

印象評価と音響的特徴量との関係性に基づく いい声判定アプリケーションの提案

菅原 衣織[†] 伊藤 貴之[†]お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース[†]

1. はじめに

私たちが生活する上で切っても切り離せないコミュニケーション。そのコミュニケーションを円滑にし、私たちの人柄をことば以上に伝えてくれるのが、声である。明るく明瞭な声で話している人にはいい印象を受けることがあるように、逆にぼそぼそとはっきりしない声で話す人には好印象を受け難い。また、メラビアンの法則[1]として従来から知られているように、コミュニケーションにおいて人の行動が他人にどのように影響を及ぼすかという実験では、言語情報が7%、聴覚情報が38%、視覚情報が55%を占めるという考え方がある。このように、声の印象が他者の評価に与える影響は少なくない。

我々は声の印象、つまり声質の観点から「いい声」を追求し、どんな音響的特徴が見られるのかを調査した。一般的に「いい声」と判断するためには、滑舌や話の速度、声の大きさなどいくつかの要素が複雑に絡んでくると考えられる。しかし、滑舌や声の大きさなどいくつかの要素は既存の手法を用いて改善される。そこで本調査では、粕谷[2]が提唱した声質の定義を更に限定的に捉え、声質という単語を「同じ音を同じアクセントで発した時に見られる個人的特徴」と定義した。

本研究では、本調査より得られた印象評価値と音響的特徴量との関係性を捉え、「いい声」を判定するアプリケーションを提案する。本アプリケーションを用いることで自身の声がどのような印象を与えているのかを客観的に知ることが可能になる。また改良を重ねることで、将来的には、相手に不快感を与えない声で話すためにテレホンセンターの研修に組み込むことや、講演前に自身の声を調整するためのチューニングツールとして利用できるのではと考えている。

2. 関連研究

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いい声マイク[3]では、いい声を「音としてしっかりと発せられていて、響いている声」と定義している。いい声であるかを判定するアルゴリズムの構築のために、バットの素振り音と俳優と新人の朗読を用いた予備調査を経て、第1、第2フォルマント領域のパワーの分散が低いこと、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ている。

また、酒造[4]は、声から「怒り」「喜び」「悲しみ」の感情を取得し、それを写真と円のグラフィックを用いて出力するアプリケーションを提案している。自身の声

の状態によってその日やその瞬間の感情や体調の変化を視認できることは非常に興味深い。

本研究ではこれらの研究を参考にし、利用する音響的特徴量などを設定した。

3. アンケートに依る聴取調査

3.1 調査方法

本調査ではまず、アンケートによる聴取評価を実施した。今回のアンケートでは、20代女性29名の単語音声を用いて、12名の聴者を対象として実施した。本調査ではTMW[5]の音源を用い、その中から母音から始まりアクセントによる個人差が少ないフレーズを選出した結果、女性の音声による「oosama」を採用することにした。

アンケートはWeb上のフォームで実施し、IPアドレス等の通信情報を一切記録しないことで匿名性を確保した。フォームでは聴者自身の性別・年齢・職業・趣味・好きな声の芸能人を入力してもらい、続いて各音源に対して、いい声だと思ったかを4段階(非常にそう思う、そう思う、あまり思わない、全く思わない)で評価してもらった。またその音声に関してどのような印象を受けたのか、Kretschmer[6]の提唱した性格の気質分類を参考にして「感受性が豊かである」「頑張りやである」「華やかで賑やかである」「強い自信がある」「冷酷である」「社会的で穏やかである」の6項目を定め、評価させた。

3.2 調査結果

4段階評価をそれぞれ、+2(非常にそう思う)、+1(そう思う)、-1(あまり思わない)、-2(全く思わない)の数値で割り当て評価値とし集計した。

これらをピアソンの相関係数にあてはめ相関を求めた。表1は「いい声」の評価値と特徴的であった項目の評価値との相関係数をまとめたものである。調査の結果から、「華やかで賑やかである」「社会的で穏やかである」という項目において強い正の相関が見られた。逆に、「強い自信がある」「冷酷である」という項目において相関は見られなかった。

表1. 「いい声」の評価値との相関関係

評価項目	相関係数	相関
華やかで賑やか	0.687	強い正の相関
社会的で穏やか	0.705	強い正の相関
強い自信	0.392	殆ど無相関
冷酷さ	-0.275	無相関

4. 音響的特徴量の解析

音声データの評価値と音響的特徴量との関係性について検証した。本調査ではMATLABを用いて特徴量を記録した。特徴量の抽出の過程において、周波数特性を強調させるための高域強調や窓関数による丸め処理を行った後に高速フーリエ変換を行っている。またフレーム分

割はフレーム長 25ms, フレームシフト長を 20ms で切り出している. 周波数については聴覚特性を考慮し, メル尺度を用いた. 取得した特徴量は基本周波数 [mel], スペクトル包絡上に見られるフォルマント周波数, フレーム数である. また今回取得したフレーム数はほぼ話速 [mora/s]と言い換えることが可能である.

4.1 基本周波数にみられる特徴

次の表 2 は高評価の音声(音声 9 と音声 27)と, 低評価の音声(音声 7, 音声 10, 音声 20, 音声 22)の基本周波数の平均とそれぞれの評価値を表したものである.

高評価のものと同評価のもの比べると, 基本周波数の平均値に差が見られることがわかる. これらと評価値との相関係数を求めたところ 0.74 となり, 正の相関が見られた. 今後の調査で理想となる基本周波数を追求したいと思う.

表 2. 基本周波数の平均値

	音声27	音声9	音声7	音声10	音声20	音声22
平均値	331.8	352.5	311.8	324.5	324.2	318.4
評価値	+22	+17	-16	-16	-17	-18

4.2 スペクトル包絡上にみられる特徴

スペクトル包絡に見られる特徴を捉えるため, 本調査ではある一定区間におけるフォルマントのピーク数の合計を求め, それらと印象評価値との関係をプロットした. その際, 評価値の分散が大きい音声は一部の聴者の趣向に左右されるためノイズとみなして外し, 評価値の分散の小さい音声のみ示した. 図 1(左)がその結果である.

ひとつ大きな外れ値が見られるが, 凡そ, フォルマントのピーク数が多いほど評価値が高くなる傾向にあるのが見て取れる.

4.3 フレーム数に見られる特徴

さらにフレーム数においても興味深い結果が得られた. 図 1(右)は印象評価値とフレーム数の関係をプロットしたものである. ある一定フレーム数を越えた途端に評価値が減少していることが分かる. これは「いい声」でない声を判定するひとつの指標になると考えられる.

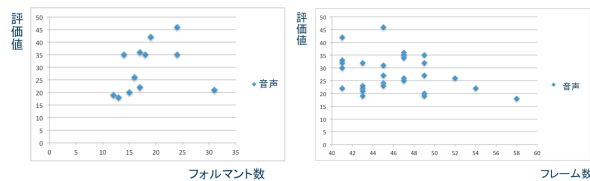


図 1. 印象評価値とフォルマント数・フレーム数の関係
(左)印象評価値とフォルマントのピーク数
(右)印象評価値とフレーム数

5. アプリケーションの提案

本研究では本調査結果に沿ったアプリケーションを用途に応じて 2 種類提案する. 本アプリケーションは自身の声質を知る手段の提供と, フィードバックを与えることによる自己啓発を目的とし, Android 環境で開発した.

5.1 周波数の特徴を捉えるアプリケーション

リアルタイムで声の特徴を捉えることができるアプリケーションである. 刻々とグラフィックが変化するため, トレーニングに利用できると考えられる.

入力音声に Visualizer クラスを用いてフーリエ変換する

ことで周波数分布を取得し, それら进行处理することで実装した. 図 2(左)は実装画面である. 図 2(右)の通り, 基本周波数を色相で, 音圧を大きさで, フォルマントのピーク数を形で表すようなグラフィックを形成させた. 一般的にフォルマントは声の豊かさと言われているため, フォルマントのピーク数が多く見られれば見られるほど丸みを帯びるデザインにした.

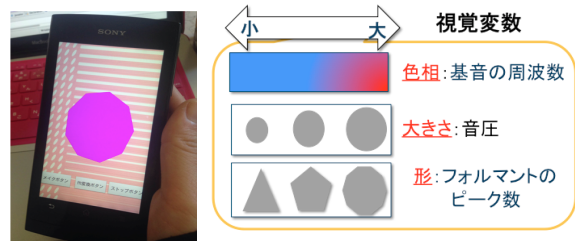


図 2. 「いい声」判定アプリケーション
(左)実装画面
(右)アプリケーションのデザイン

5.2 印象評価値を提示するアプリケーション

一定時間に入力された音声に対して「感受性が豊かである」「華やかで賑やかである」などの 6 項目で印象評価値を出力する. 3 章で得られた結果を反映しており, 診断ツールとして楽しむことができると考えられる.

TMW[5]の音源のフーリエ変換後の配列データを主成分分析で処理し, それらと入力音声との類似度からレーダーチャート方式で結果を表示するよう開発している.

6. まとめと今後の課題

本研究では, 聴取調査によってサンプル音声に対する印象評価を行い, その結果と特徴量の関係をもとに「いい声」を判定する独自のアプリケーションを提案した.

本研究の調査結果から, 比較的多くの人が「いい声」と感じる声には, 「華やかで賑やかである」「社会的で穏やかである」の項目で肯定的な評価がされており, これらと正の相関が見られた. また音響的特徴量との関係より, 「基本周波数の平均値」「単位時間毎のフレーム数」「フォルマント周波数のピーク数」が声の印象に影響を与えているのではないかと仮説が立てられた.

今後は, より大規模な聴取調査を行うことで, これらのシステムや結果を明確化させ, アプリケーションの実用を目指したい.

参考文献

- [1] A. Mehrabian: Silent messages, Wadsworth, Belmont, California (1971)
- [2] 粕谷英樹, 楊長盛:音源から見た声質, 日本音響学会誌 51(11), 869-875 (1995)
- [3] 矢島佳澄, 寛康明, 諏訪正樹:発声のメタ認知促進システム「いい声マイク」の提案, 情報処理学会インタラクティブ 2011 (2011)
- [4] 酒造正樹:スマートフォンを用いた気分の浮き沈みセンシングの一検討, 電子情報通信学会技術研究報告. MoNA, モバイルネットワークとアプリケーション 113(398), 89-91 (2014)
- [5] 音声音源コンソーシアム:東北大・松下単語音声データベース (TMW), <http://research.nii.ac.jp/src/TMW.html> (2011)
- [6] Ernst Kretschmer:Koperbau und Charakter, Springer, Berlin (1921)