検定を行ったところ、グループ間では Group B と Group C の間に差が見られた($p= 0.0345$)。フレーズ間の有意差については表 3 にまとめた。Climax (サビ, 熱唱度最大) と Ending (エンディング, 熱唱度最小) はその他のすべてのフレーズと有意な差があり、中間の3つについては相互に有意な差は見られなかった。

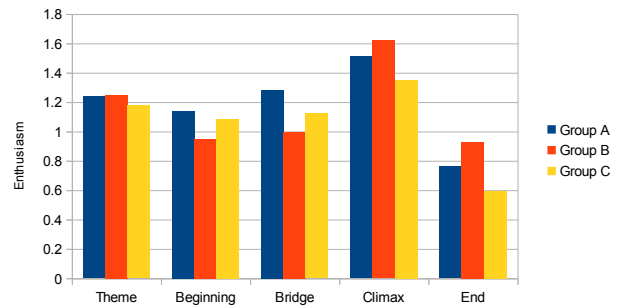


図 5 日本語歌唱のフレーズごとの評価結果

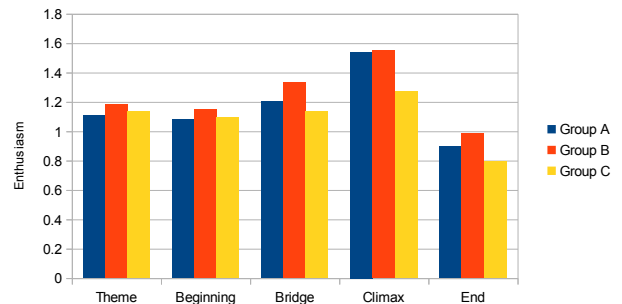


図 6 中国語歌唱のフレーズごとの評価結果

表 3 フレーズ間の有意差

	Beginning	Bridge	Climax	Ending
Theme	n.s.	n.s.	***	***
Beginning		n.s.	***	***
Bridge			***	***
Climax				***

n.s.: 有意差なし, ***: $p<0.001$

4.5 クラスタ分析

各評価者の評価値 144 次元 (Theme set + Variety set, 日本語および中国語歌唱) をベクトルとして、そのユークリッド距離を基準に、評価者のクラスタ分析を行った。アルゴリズムは Ward 法による階層的クラスタリングである。結果を図 7 に示す。Group C の評価者は比較的孤立する傾向にあるが、Group A, B, C の評価者は入り混じっており、はっきりした傾向を見いだすことはできなかった。

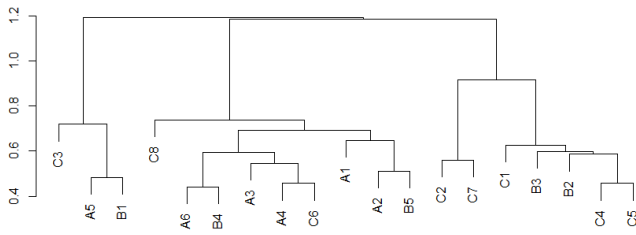


図 7 クラスタ分析結果

4.6 自動熱唱度評価との比較

3つのグループによる歌唱音声の評価値と、熱唱度の自動評価値との比較を行った。このため、文献8)の評価手法を利用して計算した自動評価値と、各グループの評価者による主観評価値の相関を計算した。結果を図8に示す。

この結果から、日本語の歌唱を日本語母語話者が評価する場合には、自動評価はある程度高い相関を示すことがわかる。日本語母語話者の評価値間の相関平均が0.4程度であるから、自動評価手法は人間による評価と同程度の結果を出していると言える。しかし、評価対象が中国語歌唱である場合、また評価者が中国語母語話者である場合には、評価値が低くなる現象が見られている。相関係数の差の検定を行ったところ、各グループでの日本語歌唱と中国語歌唱との差はすべて1%有意、日本語歌唱の中ではGroup AとGroup Cが1%有意、中国語歌唱の中ではGroup AとGroup Cが5%有意であった。

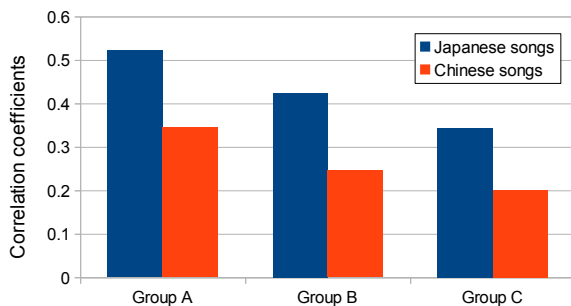


図 8 熱唱度自動評価値と主観評価値の相関

この結果を図3と比較して考える。図3から、評価者間の評価値の相関は、歌唱音声の言語によらず同程度である。これは、歌唱音声の言語が変わっても、複数の評価者がそれらに対して同じような評価をしていることを表すと考えられる。しかし図8から、日本語歌唱と中国語歌唱を同じように自動評価すると、中国語歌唱に対しては主観評価と異なる値となることがわかる。この結果から、歌唱音声の言語が変わることによって主観評価の基準が変わっており、それは複数の評価者の間で一貫性があるということが示唆される。

5. まとめ

歌唱熱唱度の言語依存性に関する調査の結果について述べた。同じメロディで日本語と中国語の歌詞で歌われる歌唱について、日本語母語話者と中国語母語話者による熱唱度評価を行い、主観的な熱唱度の値、および評価者間の熱唱度の相関について調べた。また、主観評価の結果と、熱唱度自動評価の結果について比較を行った。その結果、次のことが明らかになった。

1. 日本語または中国語を母語とする話者が、母語とそれ以外の言語で歌唱された音声聞いたときに、その熱唱度は歌唱音声の言語によらない。
2. 日本語を専門として学習する中国語母語話者の間での熱唱度知覚は、通常の日本語母語話者および中国語母語話者の中での熱唱度知覚と比較して、一貫性が低い。
3. 日本語母語話者の日本語歌唱音声知覚をターゲットとして学習した熱唱度自動評価による推定値は、同条件での主観評価値とは高い相関を示すが、評価者や歌唱言語が異なると相関が低下する。

これらの結果のうち、3は特に興味深い。先にも述べたように、人間による評価では歌唱音声の言語によらず比較的评价が一貫しているため、人間が歌唱音声の種類によって評価基準を変えているのかもしれない。

今後の課題としては、今回の結果が本当に言語の違いによるのかどうかをはっきりさせるために、日本語・中国語以外の言語での歌唱や、言語を用いない歌唱音声（スキヤットなど）と比較を行うことが必要であろう。日本語を専攻する中国語母語話者の評価傾向が異なることについては、はっきりした原因は現時点ではわからない。他の言語についても実験を行うことで、言語学習と特定言語の歌唱音声の知覚についての相互作用をさらに検討する必要がある。

参考文献

- 1) Ekholm, E., Papagiannis, G. C. and Chagnon, F. P.: Relating objective measurements to expert evaluation of voice quality in western classical singing: Critical perceptual parameters, *J. of Voice*, Vol. 12, No. 2, pp. 182-196, 1998.
- 2) Watts, C., Barnes-Burroughs, K., Estis, J., and Blanton, D.: The Singing Power Ratio as an Objective Measure of Singing Voice Quality in Untrained Talented and Nontalented Singers, *J. of Voice*, Vol. 20, No. 1, pp. 82-88, 2006.
- 3) Trainor, L. J., Clark, E. D., Huntley, A., and Adams, B. A.: The acoustic basis of preferences for infant-directed singing, *Infant Behavior and Development*, Vol. 20, No. 3, pp. 383-396, 1997.
- 4) 齋藤毅, 辻直也, 鶴木祐史, 赤木正人: 歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析, *日本音響学会誌* Vol. 64, No. 5, pp. 267-277, 2008.
- 5) 中野倫靖, 後藤真孝, 平賀謙: 楽譜情報を用いない歌唱力自動評価手法, *情報処理学会論文誌* Vol. 48, No. 1, pp. 227-236, 2007.
- 6) 竹内英世, 保黒政大, 梅崎太造: 人の主観評価に近いカラオケ採点法, *電気学会論文誌 C*, Vol. 130, No. 6, pp. 1042-1053, 2010.
- 7) Yonezawa, T., Suzuki, N., Abe, S., Mase, K., and Kogure, K.:

- Perceptual Continuity and Naturalness of Expressive Strength in Singing Voices Based on Speech Morphing, *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2007, Article ID 23807, 9 pages, 2007.
- 8) Daido, R., Ito, M., Makino, S., and Ito, A.: Automatic evaluation of singing enthusiasm for karaoke, *Computer Speech & Language*, Vol. 28, No. 2, pp. 501-517, 2014.
- 9) Balkwill, L.-L. and Thompson, W. F.: A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues, *Music Perception: An Interdisciplinary Journal*, Vol. 17, No. 1, pp. 43-64, 1999.
- 10) Morrison, S. J. and Demorest, S. M.: Cultural constraints on music perception and cognition, *Progress in Brain Research*, Vol. 178, pp. 67-77, 2009.
- 11) Patel, A. D.: Language, music, syntax and the brain, *Nature Neuroscience*, Vol. 6, No. 7, pp. 674-681, 2003.
- 12) Mendes, A. P., Brown Jr., W. S., Rothman, H. B. and Sapienza, C.: Effects of Singing Training on the Speaking Voice of Voice Majors, *Journal of Voice*, Vol. 18, No. 1, pp. 83-89, 2004.
- 13) Schön, D., Magne, C. and Besson, M.: The music of speech: Music training facilitates pitch processing in both music and language, *Psychophysiology*, Vol. 41, pp. 341-349, 2004.
- 14) Patel, A. D. and Daniele, J. R.: An empirical comparison of rhythm in language and music, *Cognition*, Vol. 87, No. 1, pp. B35-B45, 2003.
- 15) Ito, A.: Assessing Intended Enthusiasm of Singing Voice Using Energy Variance, *Proc. Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 558-561, 2014.