

# 発声のメタ認知促進システム “いい声マイク” の提案

矢島 佳澄<sup>†</sup>      笥 康明<sup>†</sup>      諏訪 正樹<sup>†</sup>

本研究は、発声という暗黙知的な要素の濃い身体スキルを向上させるために、ユーザのメタ認知を活性化させる支援システムとしての“いい声マイク”の提案を行うものである。“いい声”とは「音がしっかりと発せられ、音響として響く」声であると現時点で解釈している。本システムはユーザの発声のあるアルゴリズムによって評価し、マイク台の光の色とモニタ上のグラフで視覚化する。これこれのように発声せよというアドバイスをを行うものではない。発声に際して呼吸や身体の使い方、姿勢などをユーザが試行錯誤し、発声にまつわるメタ認知の自発的活性化によりスキル向上を支援するものである。

## e-Voice Mic: A Proposal on A System for Promoting Metacognition of Voice Production

KASUMI YAJIMA<sup>†</sup>      YASUAKI KAKEHI<sup>†</sup>      MASAKI SUWA<sup>†</sup>

This research proposes “e-Voice Mic”, a system which encourages users to develop metacognition, in order to improve vocal skill which tend to be implicit. “e-Voice” means evaluating voices on the basis of acoustic vibration. As users talk into the microphone, their voices would be evaluated with the algorithm we have composed, and is visualized by the lighting of the microphone stand and a graph on the display. It will not give any concrete advise to users. When users talk, they will be expected to adjust the way they breathe, use their body, or take their posture. It will help them upgrade their vocal skill by the spontaneous activation of metacognition which is connected to voice production.

### 1. はじめに

本研究では、人間の発声に着目し、ユーザが“いい声”を出せるように身体的スキルの向上を促すメディアの提案を目的とする。デジタルメディアを用いて、ユーザに“いい声”を出すスキルを提供するには、いくつかのアプローチが考えられる。一つ目は、ユーザの声をシステムに入力し、エフェクトをかけることで“いい声”に変換し、リアルタイムに出力する方法である。これは、ユーザの身体自体には変更を加えず、ソフトウェア（デジタル）的に出力結果を良くするものである。二つ目は、入力された声に対して、“もっと口を大きく”など、どこをどうすれば良くなるといった具体的な指針を提供する方法である。これは、具体的な変更点を指摘して、ユーザの身体スキルの向上を促すものである。三つ目は、入力された声を、“いい声”とされる指標と照らし合わせ、その結果を点数など音声とは別の形でリアルタイムに出力する方法である。これは、現在の状態をフィードバックすることにより、ユーザ自身が試行錯誤のもと因果関係を自発的に導き出せるよう、身体的スキル向上のきっかけを

与えるものである。

さて、役者等の発する声に対して、“いい声”という感想を共有することは少なくないが、“いい声”とはどのような事象を指すのだろうか。どのような声を“いい声”と判断しているのかはいわゆる暗黙知である。暗黙知的な性格を有する身体的スキルの学習を支援するには、上記の第1, 2のアプローチのように一貫して決まった尺度や指針を与えるよりも、第3のように、各ユーザがその指針自体を導きだせる支援を行うのが好ましい。

指針をユーザが体験の中で動的に決定／更新していくプロセスを支援するために、本研究では“メタ認知”の手法に着目する。ここで、メタ認知とは、身体が体感していること（既に感知できる着眼点）をできるだけ言葉にし、言葉領域の推論で新たな着眼点を得て、新たに得た着眼点を視点に加えて再度自分の身体の動きを見つめ直す（体感を感じようとする）行為であると定義される<sup>1)</sup>。更に諏訪らは、身体感覚を言語化するだけでなく、イメージやサウンドなどより抽象的なモノへの変換・外化（シンボル化）も含めた広義の手法を提唱している<sup>2)</sup>。本研究では、ユーザの声をある一定の尺度により評価し、評価値をシンボル化してユーザにリアルタイムにフィードバックする。こ

<sup>†</sup> 慶應義塾大学 環境情報学部

れにより、ユーザが主体的に声質、姿勢や所作に意識を及ぼし、“いい声”の出し方を開拓していただけるようなシステムを提案する。

図1に本研究で提案する発声のメタ認知を促すマイクシステム“いい声マイク”の概観を示す。ユーザの声をPCに入力し、その声の分析結果に伴い、マイク台の照明の色が変化する。また声の評価データの時系列変化がPC上で閲覧できる。以下、本稿では関連研究を整理し、本研究の概要、プロトタイプシステムの実装に関して述べる。

図1：いい声マイク

## 2. 関連研究

リアルタイムにスキル学習を支援するツールは従来にも様々提案されている。例として、曾我らの研究<sup>3)</sup>は、音声とテキストによってデッサンの補助・支援を行う。INSTILL<sup>4)</sup>は、英語のアクセントを力覚と視覚と聴覚でフィードバックするシステムである。石田らの研究<sup>5)</sup>では、振動を用いて、ピアノの即興演奏を支援する。これらは、いずれもスキル学習の中でも形式知の要素を扱っている。我々は発声という暗黙知の要素を多く含むスキルの向上を目的としている。

以下では、発声のスキル向上を目的とした関連研究として、1章で述べた3つのアプローチと対応づけて議論する。第1のアプローチ（声の自動変換）を採るものとして、中村らの肉伝導人工音声の変換に基づく喉頭全摘出者のための音声コミュニケーション支援システム<sup>6)</sup>が挙げられる。これは、喉頭全摘出者が発声した微弱な音声を自然な音声に変換して出力し、コミュニケーションを助けるものである。

第2のアプローチに関しては、一般的に行われるボイストレーニングがこれに該当する。これまでにデジタルメディアを用いてインタラクティブにスキル向上

を促すためのシステムの提案は多くない。

第3のアプローチに関しては、代表的なものとしてカラオケシステムの歌声採点機能が挙げられる。これも、ある尺度のもとにフィードバックして、発声の向上を促すという点で本研究の考え方と共通性はある。しかし、カラオケ採点システムの場合は、音程やリズムのお手本と比べてどれくらいずれているかにより採点がなされるものである。本研究で扱う“いい声”にはお手本がない。

本研究と目的は異なるが、同様にメタ認知的思考を活性化させることにより学習を支援するツールの例として、MotionPrism<sup>2)</sup>、ScoreIlluminator<sup>2)</sup>がある。MotionPrismは、野球の打撃練習のために身体の姿勢の推移をカラーバーによって可視化し、学習者に気付きを促すものである。ScoreIlluminatorは、オーケストラスコア上の音符の類似度を色づけによって視覚化し、スコアの意味解釈におけるメタ認知を促すものである。これらの研究は試行ごとに評価をフィードバックする手法を採っている点でリアルタイム性はない。発声は連続的な行為であることから、試行後にまとめて「いまの声は全体的にいい声だった」と採点されたとしても、どんな意識や所作がその原因であったかに関するメタ認知は活性化されにくい。発声とともにほぼリアルタイムに結果がフィードバックされる環境が望まれる。

## 3. “いい声マイク”の概要と設計

### 3.1 システム概要

“いい声”を出すためのメタ認知をユーザに促すために、発声の評価をリアルタイムに色とグラフで可視化する。ユーザは発声しながら結果を見て、リアルタイムに声の出し方や身体の使い方を工夫するというメタ認知サイクルに誘い込まれる。マイクに入力された音声は、後述するアルゴリズムによって評価され、マイク台に取り付けられたLEDの光の色に反映される。時系列の変化をLEDの発光とモニタに表示するグラフで視覚化する。

### 3.2 “いい声”判定アルゴリズムの定義

さて、“いい声”とはどんな声だろうか？その定義を規定した研究は筆者らの知る限り未だかつてない。“声がいい”かどうかの主観的判断は、聴く人の状況や感情にも依存する可能性もある。本論文は状況や感情に依存するような“いい声”を探究するわけではない。少なくとも現時点で筆者らが想定する定義は、「音としての声がしっかりと発せられていて、それが響いている」という現象である。そのような現象を判

定できる作業仮説を筆者らが如何に導出したかを以下に述べる。

### 3.2.1 予備実験 1: バットの素振り音

野球においてバットの素振りは重要な練習である。“音が低くて鋭い”ことがよい素振りであるという定説が野球界には存在する。第3著者（野球歴 40 年弱）が 10 本の素振りをを行い、振った体感と音から素振りのよさを主観的に判断した。その横でオーケストラを含む音楽経験 25 年の被験者（野球経験は皆無）も同時に“音のよさ”を主観判定した。被験者は上記の定説は知らされていない。驚くべきことに、最もよいと判定された素振りと、最も悪いと判定された素振りは、両者で完全に一致した。図 2 の上と下は、最もよい素振り、悪い素振りの音の周波数分析である。差異は低周波数のパワーにある（赤線の付近）。よい素振りは低周波のパワーピーク値が、それより高い領域にある 2 つのピークのパワーと同程度であるが、悪い素振りは明らかに低周波数のパワーが弱い。

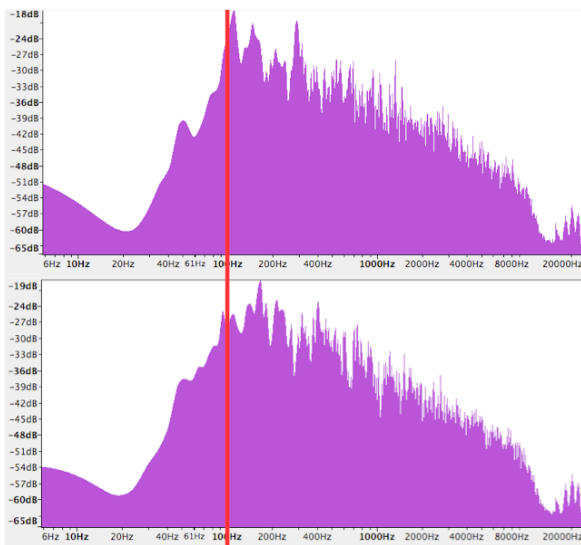


図2 素振り音の解析グラフ

### 3.2.2 予備実験 2: 俳優と素人の朗読

ある劇団の舞台俳優 4 名と一般人 3 名（発声訓練の経験なし）の被験者が同じ詩を朗読し（長さは約 1 分）。マイク録音した音声を周波数解析した。俳優の中のひとりはその劇団の主力俳優であり、25 年以上のキャリアである。3 回朗読した音声の中で、朗読後の本人の主観的判断により、最も悪くないと判定された音声の周波数分析が図 3a である。第 2, 3 著者が 3 回のなかで最もよいと判断した音声の周波数分析がそれぞれ図 3b, c である。主力俳優は、人間の声領域であると言われている周波数領域でのパワーピーク値が

ほぼ同程度に揃っているのに対し、第 3 著者は声領域の中の高周波数側でパワーが弱い。第 2 著者は比較的主力俳優に近い傾向がみられる。実は、第 2 著者は実験前に俳優陣から「いい声をしてますね」と言われていたのである。声領域以下のピークは環境固有のものであり、被験者全員に同じ周波数で存在した。

声領域でのパワーが同程度に揃っているという傾向は、良い素振りの傾向と一致している。更に、主力俳優の場合、声領域より高い周波数領域でのパワー減少度合いが綺麗な直線を描いている。他の 3 名の俳優は、いずれも主力俳優と一般人の中間的傾向を示していた（スペースの関係上グラフは割愛する）。

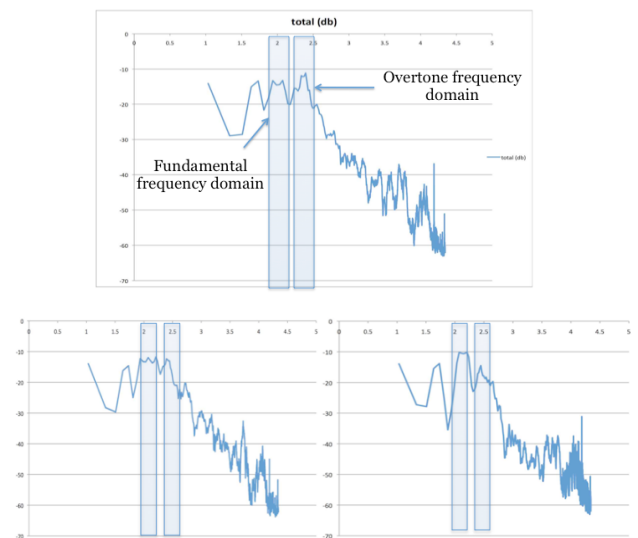


図3 詩の朗読分析（上が a, 左下が b, 右下が c）

### 3.2.3 判定アルゴリズム

そこで我々は“いい声”の判定基準の仮説として、(1)第 1, 第 2 フォルマント領域（人間の声領域と言われる周波数領域）のパワーの分散が低いこと、(2)第 3 フォルマント領域以上の直線回帰残差が低いことを挙げた。図 3a に観察できるような台形状の折れ線にフィッティングしたときの残差が小さいことをもって“いい声”と判定するアルゴリズムを考案した。

前節までに示す通り、これは少数実験における主観的判断に基づく仮説である。しかし、バットスウィングと人間の声という全く異なる領域の実験が同じ傾向を示唆していることから、本研究はこの作業仮説を基にシステム構築を開始した。本システムを生活で使用するユーザが、自分の声を開拓するメタ認知的思考を活性化するプロセスを通じて、作業仮説自体も進化し、“いい声”とは何かの探究も進める予定である。またより多数の声の採取と、多数により声の評価のための実験も今後予定している。

### 3.3 ハードウェア設計

視覚化方法はさまざまあるが、メタ認知を効果的に支援するために、LEDの光の色での視覚化を採用した。色は評価値に対応して、緑（最低値）と赤（最高値）の間に色相を変化させる。スピーカのレベルメータでよく用いられる色範囲であり、直感的な理解につなげることをねらっている。LEDはマイク台に6つ垂直に取り付け、一番上のLEDが最新の評価値を表し、下に行くほど過去の評価値を示す。これにより、ユーザに時間軸方向の評価値の遷移を表す。

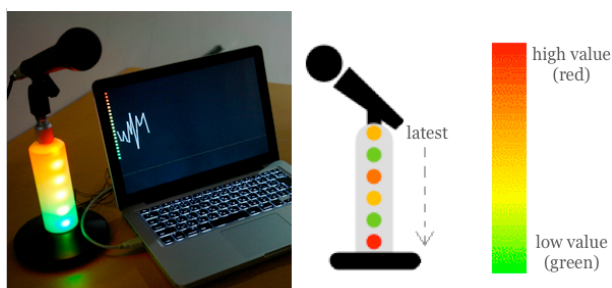


図4 マイク台のLED構成(左)と、色と評価値の関係(右)

### 4. プロトタイプと体験者の様子

マイクに入力された音声はリアルタイムにPCに取り込み、高速フーリエ変換した後、上記の判定アルゴリズムに基づいて評価値を算出し、PC画面にその時間変化を10秒ごとに折れ線グラフとして見せる(横軸が時間、縦軸が評価値)。また、評価値はArduinoに送られ、マイク台のLEDの色を評価値によって変化させる。評価値を表示する時間間隔はユーザの希望に応じて変化させることができる。

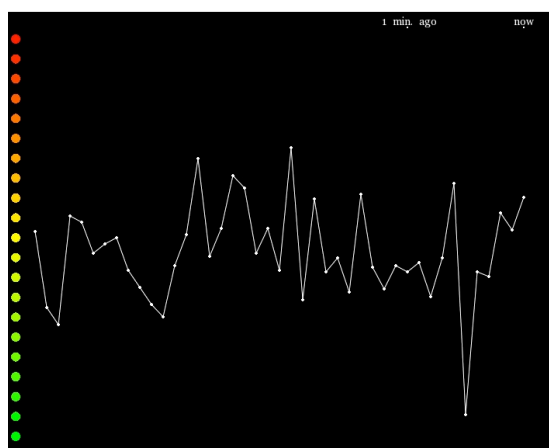


図5 モニタ上グラフ

ORF2010において本システムを展示し、100名程度の来場者が体験した。体験者からは、視覚化された評

価に関しては概ね主観評価と一致するという感想を多く得た。ただし、評価値一つのみが視覚化されていても、発声法をどう変えていいかわからないという意見や、LEDとグラフの双方に視覚化すると分かりにくいという意見もあった。評価値表示の時間幅は、どういう状況での発声か(歌う/アドリブで喋る/原稿のあるものを喋る)、またユーザの発声に関する意識のレベルに応じて、適切な秒数があると想定できる。今後評価実験を通じて探るべき要素のひとつである。

### 5. まとめと今後の課題

発声のメタ認知を促すシステム“いい声マイク”を提案・実装した。メタ認知サイクルをより自然に促すためには、視覚化(外化)手法とインタフェース、そして評価値時間幅は重要な要素であること、更に、それらはユーザの目的や反応に応じて適宜可変であることが好ましいということがわかった。システムの使用によりユーザの発声の習熟度が向上すれば、システムの可視化手法やアルゴリズムも変容させなければならない。更に、今後実証実験を重ねる過程で“いい声”の定義自体が進化する可能性までも想定しなければならない。ユーザの使用実践と通じた構成的なシステムデザインを続ける。

### 参考文献

- 1) 諏訪正樹, 赤石智哉: 身体スキル探求というデザインの術, 2009年度日本認知学会冬のシンポジウム「デザインの学と術」論文集, pp. 12-21, 日本認知科学会 (2009).
- 2) 松原正樹, 西山武繁, 伊東貴一, 諏訪正樹: 身体的メタ認知を促進させるツールのデザイン, 身体知研究会, SIG-SKL-06-03, pp. 15-22 (2010).
- 3) 曾我真人, 松田憲幸, 瀧寛和: デッサン描画中に描画領域に依存したアドバイスを提示するデッサン学習支援環境, 人工知能学会論文誌 Vol.23 No.3, pp. 96-104 (2008).
- 4) 河上朋弥, 山口武彦, 一色正晴, 佐藤誠: 力覚インタラクシオン可能な英語学習システム: INSTILLの提案, 情報処理学会インタラクシオン2008, pp. 71-72 (2008).
- 5) 石田克久, 北原鉄郎, 武田正之: 振動機能付鍵盤楽器「ぶるぶるくん」を用いた即興演奏支援システム, 情報処理学会研究報告[音楽情報科学], pp. 25-30 (2005).
- 6) 中村圭吾, 戸田智基, 猿渡洋, 鹿野清宏: 肉伝導人工音声の変換に基づく喉頭全摘出者のための音声コミュニケーション支援システム, 電子情報通信学会論文誌 D Vol.J90-D No.3, pp. 780-787 (2007).