

女性の声を例にした音響的特徴量と印象評価の関係性の調査

菅原衣織¹ 伊藤貴之¹

概要: 声色の印象が与える効果はコミュニケーションにおいて約 4 割を占めるという考え方がある。このように声の印象が日常生活に与える影響は大きいと考えられるが、どんな声がどんな状況でどんな個性を持つ他者にいい印象を与えるのかは不明瞭である。そこで本研究では聴取評価を行い、「いい声」がどのような音響的特徴量によって成立するのか調査した。調査の結果、「いい声」には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頂点数が一因するとの仮説が立てられた。

キーワード: 音響特徴量, 声質

Investigation of the Relationship of Impression Evaluation and Acoustic Feature Value as an Example the Voice of Women

IORI SUGAHARA¹ TAKAYUKI ITOH¹

Abstract:

There have been reports that voice tone covers approximately 40% of the impression in the human communication. This result suggests that the impression of voice may impact on the daily life. However, it is unclear what kind of voice give preferable impressions to any kinds of persons, and in any kinds of situation. Therefore, we conducted a questionnaire and investigated relationship of the acoustic features and "good voices". As a result of the investigation, we hypothesized that "good voices" are influenced by the average fundamental frequency, number of frames, and number of peaks of formant frequency.

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できるように、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、声の印象が自己の評価に与える影響は少なくない。

本研究では、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、滑舌は早口言葉の練習を積むことで良くなり、声の大きさは拡声器を使うことで補正が可能であるように、いくつかの要素は既存の手法を用いて改善される。そこで本研究では声質の印象に絞って議論を進める。

1.1 声質の定義

一口に声質と言っても定義が不明瞭になる可能性がある。そこで過去の研究における声質の定義を参照しながら、本研究における声質の位置づけを確認したい。Trger [2] は周辺言語の系統づけを試み、周辺言語を voice qualities(声質要素, 性状的要素) と vocalization(発声要素) からなるとし、性状的特徴をピッチ範囲・口唇・声門・調音・リズムなどの制御や共鳴・テンポなどとした。また Laver [3] は

¹ お茶の水女子大学大学院人間文化創成科学研究科

声質を「話し手が程度の差こそあれ、話している間中ずっと存在する特性で、発生された音声にほぼ永久的に存在する質」とした。これらを踏まえて 粕谷 [4] は声質を音声に含まれる個人的特徴と性状的特徴によってもたらされる聴覚的特質と定義している。

我々はアクセントの違いや滑舌など、話者の話し方が影響する個人的特徴をできる限り排除し、話者の骨格や声道の長さ、音の共鳴の仕方といった特徴を捉えたいと考えた。そこで本研究では、粕谷 [4] 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」と定義する。またこの定義に沿った研究が進められるよう、本研究で扱う音声データを、アクセントやイントネーションによる誤差が生じやすい文章単位の音声ではなく、アクセントや無声化音・鼻濁音を統一しやすい単語単位の音声もしくは単一音の音声に限定する。

1.2 本研究の方針

我々は「いい声」は一意に存在するものではないと考える。むしろ、聴者の個人性や感性、さらにはその声が使われる状況に応じて、さまざまな種類の「いい声」が存在するものと考えている。例えば、20代女性と60代男性が「いい声」だと思う声はそれぞれ異なっているのではないかと、また発表に向いている「いい声」と異性を相手にしたときの「いい声」も別の音響的特徴が見られるのではないかと仮定した。本研究では、音声データの音響的特徴量とそれに対する印象評価をもとに、聴者の個人性や使用状況を考慮した「いい声分布」なるものを追求する。本研究において聴者の個人性とは、性別・年代・職業・趣味などの個人的属性を指す。本報告では、あるフレーズの音声データを用いて、性格分析を参照した印象評価アンケートを実施し、その音響的特徴量と評価を集計した結果を示す。

2. 関連研究

2.1 「いい声」に関する研究

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発声のメタ認知促進システム“いい声マイク”の提案 [5] 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（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、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、モテ声診断 VQ チェッカー [6] というウェブサイトおよびスマートフォン用のアプリケーションでは、「声の高さ」「声の大きさ」「一音の長さ」「耳への入りやすさ」

「滑舌の良さ」の5つの項目から声の良さを定義している。

菅原 [7] は、いい声とは一意に定まるものではないと仮定し、倍音の特徴からユーザが自分の求める声の印象を目指せるような判定アプリケーションを開発した。この際に参照したのが尺八奏者の中村の著書 [8] である。中村は著書 [8] の中で、整数次倍音が強いとカリスマ性や明朗性、豊かさが印象付けられ、非整数次倍音が強いと情緒や親密感が印象付けられると述べている。この研究ではこれらの特徴を分析することで声色の印象を評価している。図 1 はアプリケーションの実装画面である。

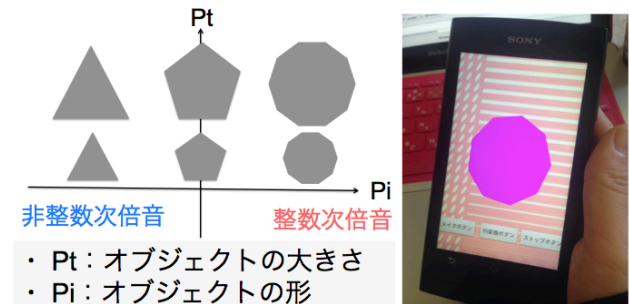


図 1 倍音による声の判定アプリケーション

上述の通り、文献 [5] ではいい声の判定基準として第1、第2フォルマント領域のパワー分散の低さを指摘し、文献 [6] では「声の大きさ」や「滑舌の良さ」などを指摘している。これらと本研究には一定の関連性を有するが、本研究では全体的な意味での声量には制約を加えず、またフォルマント領域といった特定の周波数領域に限定せず声質を分析し、いい声とは何なのかを追求するものである。

2.2 声質の表現方法に関する研究

木戸ら [9] は、通常発話の声質に関する日常表現語から統計的手法や反意語調査を経て、8対の声質表現対と反意語を持たない一つの声質表現語を抽出した。さらに聴取評価から音声の聴取印象の因子として「明瞭性因子」「迫力性因子」「美的因子」を抽出した。また木戸はその後の研究 [10] で、音声の特徴から人物を特定する音声モニタージュの可能性も示唆している。

児玉ら [11] は、声とイメージカラーの関係性を調査している。この調査では、「こんにちは」というフレーズについて、低く深みのある声・アナウンス声・のんびりした声・かわいい声・明るい声という6種類の20代女性の音声データを用意している。そして、かわいい声からはピンク、明るい声からはオレンジ・黄という、声から一定の色彩をイメージさせる傾向にあるという結果を導いている。

木戸ら [9] や児玉ら [11] の研究事例のような、音声を日常表現語や色で表すといった手段は、本研究においても有効であると考えられる。具体的には、本研究の調査結果を前提としてアプリケーションを開発する際に、これらの研

究と同様に音声を日常表現語や色で表現することが考えられる。

3. アンケートによる聴取評価

本研究では、「いい声」にはどのような音響的特徴量が寄与しているのかを導くために、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。

3.1 調査内容

本調査では音声コンソーシアムの音声コーパスに収録されたTMW[12]の音源を用いた。その音声データベース中から、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。母音から始まるフレーズにすることで、例えば子音の無声化や鼻濁音など、vocalization(発声要素)による印象の違いを排除する狙いがある。またTMW[12]には男性の音声によるフレーズも収録されているが、本調査では女性の音声のみを用いた。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以下の項目について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)の評価を入力してもらった。またその音声に関してどのような印象を受けたのか、Kretschmer[13]の提唱した性格の気質分類を参考にして項目を定めた。

- いい声である
- 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
- 社交的で穏やかである

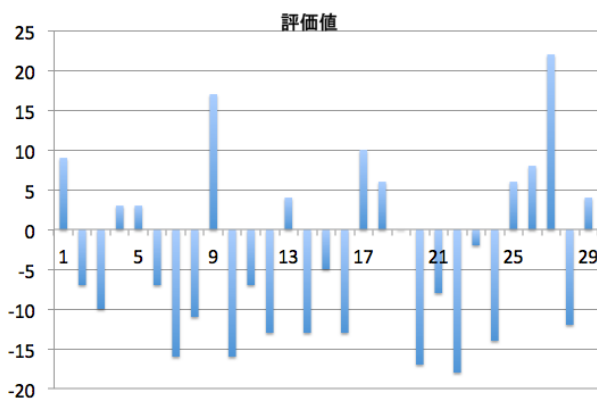


図2 評価項目：いい声である

表1 高評価の音声と低評価の音声の内訳

Table 1 A Details of Valuation.

	非常に (+2)	そう (+1)	あまり (-1)	全く (-2)	平均
音声 27	10	2	0	0	1.83
音声 9	7	4	1	0	1.42
音声 7	0	0	8	4	-1.33
音声 10	0	1	5	6	-1.33
音声 20	0	0	7	5	-1.42
音声 22	0	0	6	6	-1.5

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、集計した結果が図2である。また表1は、高い評価を得た2つの音声と、低い評価を得た4つの音声の評価の内訳人数とその評価値の合計を表したものである。評価が明らかに偏っていることから、高評価の音声と低評価の音声を比較することで、大衆にとって「いい声」と感じる音声の傾向を得られるのではないかと考える。

次項から29名分の音声のうち、特に高い評価を得た2つの音声、そして低い評価を得た4つの音声、これら6つの音声を比較して検証を進める。また、評価が良いとも悪いともいえない特異な評価値を示した5つの音声についても検証する。

3.2.1 高い評価を得た音声

質問項目「いい声である」において高評価を得た音声の、その他の6項目での評価値を図3にまとめた。いずれも「感受性が豊かである」「華やかで賑やかである」「社交的で穏やかである」の項目で一定以上の評価を得ていることが読み取れる。また反対に、「冷酷である」の項目が低評価であることも同様の傾向として捉えることができる。この結果より「いい声」であると評される声には、ポジティブな形容動詞で表せるものが多いこと、それに加えて自信を感じさせることがより良いのだとわかる。

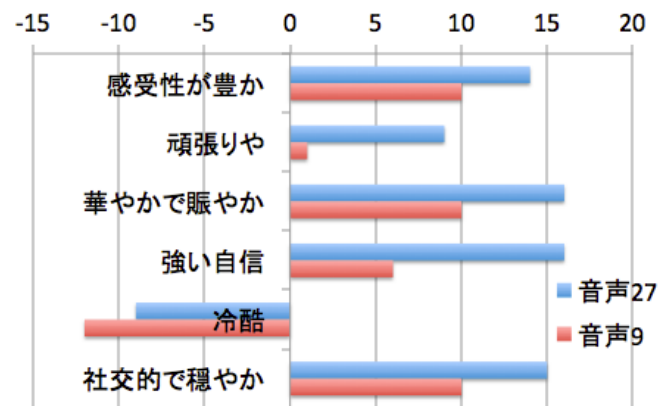


図3 高評価音声のその他の項目での評価値

3.2.2 低い評価を得た音声

質問項目「いい声である」において低評価を得た音声の、

その他の6項目での評価値を図4にまとめた。高評価と評された音声において高い数値を占めていた3つの項目の評価が、総じてあまり良くないことが見てとれる。また音声ごとに顕著な値が見られる項目がそれぞれ異なることから、低評価を得る要因には多様な傾向が混在している可能性が示唆される。ここで注目したいのは「強い自信がある」の項目である。他の要素も複雑に絡んでくるとは思われるが、「強い自信」が有り過ぎるのも無過ぎるのも低評価の一因になっている可能性がある。

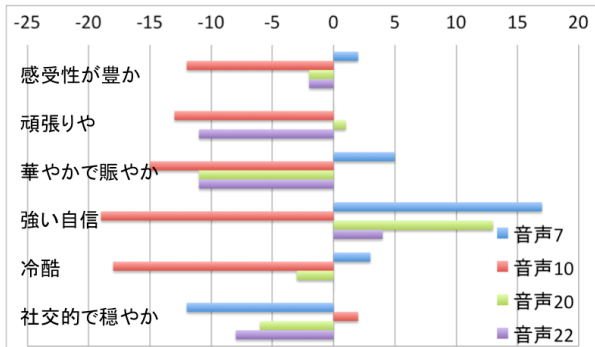


図4 低評価音声のその他の項目での評価値

3.2.3 評価が分かれた音声

質問項目「いい声である」において評価が大きく分かれた音声(音声6, 音声13, 音声15, 音声25, 音声29)の詳細を表2に記す。

我々は本報告の冒頭で、「いい声」は一意に存在するものではなく、聴者の個性や感性、使用状況に応じて複数存在するものと仮定した。もし「いい声」に関係する要素が一意であるならば、評価も一定になるはずである。しかし今回の調査では29名分の音声のうち5つの音声において評価のばらつきが見られた。

その中でも特に、音声6について面白い結果が出た。12人中10人が「いい声ではない」と評価しているのに対し、12人中2人だけ「非常にいい声だと思う」と回答している。その他の6項目の内訳を確認してみると、この2人のみ「感受性が豊かである」「頑張りやである」「社交的で穏やかである」の項目で高い評価を示していた。この2人は後述する図6の聴者1と聴者4にあたる。

表2 評価が分かれた音声の内訳
Table 2 A Details of Valuation.

	非常 (+2)	そう (+1)	あまり (-1)	全く (-2)	平均
音声6	2	0	9	1	-0.58
音声13	2	6	2	2	0.33
音声15	1	3	6	2	-0.42
音声25	3	5	3	1	0.5
音声29	3	4	4	1	0.33

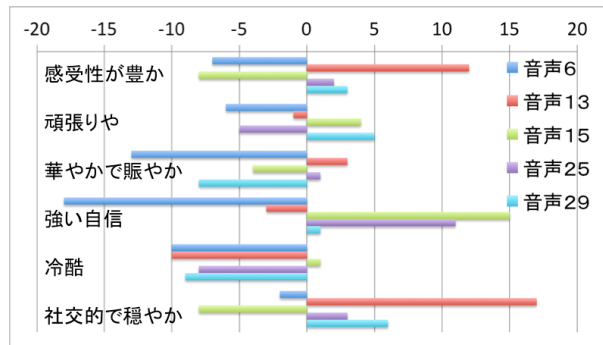


図5 評価が分かれた音声のその他の項目での評価値

参考までに、その他の6項目での評価値を図5にまとめた。評価項目および評価値が音声ごとに全く異なる分布を示していることから、聴者の個性によって評価の違いが出たと考えられる。

これらの特定の音声に対して聴者によって大きく評価が分かれた理由として、聴者の個性が関係したのではと我々は考えた。そこで、これらの音声で「非常にそう思う」「全く思わない」と顕著に回答していた聴者に的を絞って考察した。その中でも特に興味深い内容を図6に示す。

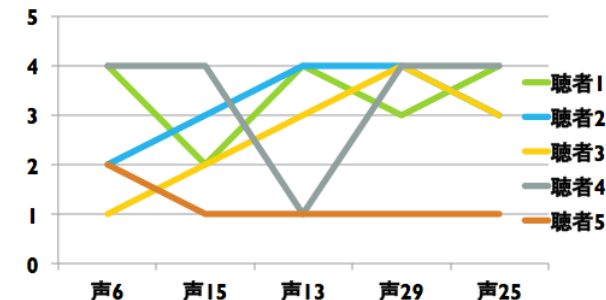


図6 個性と評価値の関係

このグラフからも、この5つの音声に関しては聴者に応じて極端に評価がわかれていることから、一般的に大多数の人が「いい声」と共感する声の他に、個人の好みによって「いい声」と判断される声があると考えられる。

例えば、聴者1はこの5つの音声に限らず全体的にどの音声に関しても肯定的な評価をしており、4段階(+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))の平均値が0.83となっている。つまり「いい声」と感じる声の許容範囲が広いと考えられる。これに関して、今回のアンケートでは個性の取得のために「好きな声の芸能人」や「趣味」を選択する欄も設けていたが、聴者1は、12名中唯一「好きな声の芸能人」を二人以上挙げていた。現時点では根拠とするには弱い要素だが、「いい声」と感じる範囲が広いことと、「好きな声の芸能人」の列挙数にも今後着目したい。聴者1とは逆に、聴者5は全体的に否定的な評価をしており、4段階の平均値は-1.24である。つまり「いい声」と感じる声の許容範囲が狭いと考えられる。これらの5つの音声の評価が分かれた要因の一つに聴者1と聴者5の存在は大きい。

もう一人の面白い聴者として、聴者4について議論した

い。図6を見てもわかるように、聴者4の評価は極端である。図7は聴者4の回答の一部である。聴者4は他の聴者とは異なり、「強い自信がある」と「冷酷である」の項目で評価の高いものを好む傾向にあった。この結果は「いい声」の一つの趣向として今後のデータ集めの際も注目したい。

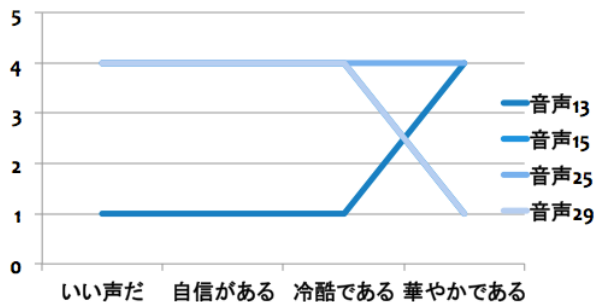


図7 聴者4のいい声とその他の項目の回答

今回のアンケートでは、一般的な好みの傾向、および特定の聴者にのみ該当する興味深い回答傾向を得られたが、個人性と評価値との相関性を論ずるには聴者数が十分ではない。本調査で新たに得られた仮定を実証するために、大規模なアンケート調査を実施する予定である。

4. 音響的特徴量との関係性

次に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する。本調査ではMATLABを用いて特徴量を記録した。周波数については聴覚特性を考慮し、メル尺度を用いた。取得した特徴量は以下の通りである。

- 基本周波数 [mel]
- スペクトル包絡
- フォルマント周波数

4.1 基本周波数にみられる特徴

基本周波数とは、音声の周波数成分の中で最も低い周波数成分のことである。次の表3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を取ったものである。参考までに項目「華やかで賑やかである」の評価値も記す。

高評価のものと低評価のものを比べると、基本周波数の平均値に差が見られることがわかる。確認が得られたわけではないが、基本周波数の平均が325Hz~330Hzの周辺が評価の分かれ目になっている可能性がある。今後の調査ではこの330Hz前後での評価の差にも注目していきたい。今回のデータは被験者がすべて20代女性であったが、これが他の年代や性別だとどのような結果になるのか非常に興味深い。

また、基本周波数はベースとなる音高を表すため、単純に基本周波数が高いほど華やかで賑やかな印象を抱くのかと想定していたが、華やかさや賑やかさとの相関はあまり

見られなかった。

表3 基本周波数の平均

Table 3 Average of Fundamental Frequency .

	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
華やか	+16	+10	+5	-15	-11	-11

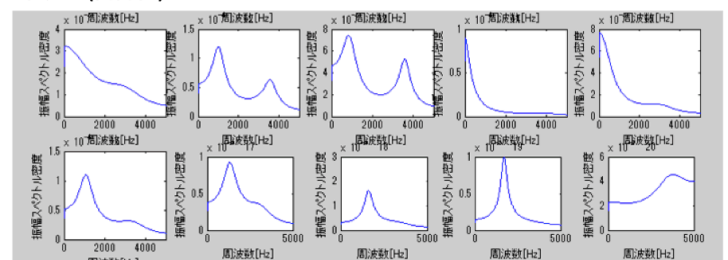
4.1.1 スペクトル包絡にみられる特徴

スペクトル包絡とは、高速フーリエ変換によって求められる振幅スペクトルの大きな周期性を表したものである。スペクトル包絡を知ることで、周波数の特性を捉えることができる。また、フォルマント周波数とは、スペクトル包絡において表れるピークの周波数のことである。図8は、アンケート結果より最も高い評価を得た音声27(図8(上))と、最も低い評価を得た音声22(図8(下))のスペクトル包絡をプロットし、その一部分を比較したものである。

同じ単語を発声していることから、似たようなスペクトル包絡を描くと思われたが、フレームの2つめ以降で大きな変化が見られる。高評価だった音声27はフォルマント周波数のピーク値がはっきりと二つ目視できるのに対して、音声22は一つしか見受けられない。

また今回の図ではわかりにくいですが、実はフレーム数にも大きな違いがある。音声27のフレーム数に対して、音声22は約1.3倍のフレーム数がある。これは音声22が他の音声に比べて間延びした発声になっているということである。以上のことから、フォルマント周波数の頂点数、及びフレーム数が評価を分ける一因になっているのではないかと考えられる。

音声27(高評価)



音声22(低評価)

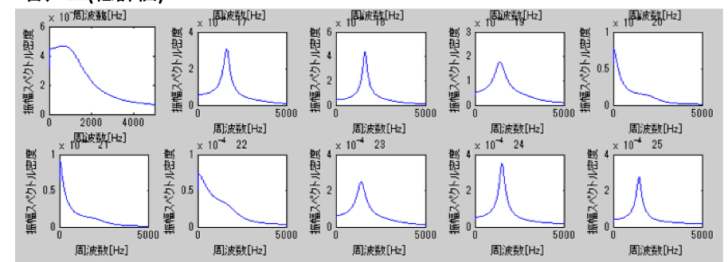


図8 高評価の音声のスペクトル包絡の一部

5. 判定アプリケーション

菅原ら [7] は、高周波数領域の倍音部分に着目して、声の判定アプリケーションを開発した。音楽理論などにおいて倍音とは、楽音の音高とされる基本周波数に対して、2以上の整数倍の周波数を持つ音の成分を指す。以上のように定義される狭義の倍音を「整数次倍音」と呼ぶのに対して、基音に対して非整数倍の周波数を持つ音の成分は「非整数次倍音」と呼ばれており、中村は著書 [8] で、整数次倍音はカリスマ性や明朗性が印象付けられ、非整数次倍音は情緒や親密感が印象付けられると述べている。菅原ら [7] のアプリケーションはこれらの知見にもとづいて、声に含まれる整数次倍音および非整数次倍音の強さを可視化することで、声質を直感的に表現している。

今回の調査から我々は、「基本周波数の平均値」「フレーム数」「フォルマント周波数の頂点数」が「いい声」の評価値の高さに影響するのではないかという仮説を立てた。この仮説にしたがって今後、菅原ら [7] が開発したアプリケーションの拡張版を開発する予定である。具体的には、倍音部分の周波数分布だけを判定に用いるのではなく、新たな指標として単位時間ごとの基本周波数の平均値やフォルマントの推定などの要素を組み込みたい。

6. 今後の課題

本調査では、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アンケート調査を実施した。足掛けとしていい結果を得られたが、これらは十分な被験者数であったとはいえない。そこで結果を基にアンケート内容に修正を加え、より多くの評価や個人性の取得を目的とした、大規模なアンケートの実施を検討している。また今回のアンケートでは聴者より「聴く順序で評価が変わりそうだ」「どういう場面におけるいい声なのかわからない」といった意見をもらった。そこで、次に実施するアンケートでは、二つの音声を比較し優劣を選択させたり、どんな場面で有用かを回答できる内容にしたいと考えている。

以上に関連して、今後のアンケートで更に大量のデータが獲得できると考えられる。そこで、もっと的確で柔軟な知見を得るために、以下の2手法による分析を検討したい。

- 聴者による印象や評価に基づいて音声を分類するための決定木の生成。
- 聴者による印象や評価と音響的特徴量との関係のモデル化。例として応答曲面法を適用する。

これらの手法を用いることで、将来的には「いい声分布」に相当する概念を導き出したい。

7. まとめ

本研究では声質の観点から「いい声」を追求している。その前提として我々は、「いい声」は一意に定まらず、音

響的特徴や聴者の個人性などの要素が絡み合うことで多様な「いい声」が存在するものであると仮定した。本報告ではその仮説を検証するために、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アンケート調査を実施した結果を示した。そして、それらの評価結果と音響的特徴量との関係をまとめた。アンケートの結果から、大多数の人が「いい声」と感じる声には、「感受性が豊かである」「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いるという共通点があることがわかった。逆に声の評価が低い理由には、多様な傾向が混在していることがわかった。そして音響的特徴量との関係を調査し、「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頂点数」が声の印象に影響を与えているのではないかという仮説を立てた。今後は、大規模なアンケートを実施することで本調査で得られた仮説を実証したい。またその仮説にしたがって、「いい声」判定アプリケーションの拡張版を開発したい。

参考文献

- [1] A. Mehrabian: *Silent messages*, Wadsworth, Belmont, California (1971).
- [2] G.L.Trager: *Paralanguage: A first approximation*, *Stud Linguist.*13,1-12(1958)
- [3] J.Laver: *The Phonetic Description of Voice Quality*, Cambridge University Press, Cambridge(1980).
- [4] 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)
- [5] 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シオン* 2011 (2011).
- [6] 声総研: モテ声診断 VQ チェッカー, <http://www.koesouken.com/vqchecker/>(2001).
- [7] 菅原衣織, 伊藤貴之: 倍音分析によるいい声作りの支援システムに向けて, *音学シンポジウム 2014* (2014).
- [8] 中村明一: 倍音音・ことば・身体の文化誌, 春秋社 (2010).
- [9] 木戸博, 粕谷英樹: 通常発話の声質に関連した日常表現語の抽出, *日本音響学会誌*,55(6),405-411(1999)
- [10] 木戸博, 粕谷英樹: 音声が入る話者の特徴情報の記憶, *音声研究*,13(1), 4-16 (2009)
- [11] 児玉朋江, 齋藤美穂: 声色のイメージカラーの分析, *日本色彩学会誌*,29-suppl,42-43 (2005)
- [12] 音声音源コンソーシアム: 東北大・松下単語音声音声データベース (TMW), <http://research.nii.ac.jp/src/TMW.html>(2011)
- [13] Ernst Kretschmer: *Koperbau und Charakter*(1921)