

音声の自己聴取音化によるものまね練習支援

嶋田 智稀^{1,a)} 井村 誠孝¹

概要：話者が自身で似ていると感じるものまねを聴者に披露しても、聴者は似ていないと感じてしまうことが多い。この現象は話者と聴者の聴き取る音声、聴取音の差異が原因であると考えられる。本研究では、聴取音の差異を取り除き、ものまね対象音源の自己聴取音化によるものまね練習の支援を目指す。提案手法では骨導音を付加したものまね対象音声を提示し、話者がその音声を模倣することで聴者との聴取音の差を取り除く。同時録音した気導音と骨導音のスペクトル比から骨導伝達関数を導出し、提示する音声を生成する。本稿では骨導伝達関数を推定し、ものまね対象音声を自己聴取音化した結果について示す。

1. はじめに

近頃、メディアや動画共有サービスでもものまねを披露する人を目にする事が多く、ものまねに対する世間の認知度は次第に高まってきている。しかし、ものまねの習得は簡単なことではない。実際、ものまねを練習し、話者自身が似ていると感じるものまねを聴者に披露しても、聴者は似ていないと感じてしまうことが多い。ものまねは、声色を真似る声帯模写と仕草や容姿を真似る形態模写の2種類に分けられる。本研究は声帯模写を対象とし、ものまねの練習の支援を目的とする。

音声には、声帯の振動が舌や口を通じて空気を振動させ聴覚器官に伝わる気導音と、声帯の振動が頭蓋骨を通じて直接的に聴覚器官に伝わる骨導音の2種類がある。一般的に、声帯模写の習得においては、話者が自己聴取音をものまね音声に近づかせるように練習を行うが、話者と聴者のそれぞれの音声伝達経路は、図1に示すように、聴者の耳では気導音のみを捉えているのに対して、話者の耳では気導音に加えて骨導音も同時に捉えている点で異なる。両者の聴取音は骨導音の影響によって差が生じるため、ものまね対象音声を真似る従来の練習手法でのものまねの習得は困難である。

本研究では、話者と聴者の間に生じる聴取音の差を取り除くため、骨導音を付加したものまね対象音声を話者に提示し、話者がその音声を真似ることで聴者との聴取音の差を取り除く手法を提案する。

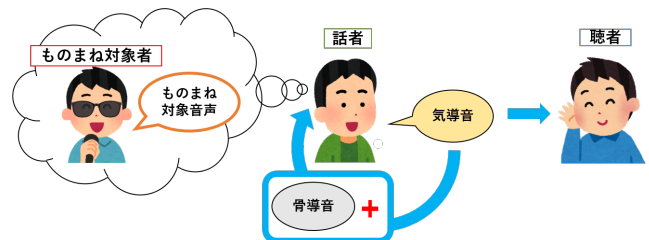


図1 話者と聴者の聴取音

2. 提案手法

本研究では骨導音を付加したものまね対象音声を提示し、話者がその音声を真似ることで聴者との聴取音の差を取り除く手法を提案する。提案手法の処理の流れを図2に示す。話者の気導音を $F_a(s)$ 、話者の骨導伝達関数を $H(s)$ とすると、話者の骨導音 $F_b(s)$ 、自己聴取音 $F_s(s)$ はそれぞれ以下の式で表される。

$$F_b(s) = H(s)F_a(s) \quad (1)$$

$$F_s(s) = F_a(s) + F_b(s) \quad (2)$$

従来の練習手法では、話者は自己聴取音がものまね音声に近づくよう練習を行うため、ものまね対象音声を $F_t(s)$ とすると、学習完了状態は以下の式で表される。

$$F_s(s) = F_t(s) \quad (3)$$

しかし、聴者に届く音声は気導音のみなので、完全な再現状態は

$$F_a(s) = F_t(s) \quad (4)$$

であり、学習完了状態と再現状態との間に話者の骨導音分の差異が生じる。提案手法では以下の式(5)で表される、

¹ 関西学院大学
Kwansei Gakuin University, Sanda, Hyogo 669-1337, Japan
^{a)} dbt47223@kwansei.ac.jp

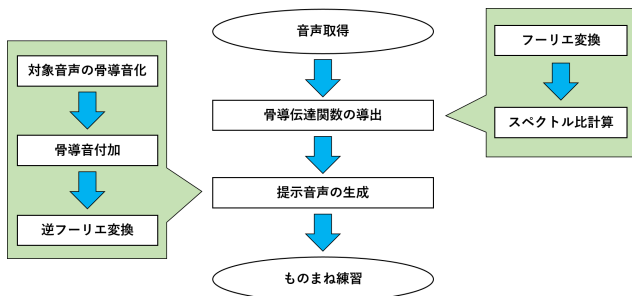


図 2 処理の流れ

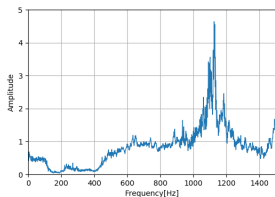


図 3 /a/の骨導伝達関数

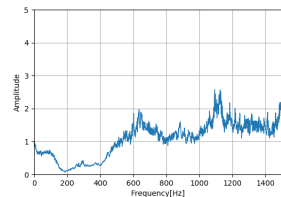


図 4 平均の骨導伝達関数

完全な再現状態時の自己聴取音 $F_t^*(s)$ を話者に提示することで、話者と聴者の聴取音の差を除去する。

$$F_t^*(s) = F_t(s) + H(s)F_t(s) \quad (5)$$

3. 実験

3.1 音声信号の取得

音声信号は、物音が少ない室内にて気導音はコンデンサマイクロホン (SONY 製 ECM-C10), 骨導音は骨伝導イヤホンマイク (TEMCO 製 EM20N-T3.5P) を使用して録音した。サンプリング周波数を 44kHz, 量子化レベルを 16 ビットとし、両音声の同期をとるため、ステレオマイクアンプの左右のチャンネルに気導音と骨導音を入力した。録音対象は日本語 5 母音/a/~/o/の気導音と骨導音とした。

3.2 骨導伝達関数の導出

提示する音声 $F_t^*(s)$ を生成するためには、骨導伝達関数 $H(s)$ を求める必要がある。骨導伝達関数の導出にあたっては、まず、自身の気導音と骨導音を複数回同時録音し、得た音声に対してフーリエ変換を行う。母音毎に各周波数のパワーの平均値を求め、気導音と骨導音のスペクトル比から骨導伝達関数を得る。

10 回の試行で得た音声から導出した骨導伝達関数の中で、母音の中でも強い特徴を示す [1] とされる/a/の結果を図 3 に示し、日本語 5 母音の骨導伝達関数を平均した結果を図 4 に示す。両者の周波数特性が 1kHz-1.5kHz 区間で大きく異なることが見てとれるため、ものまね対象音声の変換に対して、平均した骨導伝達関数を扱うことは最適でないとと言える。

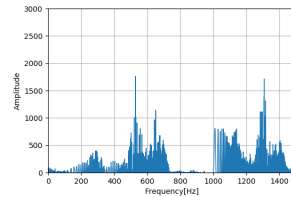


図 5 対象音声の周波数特性

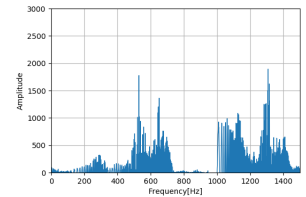


図 6 生成音声の周波数特性

3.3 適切な気導音と骨導音の比率の検証

本節では、森らの手法 [2] を参考に、自己聴取音に含まれる気導音と骨導音の比率を導出する。10 名の実験協力者に対して、気導音と骨導音の同時録音を行う。全体を 100 とした時の骨導音の割合をスライダーで操作し、その比率で気導音と混ぜ合わせた音声を生成して提示するシステムを作成した。話者自身がスライダーを操作して各母音に対する合成音を生成し、自己聴取音と比較することで、自己聴取音に含まれる骨導音の比率を選択する。全体で平均すると、気導音 71 %, 骨導音 29 % の比率であることが確認できた。

3.4 ものまね対象音声の自己聴取音化

ものまね対象音声を認知度の高いキャラクターである「マリオ」の音声として、自己聴取音化を行った。3.2 節で得た骨導伝達関数をものまね対象音声のスペクトルにかけ、3.3 節で得た比率をもとに合成音を生成した。ものまね対象音声の周波数特性を図 5、生成した音声の周波数特性を図 6 に示す。2 つの音声には差異があまり見られなかった。また、音声を聴き比べたところ、ものまね対象音声に比べて、生成した音声の音質が劣る結果となった。大きな原因として、母音毎で骨導伝達関数と、気導音と骨導音の比率が異なる点が考えられる。

4. おわりに

本稿では、ものまね対象音声の自己聴取音化によるものまね練習を支援する手法を提案した。骨導伝達関数が母音毎に異なるため、5 母音の平均した骨導伝達関数によるフィルタリングでは、十分にものまね対象音声を自己聴取音化することができなかった。今後の展望として、ものまね対象音声を音節ごとに区切り、処理を行うことが必要であると考えられる。

参考文献

- [1] Hansen, M. O. and Stinson, M. R.: Air conducted and body conducted sound produced by own voice, Canadian Acoustics, Vol. 26, No. 2, pp. 11-19, (1998).
- [2] 森幹男, 吉田千夏, 荻原慎洋ほか: 自己聴取音に占める気導音と骨導音の割合の推定, 電気学会論文誌 C, Vol. 127, No. 8, pp. 1268-1269 (2007).