

歌唱音を含む音楽音響信号を対象とした 歌唱者性別判定システムの構築

勝呂 明央[†] 三浦 雅展^{††}

近年、音楽音響信号から計算機によって歌唱音のみの抽出を行なう研究が数多く報告されている。しかし、現在流通している音楽音響データは歌唱音のみではなく、伴奏音が混在した状態で提供されることがほとんどであるため、計算機で自動的に歌唱音の特徴を抽出することは一般に困難である。ここでは、スペクトル減算及び特定の周波数帯域におけるピーク値を算出することで歌唱音から特徴を抽出し、歌唱音の抽出を試みている。その結果、ある周波数帯域におけるパワーが歌唱者の性別を判定する特徴量として用いることができる可能性が示唆されている。提案手法から得られた特徴量を基に機械学習を行なった結果、85%の推定率が確認されている。また、得られた特徴量からシステムを構築しており、16bit/44.1kHz フォーマットの音楽音響信号から自動的に歌唱者の性別を判定できることが確認されている。

Realization of a gender estimation system for singer of musical singing audio

Akio Suguro[†] and Masanobu Miura^{††}

Numerous studies has reported to extract a singing sound from acoustic signal. However, the acoustic signal is a mixed audio with singing and accompaniment sounds. Extraction of singing sound automatically by computer is usually difficult since it is not easy to separate them. This study proposed a new methodology to extract singing sound by calculating spectral subtraction and peaks from specific frequency band. Result found that power of a specific frequency band is useful to determine the gender of the singing sound. The study conducts a machine learning based on obtained feature values. The result confirmed that estimated accuracy of it is almost 85%. The study also construct a gender estimation system for singer of musical singing audio from obtained feature values. The system is possible to use on conventional computers and to accept usual CD-quality audio (16bit/44.1kHz) so that it calculates the possibility of the gender of singer as woman or man with calculated possibilities.

[†] 八戸工業大学院 工学研究科 電子電気・情報工学専攻

Doctor of Engineering Program in Electronic, Electrical and Information Engineering, Graduate School of Hachinohe Institute of Technology

1. はじめに

近年、提供されているポピュラー楽曲をはじめとする多くの楽曲には歌唱音が含まれている。歌唱音が含まれる楽曲では、歌唱音がメロディパートとして構成されることが多く、歌唱音はその楽曲における中心的な役割を担っている場合が多い。また、通常歌唱音は伴奏音と混在された状態で提供されていることが多い。歌唱音と伴奏音が混在した状態では、歌唱音と伴奏音が同じ周波数帯域に混在する場合があるため、歌唱音のみを自動的に抽出することは一般的に困難である。もし歌唱音のみを自動的に抽出できれば歌唱音の特徴、例えば歌唱者の性別などを判定することができ、ユーザーに自動的に楽曲を推薦するシステムなどに有用である。

この目的に従い、歌唱音のみを楽曲から自動的に抽出する試み[1,2]が報告されている。先行研究[1]では、歌唱音の存在する周波数帯域を強調することによって、歌唱音と伴奏音が混在した状態の音声から歌唱音の抽出を試みている。しかし、強調をする際に歌唱音のみではなく伴奏音も強調してしまい、歌唱音が抽出できなかつたと報告されている。そのため、単純に周波数成分を強調及び抑制する手法は妥当ではなく、別の手法での検討が必要であると考えられる。先行研究[2]では、歌唱音を統計的モデル化やビタビ探索などを用いて歌唱音の抽出を行っており、高い推定精度で推定できることが報告されている。ただし、機械学習やモデル化をすることで歌唱音の抽出の精度は高いものの、歌唱音の抽出までに多くの時間を要したり、データベースの作成など多くの手間と時間を要するなどの問題がある。歌唱音の自動採譜などでは歌唱音の抽出には精度も要求されるが、大量の楽曲データから歌唱音を抽出し、システムの構築などに利用するならば、短時間で計算可能であることも必要となる。また、先行研究[2]の手法では歌唱音の基本周波数に対する推定精度は高いものの、他の特徴、例えば歌唱者の性別などの歌唱者に対する特徴量を得ることができない。もし、歌唱者に対する特徴量が得られれば、歌唱者の性別などを楽曲にタグ付けするうえで有用であり、大量な楽曲を精査する上で重要な情報である。すなわち、高品質な歌唱音（ここでは歌唱音の基本周波数のみでなく、歌唱者の性別の特徴が含まれる音源を高品質な歌唱音とする）の抽出が望ましい。しかし、音楽音響信号からその高品質な歌唱音を高速に抽出する手法は存在せず、歌唱者の性別を判定するなど歌唱者の特徴量を得られる実用的なパラメータなどは報告されていない。

本報告では、短時間で高品質な歌唱音を抽出するために、スペクトル減算を用いて元音源から伴奏音を抑制する。次に、特定の周波数帯域におけるパワーのピーク値を算出し、そのピーク値を出力することで、音楽音響信号から高品質な歌唱音の抽出を

^{††} 八戸工業大学 工学部 システム情報工学科

Dept. of System and Information, Faculty of Engineering, Hachinohe Institute of Technology

試みる。次に評価実験を行ない、本報告で提案する手法で高品質な歌唱音が抽出できる可能性を客観的に評価する。その後、抽出された歌唱音が男性の歌唱と女性の歌唱において相違点が見られるかを検証し、歌唱者の性別判定に対する特徴量が得られるか否かを検証する。次に、機械学習に提案手法で得られた特徴量を用い、歌唱者の性別判定に対する特徴量として用いることができるか否かを検証する。最後に実際に得られた特徴量からシステムを構築し、音楽音響信号から歌唱者の性別を自動的に判定するシステムを構築する。

2. 先行研究について

2.1 歌唱音の強調を試みた研究

先行研究[1]は歌唱音の抽出を目的としており、歌唱音を抽出するために歌唱音の存在する周波数帯域を強調する手法を試みている。その結果、精度を定量的には示していないものの、特定の周波数帯域を強調することにより、歌唱音が強調されると同時に伴奏音も強調してしまうと報告されている。歌唱音と伴奏音が混在した状態で特定の周波数帯域を強調してしまうと、計算機は歌唱音を判別して歌唱音のみを強調することはできずに伴奏音も強調してしまうため、結果的に歌唱音の抽出は難しくなってしまう。よって、特定の周波数を強調するという手法以外の別の手法での検討が必要であると考えられる。

2.2 歌唱音のモデル化と機械学習を用いた研究

先行研究[2]も本研究と同様に歌唱音の抽出を目的としており、歌唱音の音響的特徴を混合ガウス分布でモデル化を行ない、歌唱音と歌唱音以外をモデル化している。その後、歌唱音の基本周波数を得るためにビタビ探索に基づくアルゴリズムで歌唱音の抽出を行なっている。その後、100曲に対する評価実験を行なっている。その結果、平均正解率は81.1%であり、高い精度で推定できると報告されている。しかし、モデル化や機械学習などを用いるため、大量の楽曲に基づくデータベースを構築している。そのため、単一の楽曲から歌唱音を抽出するには多くの時間を要すると考えられる。例えば音楽配信サービスなどでは、大量の楽曲から歌唱音を抽出する必要がある。そのため、単一の楽曲に対して比較的短時間で歌唱音を抽出できることが望ましい。すなわち、より短時間で歌唱音を高い精度で抽出できる別の手法が望まれる。

2.3 本報告の方針

先述したように、本研究では孤立歌唱音の抽出を目的とせず、歌唱音の特徴の高速な抽出を目的とする。そこで、元となる音源から伴奏音成分を抽出する。次に元の音源から抽出した伴奏音成分をスペクトル減算し、歌唱音成分のみを強調する。その後、

特定の周波数帯域におけるピーク値を算出し、歌唱音を抽出する。この処理は先行研究[1]と異なり比較的短時間で行なうことができると期待される。次に評価実験を行ない、歌唱音からの性別判定可能性を検討する。その後、提案手法で得られた特徴量を基に機械学習を行ない、その特徴量の妥当性を検証する。最後に実際にシステムを構築し、音楽音響信号から自動的に歌唱者の性別を判定するシステムの構築を行なう。

3 本報告で提案する手法

3.1 伴奏音成分の抽出

提供されている歌唱音を含む多くの楽曲ではステレオ音源の場合、歌唱音は左右のチャンネルから等しい音圧レベルで出力されるようにミキシングされていることが多い。また、伴奏音は歌唱音やその他のある伴奏音と周波数成分が重なってしまい、歌唱音が同時マスクングに晒されないように左右のチャンネルの片方から、または片方に寄った音圧レベルで出力されるようにミキシングされていることが多い。本報告では、このミキシングでの性質を利用する。まず、元の楽曲における左チャンネルと右チャンネルの音を得る。その後、左のチャンネルの振幅値から右のチャンネルの振幅値を減算する。この処理によって、左右のチャンネルから等しい音圧レベルの等しい音が抑制される。すなわち、歌唱音成分が消され、伴奏音成分が抽出できる。

3.2 歌唱音成分の強調

次に3.1で抽出した伴奏音成分を用いて元音源から歌唱音を強調する。具体的には、元音源の左チャンネルと右チャンネルを波形合成し、モノラル音源を作成する。その音源から3.1で作成した伴奏音成分をスペクトル減算する。この処理によって、元音源の周波数成分から、伴奏音成分のみが抑制される。すなわち、元音源から伴奏音成分が抑制され、歌唱音成分のみが強調された音源が作成できる。

3.3 歌唱音の抽出

歌唱音が強調された音源から特定の周波数帯域におけるパワーのピーク値を算出することによって、歌唱音の抽出を試みる。この根拠は、歌唱音は他のパートと比べてより強調して含まれていることが多い、という仮定に基づく。具体的には、伴奏音成分がスペクトル減算により伴奏音が抑制された音源のパワーのピーク値は歌唱音であると考えられる。すなわち、伴奏音が抑制された音源のパワーのピーク値を算出すれば、歌唱音が抽出できると期待される。ここで、このような手法で歌唱音を抽出した例はなく、この手法で得られるピーク周波数が歌唱音として抽出できるかは不明である。そのため本報告では、試験的に約624 Hzを下端とする幅約214 Hzの周波数帯

域を考え、それを約 214 Hz ずつずらした 4 つの周波数帯域からピーク値を算出する。その後、そのピーク値を出力することで歌唱音の抽出を試みる。この手法では基本周波数だけでなく基本周波数の倍音も抽出できる可能性が期待できる。すなわち、単純な音高情報のみではなく、性別判定が可能な特徴の抽出が期待できるため、楽曲のタグ付けなどに用いる特徴量の算出が期待できる。

4 評価実験

4.1 評価実験の目的

本報告で提案する高品質な歌唱音を抽出する手法を述べたが、実際に提案手法でどれほどの歌唱音を抽出できているかは不明である。そのため、単純に歌唱音を提案手法で抽出できているかを客観的に評価する必要がある。そこで本報告では、実際に提案手法で抽出した歌唱音を聴取者に聴取させ、歌唱音の抽出度を評価させる実験を行った。

4.2 実験刺激

実験刺激には男性のみの歌唱音を含む楽曲 10 曲と女性のみの歌唱音を含む 10 曲の全 20 曲を用いた。楽曲リストを表 1, 2 に示す。表 1 は男性の歌唱からなる楽曲リスト、表 2 には女性の歌唱からなる楽曲リストである。男性と女性の歌唱音の差を検証するために、男性と女性の両方が歌唱している楽曲は実験刺激には選出しなかった。また、用いる楽曲はジャンルなどを限定せず、無作為に選出した。また、実験刺激の時間長は 30 sec とし、楽曲が始まってから最初に歌唱が始まった地点から 30 sec を抽出し、実験刺激とした。

4.3 実験環境

聴取者は本学に在籍する 10 名 (男性 9 名, 女性 1 名, 平均年齢 22.1, 標準偏差 1.30) で行なった。実験場所は八戸工業大学システム情報工学専門棟 I108 にて行ない、実験機材は、PC は NEC 社製の PC-GN257AGG3、ヘッドフォンには SONY 社製の MDR-1ADAC を用いた。なお、聴取者には椅子に座った状態でヘッドフォンを装着し、実験刺激を聴取させた。

表 1 男性の歌唱からなる楽曲リスト

ID	楽曲名	アーティスト名
1	Goodbye To Say Hello	HEY-SMITH
2	Gorilla	BRUNO MARS
3	KILLING ME SOFTLY WITH HIS SONG	平井堅
4	My Instant Song	MONOYES
5	ROCKIN'OUT	SPYAIR
6	STAY TUNE	Suchmos
7	アイネクライネ	米津玄師
8	いとしのエリー	サザンオールスターズ
9	ブルーバード	キリンジ
10	田園	玉置浩二

表 2 女性の歌唱からなる楽曲リスト

ID	楽曲名	アーティスト名
1	Catch the Moment	LiSA
2	Magia	kalafina
3	サラバ愛しき悲しみたちよ	ももいろクローバーZ
4	プレイバック Part2	山口百恵
5	ミカヅキ	さユリ
6	君と夏フェス	SHISHAMO
7	真夏の太陽	大原櫻子
8	未来へ	Kiroro
9	涙そうそう	夏川りみ
10	贅沢ないいわけ	パスピエ

4.4 提示方法

まず聴取者に、「最初に処理が加えられていない音源が 30 秒、その音源から歌唱音を抽出した音源が 30 秒流れます」と教示し、最初に歌唱音を抽出する前の音源を 30 秒呈示し、その後その音源から歌唱音を抽出した音源を呈示した。聴取者には 1 から 7 の 7 段階で評価させ、数字が大きいほど歌唱音が抽出できていると評価させ、提案手法でどれだけの歌唱音が抽出できているかを検証した。

4.5 評価実験の結果

男性のみの歌唱音を含む楽曲に対する評価の平均値を図 1 に示す。また、女性のみの歌唱音を含む楽曲に対する評価の平均値を図 2 に示す。縦軸は全被験者の回答に対する平均値、横軸は楽曲 ID を表す。また、図 1,2 内のティックは 95%信頼区間を表す。図 1 と図 2 の平均値を比較した結果、女性のみの歌唱音を含む楽曲に対する評価の方が 1.3 ポイント高かった。すなわち、本報告で提案した手法では、女性のみの歌唱音に対する楽曲の方が推定精度が高いことが確認できる。また、図 1 と図 2 の結果に対して t 検定を行なったところ、5%有意水準 ($p=0.033$) で有意であった。すなわち、男性と女性の歌唱音を含む楽曲に対する歌唱音抽出に対して差があり、女性のみの歌唱音を含む場合の方が評価が高いため、本報告の提案する手法は女性のみの歌唱音を

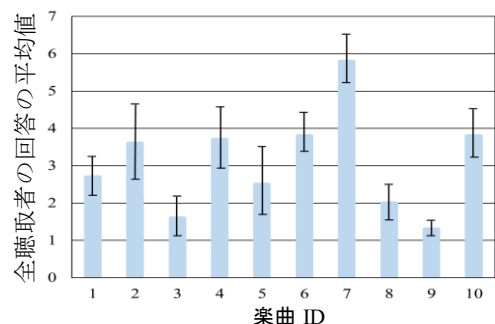


図1 男性のみの歌唱音を含む楽曲の評価

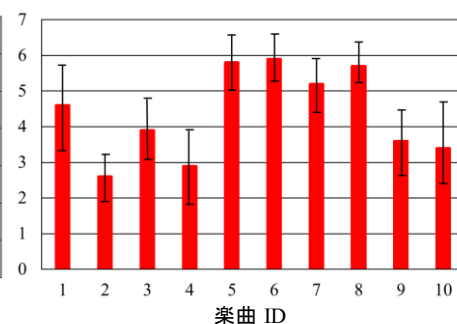


図2 女性のみの歌唱音を含む楽曲の評価

含む楽曲に対して有効であることが確認された。

5. 提案手法の処理時間について

4.2に示した表1及び表2の楽曲を用いて歌唱音を抽出するまでの時間を計測した。その結果、全20曲に対して歌唱音を抽出するのに要した時間の平均は10.689 secであった(なお、いずれの音源ファイルも44.1kHz/16bitのWAVE形式、平均時間長は264.24 sec, Corei7 4.00 GHz, Windows10 環境下で実施)。この処理時間であれば、システム化に用いるうえでも有用な処理時間であると考えられる。

6. 検証実験

6.1 検証実験の目的

先行研究[3]では、女性と男性の歌唱では基本周波数が多少異なると述べている。先行研究[3]でその差異は、主に音高や基本周波数であると述べている。そのため、本報告の提案手法で歌唱音を抽出した場合、正しく歌唱音が抽出できれば、男性の歌唱音と女性の歌唱音を抽出した際に差異が見受けられるはずである。その差異を機械学習に用いれば、歌唱者の性別を判定する際の重要なパラメータとなることが期待できる。そこで、男性のみの歌唱音を含む楽曲と女性のみの歌唱音を含む楽曲を用いて、男性と女性の歌唱音を抽出した際の差異を表す特徴としての妥当性を検証した。

6.2 実験刺激

実験刺激には4.2に示した表1及び表2の合計20曲を用いた。また、評価実験では30 secを切り出して用いたが、検証実験では切り出さずに音源の全てを用いた。

6.3 検証実験の結果

全実験刺激から、3で述べた手法により算出したパワーをそれぞれの周波数帯域で算出した。その後、男性ボーカル抽出音の平均音響パワーと、女性ボーカル抽出音の平均音響パワーを10曲にわたって比較し、それらの平均をdB値として算出した。算出手順を式(1)に示す。

$$Ave = 10 \log_{10} \left(\frac{\sum_{i=1}^{10} \sum_{j=1}^a \frac{1}{10} x_w}{\sum_{i=1}^{10} \sum_{j=1}^b \frac{1}{10} x_m} \right) \cdots (1)$$

x_w は女性のみの歌唱音を含む楽曲に対して抽出したパワー値、 a はそのピーク値に対するID、 x_m は男性のみの歌唱音を含む楽曲に対して抽出したパワー値、 b はそのピーク値に対するIDを表す。式(1)で算出した結果を図3に示す。縦軸はAve[dB]、横軸はピーク値を抽出した周波数帯域を表す。図3より、1054 Hzから1268 Hzにおいて1.85 dBの差が見受けられる。男性と女性の歌唱音を本報告の提案手法で抽出した場合、1054 Hzから1268 Hzにおいて差が見られた。すなわち、1054 Hzから1268 Hzのパワー値を特徴量とし、機械学習のパラメータとして用いることで、歌唱者の性別判定をするうえで重要なパラメータになると考えられる。また、1269 Hzから1483 Hzのパワー値では男性と女性の歌唱において差は見られなかった。本報告の提案した手法では試験的4つの周波数帯域を提案したが、歌唱者の性別を判別する特徴量としては1269 Hzから1483 Hzの周波数帯域は用いる必要がないものと考えられる。

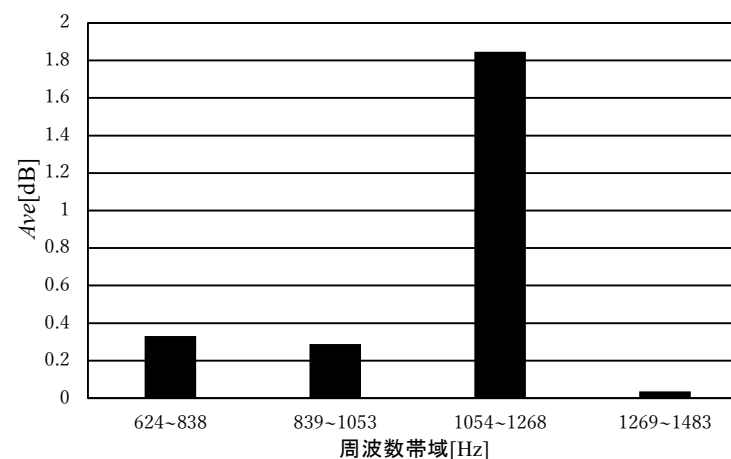


図3 男性ボーカル抽出音に対する女性ボーカル抽出音の音響パワーの比較結果

7. システムの構築

7.1 システムを構築する目的

6.3 より本報告で提案した手法で得られる値は男性の歌唱であるか、それとも女性の歌唱であるかを判別できる特徴量になると考えられる。そこで本報告では、実際に機械学習を用いて提案手法より得られた特徴量をパラメータとして用いることで、音楽音響信号から歌唱者の性別を判定することを目的とし、検証を行なった。

7.2 システムに用いるパラメータセット

用いる機械学習法は、本報告と同様に音楽音響信号を対象としている先行研究[4]で用いている Logistic regression を用いることとした。その理由として、先行研究[4]でシステム化が容易と述べられているためである。また、本報告では先述したパラメータに加え、提案手法より算出した 624 Hz から 838 Hz, 839 Hz から 1053 Hz, 1054 Hz から 1268 Hz のパワー値の平均、標準偏差、尖度、歪度をそれぞれ算出した 12 パラメータを用いる。1269 Hz から 1483 Hz は 6.3 で先述したが、男性の歌唱と女性の歌唱で差が見られないため、パラメータセットの選出から除外した。しかし、用いるパラメータは本報告で提案したパラメータのみでは少ない可能性が示唆されるため、先行研究[4]で用いられている 51 種類のパラメータか 35 種類のパラメータを選出し、本報告のシステムのパラメータとして用いる。以下に用いる音響パラメータを示す。

- ・スペクトルセントロイド
- ・スペクトルフラックス
- ・スペクトルフラットネス
- ・スペクトルロールオフ
- ・バンド幅 (周波数帯域幅)
- ・RMS (Root Means Square)
- ・テンポ

以上 7 類の音響パラメータに対し、平均、標準偏差、尖度、歪度、回帰直線の傾きをそれぞれ算出し、計 25 のパラメータとして用いる。すなわち、用いるパラメータの総数は 7 (音響パラメータ) × 5 (統計量) の計 35 パラメータを用いる。すなわち、本報告のシステムに用いるパラメータは 12 (提案パラメータ) + 35 (音響パラメータの総数) の合計 47 パラメータをパラメータセットとして用いる。教師データには 4.2 で用いた 20 曲を用いる。

7.3 クローズドテストによる検証

先述した計 47 パラメータを用いて Logistic regression によるクローズドテストを行

なった。その結果、推定精度は 100 %であった。このことから、教師データを推定するためのパラメータは十分であると考えられる。そのため、Logistic regression 及び 7.2 で述べたパラメータセットを本報告のシステムに用いることとする。

7.4 オープンテストによる検証

オープンテストには先行研究[4]の手法と同じ 10-fold CV(Cross Validation)を用いる。教師データとパラメータ値は 7.2 で述べたものを用いた結果、推定精度は 50 %であった。これは過学習により、推定精度が下がったものと考えられる。そこで、計 32 のパラメータから単一毎にパラメータによる推定を行ない、カッパ係数が低いパラメータを除去し、最適なパラメータセットでの推定を試みた。その結果、バンド幅の尖度、スペクトルフラックスの回帰直線の傾きの 2 つのパラメータセットを用いた場合推定精度が 85 %となり、検証したパラメータセットの中では最も高い推定精度が確認された。そのため、このパラメータセットを用い、システムの構築を行なう。

7.5 構築したシステムの概要

入力された楽曲の歌唱者の性別を判定するシステムを構築する。使用するパラメータセットは 7.3 で述べた精査したパラメータセットを用いた。構築したシステムを図 4 に示す。図 4 左上の「INPUT FILE(*.wav)」というボタンを押し、システムに入力する楽曲、すなわち歌唱者の性別を判定したい楽曲を選択する。なお、入力される音源の形式は WAVE 形式 (44.1 kHz/16 bit) に対応している。次に、図 4 右上の「Man/Woman Estimation」というボタンを押すと推定が始まる。推定結果として図 4 の中央部分に各性別の事後確率がグラフ化され、その結果を「Result」上に表示する仕様となっている。図 4 では入力された楽曲は女性の歌唱者であると推定している。図 4 左下の「Pray」というボタンを押せば入力した音源が再生され、図 4 右下の「Stop」のボタンを押せば、再生されている楽曲を停止できる仕様となっている。なお、図 4 中央上の「State」にはシステムの現在の状態が表示される仕様となっており、

- ・ Please select "INPUT FILE"
楽曲ファイルの入力待ち画面
- ・ You can play or estimate
再生又は性別判定の実行待ち画面
- ・ Playing music
楽曲を再生中の画面
- ・ Estimating…
歌唱者の性別判定中の画面

の 4 つの状態を表すことができる。よって、音楽音響信号から自動的に歌唱者の性別

を判定することができるシステムの構築に成功した。

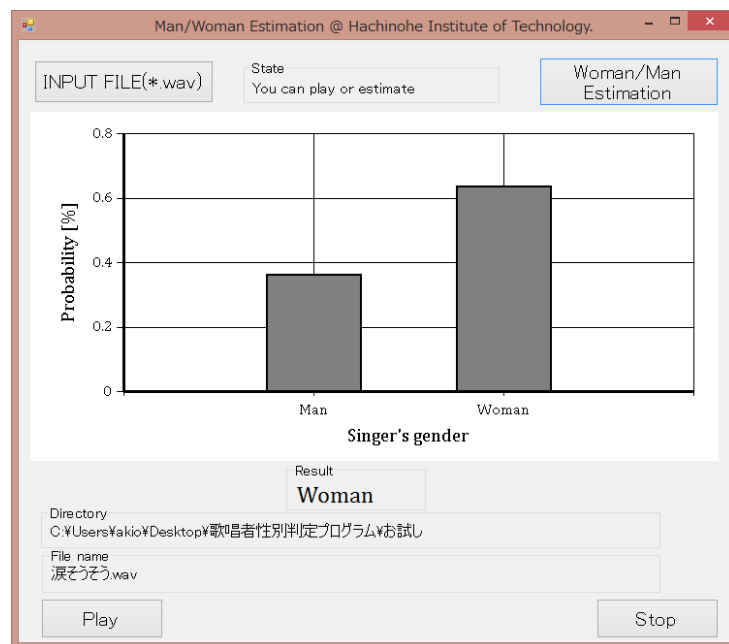


図4 構築したシステムの概要図

8. 考察

6.3 では、評価実験において女性の歌唱音を含む楽曲の方が高い評価値であることが確認されている。すなわち、本報告で提案した手法は女性の歌唱音に対して有効である可能性が考えられる。しかし、楽曲 ID=9 の様に男性の歌唱音であっても評価値が高い楽曲も存在することが確認できる。そのため、ピーク値を算出する周波数帯域についてさらに検討する必要があると考えられる。

7.3 より、正答率は 85% となり、そのチャンスレベルの 50% よりも高い結果が得られ、信号検出理論での 2 群の差が 76% 以上を超える場合とされていることを考慮してもそれを上回る結果と言える。しかし、まだ 15% のエラーがあることから検討の余地がある。特に、用いた教師データが 20 曲と少ないため、教師データの拡充をすることで推定精度の向上が期待できる。また、最適なパラメータセットは 2 つと少なく、各パラメータでの交互作用が見受けられないと考えられる。よって、パラメータの拡充を行なう必要があると考えられる。

9. まとめと今後の課題

本報告では、歌唱音と伴奏音が混在した状態の音源から歌唱音のみを短時間で抽出する手法を提案した。評価実験では、女性の歌唱音を含む楽曲の方が評価が高いことが確認され、本報告で提案した手法では女性の歌唱音に対して高い精度で抽出できる可能性が示唆された。検証実験では、短時間で提案手法が実行できることが確認された。また、1024 Hz から 1268 Hz において男性と女性の歌唱音において差が見られることが確認された。また、機械学習を用いて歌唱者の性別を推定した結果、推定精度が 85% であった。最後にそのパラメータセットを用いたシステムを構築し、正常に動作することを確認した。

今後は、男性の歌唱音に対しても達成度評価を向上させるため、ピーク値を算出する周波数帯域を変更して歌唱音抽出を行ない、男性の歌唱音に対する推定精度の向上を試みる。男性の歌唱音に対する推定精度が向上すれば、機械学習に特徴量として用いる際にさらに推定精度を高める特徴量として用いられる可能性が期待できる。また、先行研究[4]で提案された音響パラメータは音楽音響信号に対するパラメータであり、歌唱音を抽出した音源には適さない可能性も考えられる。よって推定精度を高めるには新たなパラメータを用いる必要がある。そのため、歌唱者の性別を識別する際に有効なパラメータの精査を行なっていきたい。また、パラメータの精査と同様に教師データの拡充も行ない、さらに推定精度の向上も行なっていきたい。

謝辞

本研究の一部は、科研費(15H02882)の援助を受けた。また、本研究を進めるうえで貴重なご意見をいただきました龍谷大学情報メディア学科桶本まどか女史に感謝いたします。

参考文献

- [1] 佐々木ら, “混合音におけるボーカル音声強調のための基礎的検討”, 電気関係学会東北支部連合大会講演論文集 pp.-201 (2010)
- [2] 藤原ら, “歌声の統計的モデル化とビタビ探索を用いた多重奏中のボーカルパートに対する音高推定手法”, 情報処理学会誌 Vol.49 No.10 pp3682-3693 (2008)
- [3] 櫻庭ら, “女性と判定される声の特徴—性同一性障害者の話声位—” 音声言語医学 Vol.50 No.1 pp.14-20 (2009)
- [4] 岡田ら, “ポピュラー音楽の音響信号を対象とした年代推定システムの提案”, 日本音響学会音楽音響研究会・日本音楽知覚認知学会 合同研究発表会, 日本音響学会研究資料, MA2016, 39 (2016)