修士論文

ボイストレーニングのための SVMを用いた歌唱音声評価 —YUBAメソッドの初期段階に 限定したシステムの構築—

平成28年度

三重大学大学院　工学研究科
博士前期課程　物理工学専攻

酒井　彰史

目 次

第1章　序論

1.1　研究の背景と必要性

最近は若者だけでなく、中高年でも趣味でカラオケを楽しんだり合唱サークルに所属して歌を歌う人が多い。また、その人達が歌いたいと思う曲には高音域の発声（一般に裏声あるいはファルセットボイスと言われる）を必要とするものも多く、彼らは「どうすればプロ歌手のように高音をきれいに発声できるのか」ということに強い関心を持っている。その中で、数年程前からYUBAメソッドという発声トレーニング法がテレビなどのメディアでよく紹介されて注目を集めている[1–4]。YUBAメソッドとは三重大学教育学部教授弓場徹が提唱する歌唱トレーニング法（第2章参照）であり、本研究はこれに関連する弓場との共同研究の一部として実施されたものである。

ここで、まずYUBAメソッドのトレーニング法について簡単に説明する。YUBAメソッドでは最初に音域の拡張を目的に裏声と表声（地声ともいう）を分離して発声する訓練を行う。その後、表声が声帯の振動様態の異なる裏声に切り換わる音域つまり換声域[5]での音色の急激な変化や音程の乱れ（換声点ショックという）を目立たせないように裏声と表声を滑らかに変化させる訓練に移行する。このようなYUBAメソッドの普及を目的に、その具体的な方法を解説した書籍[6–11]、CD[12,13]、DVD[14–18]が多数出版・販売されている。またYUBAメソッドを利用した歌唱トレーニングでの音痴克服や安定した歌唱習得の成果も発表[19,20]されており、YUBAメソッド自体の有効性は既に確認されている。

YUBAメソッドのトレーニングでは表声と裏声をしっかりと出し分けられているか、裏声と表声が滑らかに変化しているか（換声点ショックが小さいか）について本来は熟練した指導者が耳で聞いて判断することが望ましく、発声者自身もそれらの点を意識することが重要とされている。しかし、指導者がいない状況で書籍などを購入した初心者が自身の発声を評価することは難しく、トレーニングの導入の妨げになっていた。そこで、当研究室では個人による歌唱トレーニングを効率よく実施できるように、弓場との共同研究により機械学習を利用した表声/裏声の自動判別のためのシステムの構築を試みてきた[21–23]。一方、歌唱トレーニングにおいては表声/裏声の判別以

外にも「息の漏れ度合」を評価することも重要視されている。例えば同じ裏声でも、息漏れの少ないいわゆる「歌える裏声」と、息漏れの多い「息漏れの裏声」の区別があり、前者は歌唱に適した発声である。一方、後者は歌唱には適さないものの、音程をとるために働く輪状甲状筋を効率よく鍛えるための発声であり、YUBA メソッドの初期段階では特にこの発声が求められる。したがって、表声/裏声の判別に加え息漏れの度合を評価することにより、正確で信頼できる声質の評価が可能となり、より効率的なトレーニングが可能になると考えられるが、これまでの研究では息漏れ度合の評価に関しては、あまり高い精度が得られていなかった [23]。

そこで本研究では研究方針を見直し、機械学習による判別を YUBA メソッドの初期段階で発声される音声に限定することで、息漏れ度合を含めた声質の判別精度を高くできる可能性がないかを確かめることにした。つまり機械による音声の判別の用途を限定することで、判別精度が向上するのか、また実際のトレーニングにおいて専門の指導者と同程度の判別精度が期待できるのかを探ることにした。そのため本研究では YUBA メソッドに精通している専門家 1 名による声質の再評価結果との比較も試みる。ちなみに YUBA メソッドでは練習の初期段階においても、表声と裏声の出し分けだけでなく、息漏れ度合も意識して発声する必要がある。具体的には「息漏れのない表声」と「息漏れのある裏声」の発声コントロールが求められる。

1.2　YUBA メソッドで必要とされる音声とその評価法

YUBA メソッドでは人の発声は男声、女声ともに 2 声区 (表声・裏声) しか存在しないものとしている。しかし専門家が歌唱音声を評価する場合、「裏声の混ざった表声」や「表声の混ざった裏声」などといった表現をする場合がある。トレーニングにおいても「しっかりした表声」「息漏れた裏声」など声楽家は様々な表現を使用する。これは発声者が感覚的に理解しやすいように指示する言葉を選んでいるためと考えられるが、歌声の評価を自動化するためにはこれらの表現を整理して工学的に取り扱い易いように指標化することが必要である。

そこで従来までの単なる表声/裏声のみの 2 段階評価ではなく、表声にどのくらい裏声が混ざっているのかを表す独自の指標 Falsetto Mixing Ratio(以後 FMR と記述) を導入した。また前述した「息漏れ度合」の評価のための指標 Breathy Strength(以後 BS と記述) も導入し、歌唱音声の評価に用いることにした。これらの指標は 0 から 1 までの数値で表され、FMR=0 が「完全な表声」、FMR=1 が「完全な裏声」であることを意味し、BS=0 が「息漏れのほとんどない歌声」、

BS=1 が「息漏れが最も多い歌声」であることを意味している。これらの指標は専門家の耳による感覚を頼りに単音ごとに値を割り当て、SVM による機械学習のための教師データとして用いている (第 3〜4 章参照)。なお本研究で取り扱う YUBA メソッドの初期段階では、1 音毎に発声した音声の判別に重点が置かれ、発声する母音も/a/と/o/に限定される。

1.3　研究概要

従来の研究において SVM を用いて前述の FMR と BS の予測が試みられている [23] が、FMR の平均正解率は 66%、BS の平均正解率は 55%にとどまっていた。そこで本研究では以下に挙げる内容を実施する。

- データベースの再構築

  従来の研究で用いたデータベースには総計 10,265 音におよぶ 20 代から 50 代男性の音声サンプルが収録されていたが、その中にはきちんと発声しておらず音程が不安定な音声も含まれていた。本研究では、このような音声の他に、物音や他の人の声が一緒に録音されているもの、編集ミスにより重複して収録されているものを取り除いてデータベースを再編する。また、全ての収録音声を 1 音毎に切り出してファイル化し、改めて音声パラメータの解析も行う。そして、各音毎に算出された音高 (ピッチ周波数)、音量、高周波比率などの物理的評価値を専門家 1 名による FMR と BS の評価結果と共にデータベースに収録する。本研究では、このデータベースを利用し、YUBA メソッドの初期段階で発声される音声について、FMR と BS を予測する SVM をそれぞれ構築し評価精度の検証を行う。

- 評価方法の変更

  従来の研究では、年代を考慮してデータベースの音声を無作為に学習用と評価用に二分して判別システムの予測精度を評価していた。これはいわゆるホールドアウト検証であり、抽出された音声に学習が依存してしまう危険性がある。そのため、本研究ではデータを 10 グループに分けて 10 分割交差検証を行う。

- 学習を限定した SVM の予測

  データベース中から YUBA メソッドの初期段階では発声しない連続歌唱音声を取り除き、評価対象母音を/a/と/o/に限定したデータを用いて SVM を学習・予測する。この際、限定

した音声データの数は 2,626 件である。そして、この予測結果を学習データを YUBA メソッドの初期段階に限定しない場合の SVM の予測結果と比較する。

- 専門家の判断との比較

  本研究では、SVM による評価の比較、検討だけでなく専門家 1 名に再度同じ音声について $FMR \cdot BS$ の評価を依頼し、ヒトの判断の再現性や SVM による予測結果との一致性について調査する。

1.4　他の研究との比較

ここで、歌唱音声の評価に関しては、YUBA メソッドを用いた類似研究が実施されているので、それらとの相違点について触れておく。類似研究は浅野らによる “裏声判別指標を用いたボイストレーニングソフトウェア”[22] である。この研究では、YUBA メソッドのボイストレーニング初期段階における裏声と表声を、被験者に発声してもらい、その歌声について裏声と表声をしっかりと出し分けられているのかを、LPC 分析によって判別していた。これに対して、本研究では歌声を音声解析ソフトの VoiceSauce[24] により解析し、算出されるパラメータを用いて SVM で学習・評価する。また、表声と裏声の判別について FMR という独自の指標を導入する。さらに、表声と裏声の判別だけでなく、発声した音声の息漏れ度合についても評価するため、息漏れ度合の判別指標として BS という指標を導入する。これらの点が浅野らの研究とは異なっており、表声と裏声の判別だけでなく息漏れ度合の判別もできるシステムである。浅野らの研究の他にも、岩本らによる “機械学習に基づく歌唱音声の声質評価システムの構築”[23] が類似研究として挙げられる。この研究では、SVM を用いて被験者の発声した歌声に対する表声/裏声の判別と息漏れ度合の判別を行うというものである。表声/裏声の判別は平均正解率が 66%であったが、息漏れ度合の判別については、平均正解率が 55%と高い精度は得ることができないという結果であった。本研究では、岩本らが使っていた音声データベースを再構築し、SVM の学習のために用いた入力パラメータも変更している。また、岩本らの研究では、音声データベースの全ての音声を使って学習・評価を行っていたが、息漏れ度合の予測精度が良くなかったため本研究では連続歌唱音声を除き、YUBA メソッドの初期段階に絞って評価精度の向上を目指している。

1.5　本論文の構成

以下に本論文の構成を示す。

第 1 章では、研究背景・目的

第 2 章では、発声メカニズムと YUBA メソッド

第 3 章では、歌唱音声データベースの再構築

第 4 章では、予測評価手法

第 5 章では、FMR の評価精度に関する検討

第 6 章では、BS の評価精度に関する検討

第 7 章では、専門家の再評価精度に関する検討

第 8 章では、総括と今後の課題

について述べる。

第2章　発声メカニズムとYUBAメソッド

本章では、ヒトの音声の特徴とともに研究の遂行に必要となる裏声・表声の発声メカニズムと歌唱トレーニング法『YUBA メソッド』について概説する。

2.1　ヒトの発声メカニズム [25]

歌唱音声に限らず、ヒトが発する様々な声の多く（有声音として母音が代表的）は、肺から送られた呼気流によって声帯（声門）が振動する（閉じたり開いたりする状態を繰り返す）ことで生じた音（声帯音源という）によって作り出されている。声帯音源は気流の断続で生ずる波形（三角波に近い形状）で、我々が普段耳にする声とは異質のブザー音のようなものである。しかし、これが口腔・咽頭・喉頭・鼻腔・副鼻腔で構成される断面形状が長手方向に沿って複雑に変化する管（音声学的には声道という）を通ることで特定の周波数成分が強調されたり抑圧されたりして（周波数スペクトルに変化が生じ）、口や鼻孔から聞き慣れた声として大気中に放射されている。つまり声道は声帯原音のスペクトルを変化させて声に変換するフィルタ装置と見なすことができ、これを声道フィルタと呼ぶ。要約すれば、ヒトの声は声帯で発声した声帯音源を声道フィルタに通すことで得られる音といえる。図 2.1 は声帯音源から音声が作られるイメージを図示したものである。

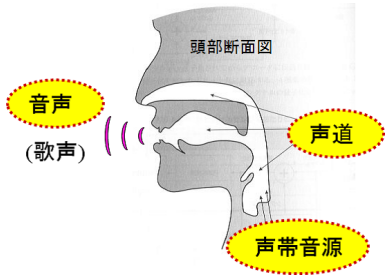


図 2.1: 人の発声過程の図

2.2　音韻と音程の違い

ヒトの声を特徴付けるものとして、大きさ、音韻、音高（ピッチ）がある。音声の大きさの変化が声帯音源の大きさに依存していることは自明である。

これに対して、「あ」「い」「う」のような音韻の認識の違いは音声のスペクトルのエンベロープのピーク、すなわち声道フィルタの局所ピーク（フォルマントと呼ばれる）の相対的なレベルとその位置関係（フォルマント周波数の組み合わせ）によるものと考えられている。

また、声の高さ（ピッチ、音高）は音声のフォルマントとは関係なく声帯原音の周期に依存しており、その逆数である基本周波数で決定される。つまり、音の高さはフォルマント情報には関係なく声帯の振動周期のみに依存していることになる。

図 2.2 に音声波形とスペクトルのイメージを、図 2.3 にフォルマントのイメージを示す。本論文では図 2.2 に示すようにピッチ周波数（単位 Hz）を f_0、基本波のスペクトル強度（単位 dB）を H_1 で表し、その高調波である 2〜n 倍音のスペクトル強度を H_2〜H_n で表すことにする。同図よりスペクトルの細かな周期構造がピッチを決める要因になっていることがわかる。また、図 2.3 に示すようにスペクトル包絡に現れるピークがフォルマントであり、低い周波数の方から順に第 1、第 2 … フォルマントと呼ばれる。本論文中でのそれらのピーク周波数（単位 Hz）をフォルマント周波数として記号 F_1,F_2, … で表す。またそれぞれのピーク値（スペクトル強度、単位 dB）を A_1,A_2, … で表す。このようなゆるやかなスペクトル包絡の形状が音韻を決める要素になっている。

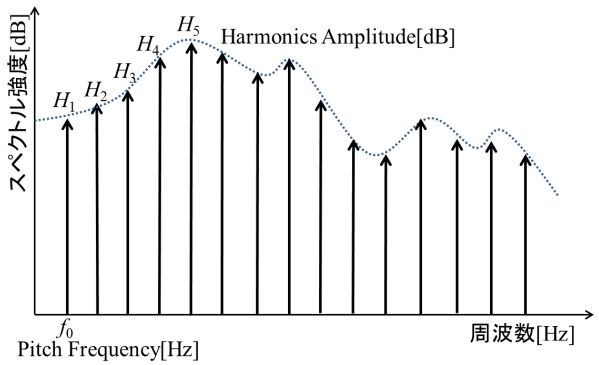


図 2.2: 発声に対する倍音調波構造のイメージ図

ところで、会話音声のピッチ（声の高さ）は声帯が最も効率よく振動する周波数で決定されており、個人（特に男女）間のピッチ差は声帯の長さ・質量・張力などと関連がある。通常の会話

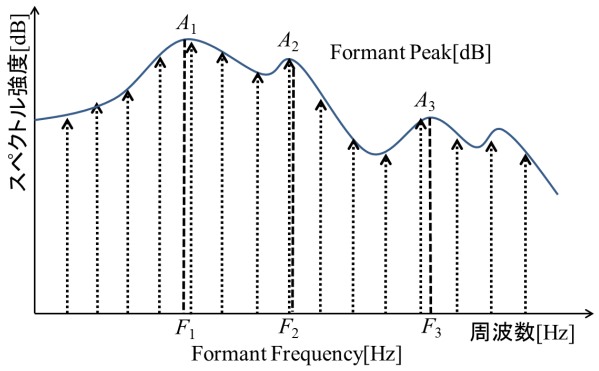


図 2.3: 発声に対するフォルマントのイメージ図

音声の場合、ピッチ周波数は男声で 60∼260Hz、女声で 120∼520Hz に分布するが、通常の会話で各個人が変化させる範囲はせいぜい 100∼200Hz 程度である。しかし、歌を歌う場合にはこのピッチをメロディに合わせて、より広い範囲で変化させることが必要となる。当然、通常の会話音声の発声とは異なる声帯の振動が必要とされる。後述するように特に高音を発声する場合には声帯のコントロールが難しくなり、発声ができなかったり、音程を外す原因となる。

2.3　裏声と表声の違い

弓場の著書 “奇跡のボイストレーニング BOOK（主婦の友社,2004）” によれば、裏声と表声の発声法の違いには内喉頭筋群が関係している。内喉頭筋群とは声帯を引っ張ったり、声門（左右の声帯のすき間）を閉じたり開いたりして、声帯の動きをコントロールしている喉にある一連の筋肉群のことであり、喉ぼとけや甲状軟骨に付随する閉鎖筋群や開大筋がある（図 2.4 参照）。

弓場はこれらの筋肉の中でも声帯を引っ張り伸ばす筋肉や声門を閉じる筋肉ことを、歌うことの中心的な役割を担っているので「歌う筋肉」と呼んでいる。

これら筋肉のうち、音の高さを変えるのに主役となって働くのが輪状甲状筋である。この筋肉は気管の一番上にある輪状軟骨と甲状軟骨（突出したところを一般に喉ぼとけと呼ぶ）をつないでいる。この筋肉が働くと、甲状軟骨と輪状軟骨が近づいて声帯が引き伸ばされこの時声帯の傾きが弱く声帯の質量が小さいと音が高くなり裏声が出る。一方、閉鎖筋群が輪状甲状筋に対して優勢に働き、声帯筋の働きにより声帯の質量が大きい状態で声門が閉じられると息漏れの少ない表声になる。

したがって表声か裏声かは、内喉頭筋の筋肉運動による声帯の振動状態の違いで決まるのであっ

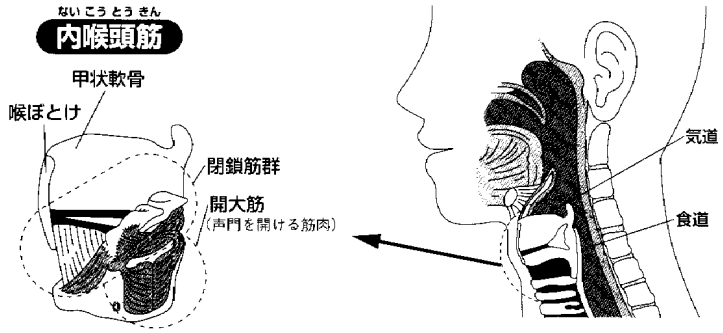


図 2.4: 内喉頭筋の様子 [5]

て、声の響きの状態で決まるわけではない。図 2.5 に裏声発声時の輪状甲状筋の働きを示す。

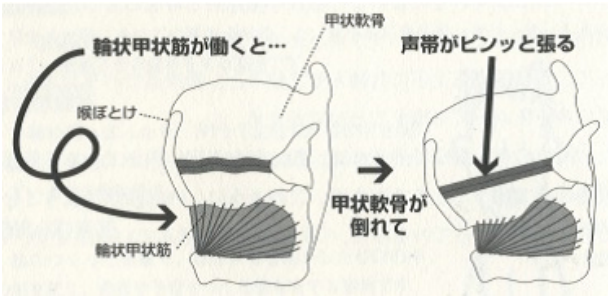


図 2.5: 裏声発声時の輪状甲状筋の働き [5]

2.4　YUBA メソッド

YUBA メソッドとは弓場が提唱しているボイストレーニング法のことである。このトレーニング法は、ヒトは内喉頭筋を直接意識してコントロールすることはできないが、出す声によってどこの筋肉が働くかはおおよそ予想できるため、モデルとなる声をまねて発声することにより間接的に歌うために必要な筋肉を効率よくコントロールできるようになるという考え方（YUBA 理論、発声制御理論）に基づいている。トレーニングの簡単な流れは図 2.6 に示す通りである。

図 2.6 中のそれぞれの Stage の目的と練習内容は次のように定義されている。

- **Stage 1:** 裏声と表声をはっきりと分けて出す

  例 1：息漏れのある高い裏声を出す

  フクロウの鳴き声「ホー」や犬の遠吠え「ウォー」等をまねて発声し、裏声を出すことに慣れる。

例 2：息漏れのない表声を出す

口を「あ」の形に開け、息を止めてからひと息で「アー」とはっきりした息漏れのない（息が効率よく声帯振動に変わる状態）低めの声で 2~3 秒声を出す。

- **Stage 2:** 裏声・表声でいろいろな高さの音を出す

例：Stage1 で発声した音を様々な音程で歌唱する

「ホー」と高めの裏声で始め、「ホー・ホー・ホー・ホー」と一声ずつ音の高さを変えて出す。次に「ホー」を表声の「アー」に変えて行う。

- **Stage 3:** 裏声・表声で簡単なメロディを歌う

例：「かえるの合唱」などの簡単なメロディーを高い音域の裏声「オー」（または「ウー」）で歌う。息漏れを少なくし、一息で長めのフレーズを歌う。次に音域を下げて低めの息漏れのない表声「アー」で同じメロディーを歌う。

- **Stage 4:** 裏声と表声の両方の声を行き来して歌う

例：「ドーシーラーソーファーミーレードー」と高い音から「裏声→表声」に向かって歌い、反対に低い音から「表声→裏声」でも練習する。途中換声点で声がひっくり返ったり、出しにくくなっても音程が外れなければ良好な状態と判断する。

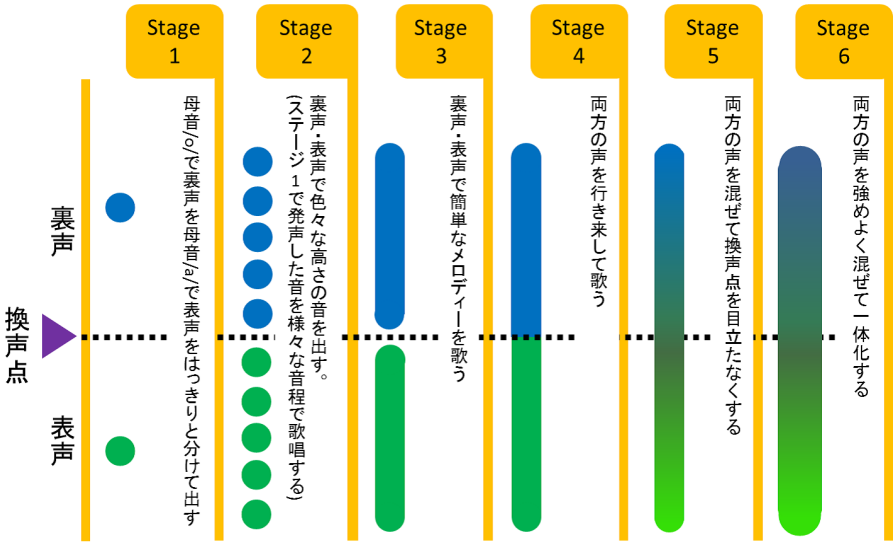


図 2.6: YUBA メソッドのトレーニング段階図（弓場によるイメージ図）

- **Stage 5:** 両方の声を混ぜて換声点を目立たなくする

  例：出来るだけ高めの息漏れのない裏声を「オー」（息漏れするようなら「アー」）で歌い始め、表声に向かって 2 オクターブ（ドーシラソファミレドーシラソファミレドー）下げていく。

- **Stage 6:** 両方の声を強めよく混ぜて一体化する

  例：さらに喉の筋肉トレーニングが進んで Stage 5 がより発展した状態である。

このボイストレーニングを実施することで、表声と裏声の境目である換声点での急激な音質や音量の変化を減らし、広い音域をなめらかに発声することが可能になる。インストラクタの模範発声をまねて実践的にボイストレーニングできるトレーニング本（CD 付）や CD,DVD[11–17] が出版されている。

2.5　普及のための課題

本章で紹介した YUBA メソッドの発声・歌唱教育上の効果の高さは既に検証されている [21] が、第 1 章で述べたように、個人で本（CD 付）や DVD を購入してトレーニングする場合を考えると、発声状態の確認は自己判断に委ねられるため練習が効率的に進まないことが多々ある。そのため、より効率的にトレーニングが進むように、個人レベルで客観的に自分の発声が裏声なのか表声なのか息がどのくらい漏れているのかという情報を発声者にフィードバックすることが重要になってくる。そこで、その判別のために FMR や BS などの声質判別指標を導入することが求められている。また、このような指標を利用した個人で簡単かつ効率的にトレーニングできるアプリケーションの開発も期待されている。

第3章　歌唱音声データベースの再構築

本章では本研究の遂行のために新たに再構築するデータベースの作成手順と収録されるデータについて解説する。従来の研究に用いられていたデータベースには総計 10,265 件におよぶ男声のサンプルが収録されていた。しかし、改めてその内容を精査したところ発声が安定しておらず音程が不安定な音声サンプルが含まれていることがわかった。そこで、このような音声サンプルを取り除くために音声データベースを再構築する。

3.1　収録音声の切り出し

データベースを再構築するために、まず従来の研究で用いられていたデータベースから音程が不安定な音声を 1 音ずつ再生して耳で確認しながら取り除いた。削除された音声として、以下のものが挙げられる。

- 歌唱音声のピッチが不安定なもの
- 発声している最中に物音や他の人の声が一緒に録音されているもの
- 編集ミスにより同じ音声が重複して保存されていたもの

この作業で音声サンプル数は、10,265 件から 9,329 件まで減少したが、より分かりやすい音声サンプルを集めたデータベースとなった。データベースに収録されている音声について専門家が評価した FMR と BS の値の内訳を図 3.1 に示す。図 3.1 をみると表声傾向で息漏れの多い（FMR が小さく、BS が大きい）歌唱音声の割合が少ないことがわかる。これは第 2 章で説明したように、表声発声時は声門が閉じている状態にあり、表声では息漏れがあまり発生しないためであると考えられる。

なお、音声サンプルは発声形態により以下の 3 つのグループに分類できる。

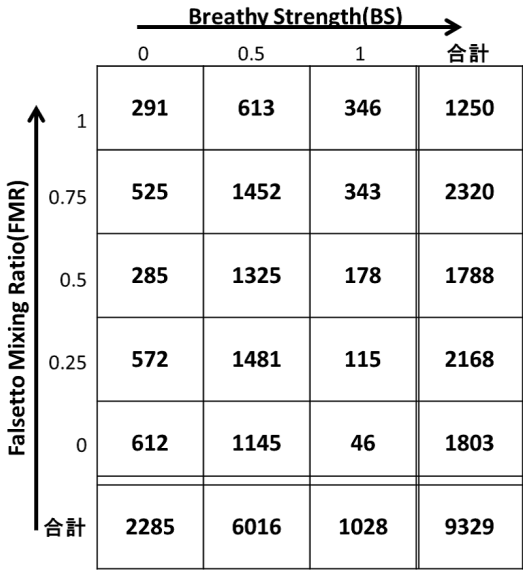


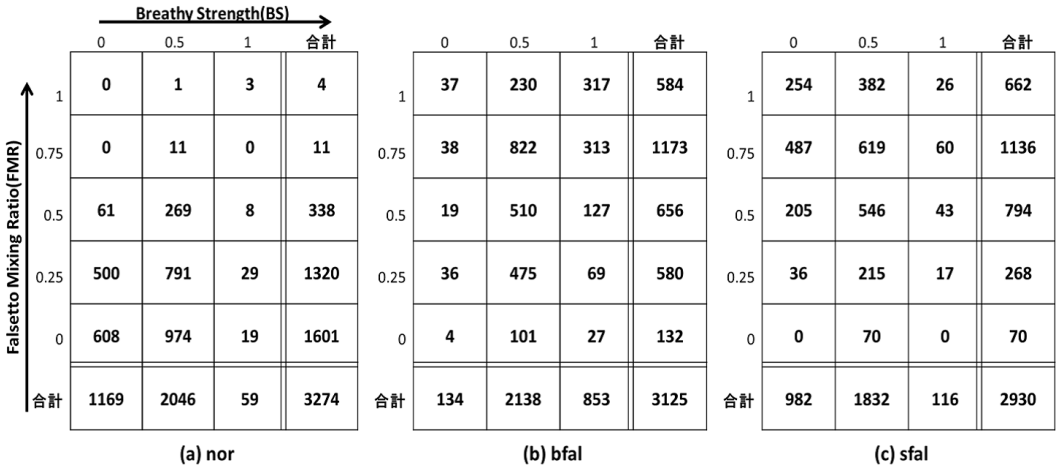
図 3.1: FMR・BS 値別で再構築したデータベース内訳（単位：個）

- **グループ 1:** 表声を 1 音ずつ区切って発声したもの。normal voice の略で以後、nor と表記する。この発声は YUBA メソッドの Stage 1〜5 で求められる。また YUBA メソッドの初期段階では BS=0 で FMR ＜ 0.5 の発声が要求される。

- **グループ 2:** 裏声を 1 音ずつ区切って発声したもの。breathy falsetto の略で以後、bfal と表記する。この発声は YUBA メソッドの Stage 1〜2 で求められる。また YUBA メソッドの初期段階では FMR＞ 0.5 の発声が要求される。

- **グループ 3:** 音をつなげて一息で発声したもの（連続歌唱音声とも呼ぶ）。singable falsetto の略で以後、sfal と表記する。この発声は YUBA メソッドの Stage 3〜6 で求められる。今回のデータベースの再構築において、このグループの音声はフレーズ単位で録音されたものを、音程の安定した区間を特定して音声が不自然にならないようにテーパー状になった時間窓を掛けて 1 音ずつ切り分けた。

データベースに含まれる音声の上記 3 グループ別の FMR、BS の構成割合を図 3.2 に示す。図より各グループの音声は約 3,000 件であり、発声形態の異なる音声をバランスよく収録していることがわかる。この他に、nor の音声には FMR と BS が共に小さいもの、bfal の音声には FMR と BS が共に大きいもの、sfal の音声には FMR が大きく BS が小さいものが多いこともわかる。



FMR \ BS	0	0.5	1	合計
1	0	1	3	4
0.75	0	11	0	11
0.5	61	269	8	338
0.25	500	791	29	1320
0	608	974	19	1601
合計	1169	2046	59	3274

(a) nor

FMR \ BS	0	0.5	1	合計
1	37	230	317	584
0.75	38	822	313	1173
0.5	19	510	127	656
0.25	36	475	69	580
0	4	101	27	132
合計	134	2138	853	3125

(b) bfal

FMR \ BS	0	0.5	1	合計
1	254	382	26	662
0.75	487	619	60	1136
0.5	205	546	43	794
0.25	36	215	17	268
0	0	70	0	70
合計	982	1832	116	2930

(c) sfal

図 3.2: FMR・BS 値別での各データベース内訳（単位：個）

3 グループの音声は今回、1 音 1 音個別に切り出されてファイル化された。これは今後、様々な音声分析ソフトを使い、そこから得た分析結果でデータベースを拡充する際に役立つものと考えられる。このように切り分けた音声は平均して 0.5s の継続時間を有する。

3.2　データベース収録の音声パラメータ

新たに 1 音毎に分離されたデータベース中の音声を改めて音声解析ソフト VoiceSauce により分析した。VoiceSauce によって 1ms 毎に算出（フレーム長 25ms、シフト量 1ms による分析）される音声のパラメータ [26] を表 3.1 に示す。データベースに収録されている音声の平均的な継続時間は約 0.5s であるので、1 つの音声について、これらのパラメータが 500 個程度算出されることになる。そこで 1 音ずつの代表値としてその中央値や平均値、第三四分位数、パワー平均、分散を求めた。このようにして抽出され、データベースに収録されたパラメータを表 3.2 に示す。なお、H_1-H_2 と A_1-A_3 は VoiceSauce で解析されるパラメータを用いて、独自に算出したものである。

また各音声について YUBA メソッドに精通した専門家により FMR と BS の値ならびに音名（フィールド名 MIDI）が評価された。今回切り分けた音声 1 音毎に評価された FMR と BS については第 1 章で述べたように 0 から 1 までの数値で表し、専門家の意見に基づき FMR は {0, 0.25, 0.5, 0.75, 1} を 5 段階、BS は {0, 0.5, 1} を 3 段階で評価されている。FMR と BS の評価イメージを図 3.3 に示す。この図は専門家のおおよその感覚を図的に表したものであり、FMR と BS の評価軸は直交するものと仮定している。切り分けた音声データが図のどの位置の音声であるのか

表 3.1: VoiceSauce で算出されるパラメーター一覧

記号 (フィールド名)	内容
f0	基本周波数 f_0[Hz] 計算には STRAIGHT(kawahara et al. 1998) を使用 他にも Snack Sound Toolkit(Sjolamder 2004) を用いて計算したものもある
H1,H2,H4	基本周波数、第 2、第 4 倍音の振幅 [dB] H_1 と H_2 の値はフォルマントの影響を受けて歪むこともある
A1,A2,A3	第 1、第 2、第 3 フォルマントのスペクトル強度 [dB] A_1、A_2、A_3
H1H2c,H2H4c	H_1-H_2、H_2-H_4 をフォルマントの値に基づいて修正した値
H1A1c,H1A2c,H1A3c	H_1-A_1、H_1-A_2、H_1-A_3 をフォルマントの値に基づいて修正した値
sF1,sF2,sF3,sF4	第 1、第 2、第 3 フォルマント周波数 [Hz] F_1、F_2、F_3
Energy	エネルギー　the Root Mean Square(RMS) で求める (RMS は音圧の実効値である)
CPP	Cepstral Peak Prominence Hillenbrand et al. (1994) のアルゴリズムに基づいて計算される
HNR	倍音-ノイズ比率 (Harmonic Noise to Ratios) de Krom(1993) のアルゴリズムで求める

を専門家の耳の感覚でプロットしてもらい各音毎にラベリングを行った。図 3.3 で色が変わっている部分が YUBA メソッドの Stage 1 で発声が求められる理想の音声の範囲である。

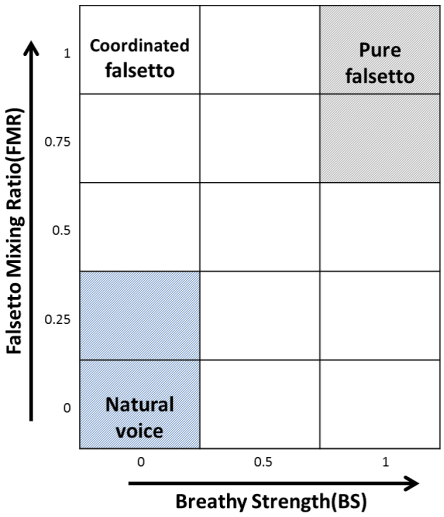


図 3.3: $FMR \cdot BS$ 評価のイメージ図

最終的に、専門家による評価と VoiceSauce による各パラメータの算出結果を結合しデータベースを構築した。このデータベースはさまざまな年代の音声データを網羅しており汎用性が高く貴重であるといえる。

表 3.2: 内包されるパラメータ一覧

記号 (フィールド名)	内容
SUB	歌唱者番号
VOWEL	母音コード (1:/a/,2:/i/,3:/u/,4:/e/,5:/o/)
FMR	Falsetto Mixing Ratio の専門家による評価値 (0〜1)
BS	Breathy Strength の専門家による評価値 (0〜1)
MIDI	音名 (MIDI ノート番号)
pf0,sf0,strf0(.avr, .med)	基本周波数 f_0[Hz] それぞれ Praat,Snack,STRAIGHT での計算結果の 平均値 (.avr) と中央値 (.med)
H1H2c,H2H4c(.med, .pow, .q3)	H_1-H_2、H_2-H_4 の修正値の 中央値 (.med) とパワー平均値 (.pow) と 第三四分位数 (.q3)
H1A1c,H1A2c,H1A3c(.med, .pow, .q3)	H_1-A_1、H_1-A_2、H_1-A_3 の修正値の 中央値とパワー平均値と第三四分位数
H2K(.med, .pow, .q3)	2kHz 付近での倍音レベル H_{2k} の 中央値とパワー平均値と第三四分位数
H42Kc(.med, .pow, .q3)	H_4-H_{2k} の修正値の 中央値とパワー平均値と第三四分位数
H5K(.med, .pow, .q3)	5kHz 付近での倍音レベル H_{5k} の 中央値とパワー平均値と第三四分位数
H2KH5Kc(.med, .pow, .q3)	H_{2k}-H_{5k} の修正値の 中央値とパワー平均値と第三四分位数
A1,A2,A3(.med, .pow, .q3)	第 1、第 2、第 3 フォルマントのスペクトル強度 [dB] A_1、A_2、A_3 の 中央値とパワー平均値と第三四分位数
H1,H2,H4(.med, .pow, .q3)	基本周波数、第 2、第 4 倍音の振幅 [dB] H_1、H_2、H_4 の 中央値とパワー平均値と第三四分位数
HNR05,HNR15, HNR25,HNR35 (.avr, .med, .pow, .q3, .std)	Harmonic to Noise Ratio (05 は 0〜500Hz、15 は 0〜1500Hz, 25 は 0〜2500Hz、35 は 0〜3500Hz までの測定値) の 平均値と中央値とパワー平均値と第三四分位数と分散
H1H2,A1A3(.med)	H_1-H_2 と A_1-A_3 の中央値

3.3　まとめと課題

本章では実際に再構築したデータベースの作成手順と収録されているデータの内訳について説明した。歌唱音声は発声形態別に 3 つのグループ (nor、bfal、sfal) が保存されており、用途に応じて区別できるようになっている。また、データベース収録の音声パラメータは、同じものであっても音声ごとに中央値や平均値など様々な値を算出しており、汎用性が高く貴重である。

なお現在は FMR と BS の評価を YUBA メソッドに精通した 1 名の専門家により行っているが、複数の評価者による平均化も今後検討しなければならない。

第4章　予測評価手法

本章では、前章で述べたベータベースを用いて FMR と BS を予測・評価するための学習データと評価データの作成法、また、そのために必要な SVM の構成について説明する。

4.1　SVM[27] と使用ソフト

SVM は教師あり学習に用いる識別手法の一つであり、現在知られている多くの手法の中で一番認識性能が優れた学習モデルの一つといわれている。SVM ではまずグループ分けした学習のグループパターンをコンピュータに学習させ、その学習結果をもとに評価データをグループ分けすることでパターン認識を行う。また SVM の最大の特徴としてマージン最大化がある。これは最も適したグループ分けを行うために学習データの中で最も他クラスと近い位置にいるものを基準として、そのユークリッド距離が最も大きくなるような位置に識別境界を設定する。このノンパラメトリックな手法で明確な基準で識別境界をを与えることができるのが、SVM の最も優れた部分である（図 4.1 参照）。本研究では WEKA[28] と呼ばれるアプリケーションソフトウェアを使用し SVM を構築した。

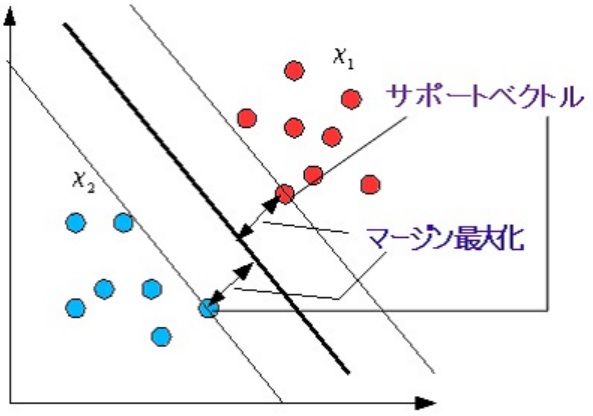


図 4.1: SVM によるマージン最大化イメージ図 [27]

4.2　SVM の構成

ここではまず学習させた SVM の構成についてまとめておく。

4.2.1　入力要素

入力要素には従来の研究 [23] と息漏れについて記述されている論文 [29] を参考にして、第 3 章で説明したデータベースに収録されている以下 6 つのパラメータを使用した。

- f_0：基本周波数 [Hz]
- H_1-H_2：基本波成分と第 2 倍音（$2f_0$）のレベル差 [dB]
- H_1-H_{2c}：H_1-H_2 の修正値 [dB]
- A_1-A_3：100〜1000Hz と 1800〜4000Hz の間のそれぞれ最も高い振幅値の差 [dB]
- HNR_{15}：0〜1500Hz までの雑音比率 [dB]
- HNR_{35}：0〜3500Hz までの雑音比率 [dB]

熟練者が耳で聞いて音声を評価する場合、基本周波数は最も重要なパラメータであることは明白である。また、H_1-H_2 は倍音成分と基本波成分とのレベル差であり裏声や息漏れの判別に有効とされているパラメータである。H_1-H_{2c} はこれをフォルマント周波数とその帯域幅を使って修正したものである。H_1-H_2 と情報として重複する部分が大きいが、音声研究の分野で広く使われる指標であるため今回入力要素として採用した。A_1-A_3 は息漏れに関する文献 [29] と、以前の研究結果（酒井、2016）[30] から歌声の判別に対して有効であると判断し、入力要素として加えることにした。HNR_{15} と HNR_{35} に関しては従来の研究（岩本、2015）[23] から HNR が有効であることがわかっており、その中でもこの 2 つのパラメータが有効であるとされていることから入力要素に採用した。

4.2.2　出力と学習アルゴリズム

SVM の出力形式には離散値タイプを選択し、専門家の判断と同じく FMR については {0, 0.25, 0.5, 0.75, 1} の 5 段階の数値で、BS については {0, 0.5, 1} の 3 段階の数値で出力を得る。SVM

の学習アルゴリズム [28] には SMO を選択した。FMR と BS はそれぞれ別々に予測するための個別の SVM を構築した。

4.2.3　学習データと評価データ

前章で構築したデータベースには 9,329 件の音声データが含まれている。従来の研究では、1 組の学習データと評価データで予測と評価を行うホールドアウト検証を実施した。しかし、本研究では 1 つの学習・評価パターンだけではなく、より多くのパターンで算出することで評価結果の正確性を高めることを考えた。そのために計 10 個の学習・評価パターンから SVM の評価結果の平均をとって判断する 10 分割交差検証を行った。

また、YUBA メソッドの初期段階に限定して予測を行うために、必要な音声データを選別した。本研究でいう YUBA メソッドの初期段階とは具体的に第 2 章で述べた YUBA メソッドの Stage 1 に相当するものである。Stage 1 の例 1 で発声する「ホー」や「ウォー」を母音/o/で、例 2 で発声する「アー」を母音/a/で予測するために、データベースから母音/a/と/o/のみの音声を抽出した。また、Stage 1 では歌唱ではなく、正しく 1 音 1 音が発声できているかを評価するため、連続で歌唱している音声である sfal を取り除いた。

さらに交差検証用に音声を FMR と BS でバランスよく分けるために、図 4.2 のようにデータ（音声やそのパラメータ）を 10 等分して 10 個のデータセットを作成した。この 10 個のデータセットから 9 個を学習用に、残りの 1 個を評価用に採用する方法で、全 10 パターンの評価結果を算出した（交差検証）。

また、学習データセットで FMR と BS 共にその評価値の分布に偏りがあり、偏りがあるまま SVM で学習させると学習がデータ数の多いものに偏り、予測も偏ったデータに左右されてしまう [23]。そのため、データ数のバランスをとる（分布を均一化する）ために各評価値の中でデータ数が最少のものに一致するように他のデータを無作為に選別し、図 4.3 のように均一化を施した上で SVM の学習に用いた。また、この作業によって排除された余分なデータがあるが、その使われないデータを無駄にしないために、10 分割交差検証で全体の平均をとり、各パターンで均一化を行うことで使われないデータを極力少なくし、また全データを評価データにも使えるようにした。

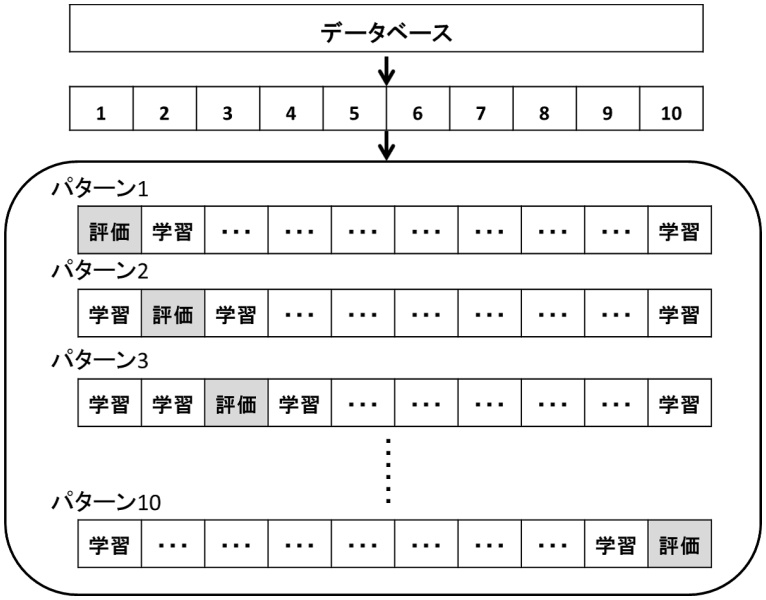


図 4.2: 10 通りのデータを作成するイメージ図

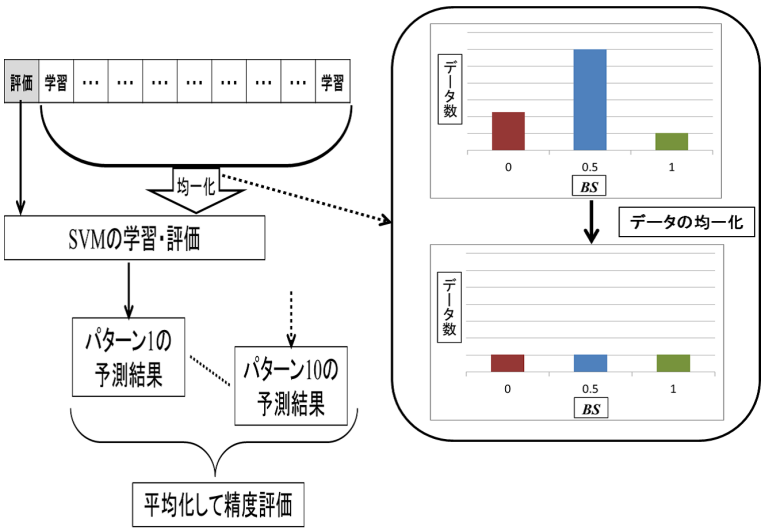


図 4.3: SVM での学習・評価のイメージ図

4.3　まとめ

本章では表声/裏声判別と息漏れ度合の判別に関して、第 3 章で作成した YUBA メソッドの初期段階に用いるための第 3 章で述べた歌唱音声データベースを使用して SVM を構築・評価する方法を示した。具体的には、従来の研究では 1 パターンのみの学習・評価であったものを、本研究では複数パターンの学習・評価をすることで、より正確な評価結果を導出できるように配慮した。

第5章　FMRの評価精度に関する検討

本章では、前章で述べたSVMを用いてFMRの判別精度を検証する。

5.1　従来のデータ構成での出力結果

まず、従来の研究で行なっていたように、YUBAメソッドの初期段階に限定しない学習データセットを用いた。このデータセットから構築したSVMを以後SVM0-$_{FMR}$と記す。評価データにはYUBAメソッドの初期段階に限定したデータと比較するために連続歌唱音声を除き、母音を/a/と/o/に絞ったものを用いた。この際に用いたデータベースは、学習データ、評価データ共に本研究で再構築したデータベースである。それを10グループのデータセットに分けて、10分割交差検証を行った。YUBAメソッドの初期段階においては、FMR=0とFMR=0.25の表声の区別、またはFMR=0.75とFMR=1の裏声の区別はあまり重要ではないので、FMRを5段階から3段階｛＜0.5、=0.5、＞0.5｝にカテゴライズした結果をみることにした。また、平均正解率、平均適合率、F値をそれぞれ算出した。F値は適合率と正解率の総合的な評価に使用し、F値が高いとバランスよく両方の精度が良いということを測ることができる。今回の場合、F値がもっとも高くなるのは適合率と正解率が100% のときの値で100である。図5.1はSVMの離散値出力での予測の正解率（データベース収録の専門家の評価を正解とする）、データ数、平均正解率、平均適合率、F値の各値の算出結果を表したものであり、10通りの評価から予測された各値のデータ数の総数と、そのデータ割合を10通りの評価の平均で整理した。また、対角線上の割合は各行のFMRに対しての専門家の判断とSVMの出力の正解率の平均を示している。図5.1をみると、表声（FMR＜0.5）の正解率は91.0%と高い精度を確保できたが、裏声（FMR＞0.5）の正解率は73.4%とそこまで高い精度は確保できないという結果になった。また平均正解率、平均適合率とF値をみると平均正解率は69.7%で平均適合率は69.2%、F値が69.4という結果になった。

5.2　本研究のデータ構成での出力結果

次に、YUBA メソッドの初期段階に限定した学習データセットを用いた。このデータセットから構築した SVM を以後 SVM1-$_{FMR}$ と記す。予測には SVM0-$_{FMR}$ と同じ評価データを用いた。図 5.2 は離散値の結果を 5 段階から 3 段階にカテゴライズした場合の結果と平均正解率（データベース収録の専門家の評価を正解とする）などを同じように示している。結果をみると表声（$FMR < 0.5$）の一致率は 87.1%、裏声（$FMR > 0.5$）の一致率は 90.7%となりいずれも 9 割近い正解率で非常に高い精度を確保することができるという結果なった。

SVM0-$_{FMR}$ と比較すると表声の正解率は 91.0%から 87.4%と少し低くなってしまうが、裏声の正解率は 73.4%から 90.7%まで大幅に上昇することがわかり、裏声判別の精度は良くなったといえる。しかし平均正解率をみると、SVM0-$_{FMR}$ は 69.7%で SVM1-$_{FMR}$ が 70.3%、平均適合率は SVM0-$_{FMR}$ は 69.2%で SVM1-$_{FMR}$ が 67.7%、F 値は SVM0-$_{FMR}$ は 69.4 で SVM1-$_{FMR}$ が 69.0 であり大きな差はない。このことから、SVM0-$_{FMR}$ と SVM1-$_{FMR}$ には詳細にみれば違いはあるものの、平均正解率などには有意な差はなく、SVM の学習に用いる音声を限定することの効果は得られなかったといえる。これは、FMR=0.5 のときの正解率が悪化していることが原因として挙げられ、この値の改善が今後の課題といえる。

5.3　まとめ

本章では、FMR の予測について YUBA メソッドの初期段階に限定しない学習データセットから構築した SVM（SVM0-$_{FMR}$）と、YUBA メソッドの初期段階に限定した学習データセットから構築した SVM （SVM1-$_{FMR}$）との比較を行った。その結果としては、SVM1-$_{FMR}$ の方が、裏声の判別精度が非常に高いという結果になったものの、平均正解率と平均適合率、F 値を比較してみるとほとんど変わらないという結果になった。このため、SVM に用いる音声を限定することの効果は得られないと判断される。しかし、従来の研究の平均正解率は 66%であったため、データベースを再構築したことにより少しではあるが精度は改善されている。

課題としては、FMR=0.5 の音声の判別精度が良くなかったため、その点を改善することが重要である。

FMR		SVMによる推定値			合計
		<0.5	=0.5	>0.5	
専門家	>0.5	1.0%(8)	25.6%(172)	73.4%(439)	100%(619)
	=0.5	34.3%(103)	44.6%(138)	21.1%(65)	100%(306)
	<0.5	91.0%(1554)	7.0%(115)	2.0%(32)	100%(1701)

平均正解率	平均適合率	F値
69.7%	69.2%	69.4

図 5.1: SVM0-$_{FMR}$ の正解率（データ総数）など

FMR		SVMによる推定値			合計
		<0.5	=0.5	>0.5	
専門家	>0.5	0.7%(6)	9.3%(66)	90.0%(547)	100%(619)
	=0.5	27.9%(83)	33.7%(106)	38.4%(117)	100%(306)
	<0.5	87.3%(1493)	8.3%(137)	4.4%(71)	100%(1701)

平均正解率	平均適合率	F値
70.3%	67.7%	69.0

図 5.2: SVM1-$_{FMR}$ の正解率（データ総数）など

第6章　BSの評価精度に関する検討

本章では、SVM を用いて BS の判別精度を検証する。

6.1　従来のデータ構成での出力結果

前章の FMR の出力結果と同様に BS の結果を示す。従来の研究で行っていたように、YUBA メソッドの初期段階に限定しない学習データセットを作成した。このデータセットから構築した SVM を以後 SVM0-$_{BS}$ と記す。評価データは FMR 同様、YUBA メソッドの初期段階に限定したデータセットを用いて、10 分割交差検証を行った。図 6.1 は、離散値で出力し 3 段階｛=0、=0.5、=1｝にカテゴライズした結果と平均正解率（データベース収録の専門家の評価を正解とする）と平均適合率、F 値を算出したものを示している。結果をみると息漏れのある歌声（BS=1）の正解率は 80.6%となり高い精度を確保できた。しかし、息漏れのない歌声（BS=0）の一致率は 63.7%と、少し低い結果となった。また、平均正解率と平均適合率、F 値をみると平均正解率が 60.2%、平均適合率は 49.5%となった。また F 値は 54.3 となった。これを、本研究で用いたデータの評価結果と比較する。

6.2　本研究のデータ構成での出力結果

次に、YUBA メソッドの初期段階に限定した学習データセットを作成した。このデータセットから構築した SVM を以後 SVM1-$_{BS}$ と記す。予測には SVM0-$_{BS}$ のときと同じ評価データを用い、10 分割交差検証を行った。その結果を図 6.2 に示す。SVM の離散値の出力結果を 3 段階にカテゴライズした場合の結果と平均正解率（データベース収録の専門家の評価を正解とする）などの各値を示している。結果をみると息漏れのない歌声（BS=0）の正解率は 82.6%、息漏れのある歌声（BS=1）の正解率は 83.6%となりいずれも高い精度を確保できた。また、BS=0、BS=1 の正解率がどちらも SVM1-$_{BS}$ の方が精度が良いという結果になった。特に BS=0 の正解率は 67.3%か

ら 82.6%と大幅に向上している。しかし、全体の平均正解率をみると SVM0-$_{BS}$ が 60.2%、SVM1-$_{BS}$ が 60.6%、平均適合率は SVM0-$_{BS}$ が 49.5%、SVM1-$_{BS}$ が 49.3%、F 値は SVM0-$_{BS}$ が 54.3、SVM1-$_{BS}$ が 54.4 であり、どれも従来の研究の結果とほとんど差がないという結果になった。このことから、SVM0-$_{BS}$ と SVM1-$_{BS}$ には詳細にみれば違いはあるものの、平均正解率などには有意な差はなく、SVM の学習に用いる音声を限定することの効果は得られなかったといえる。これは、BS=0.5 の正解率の悪化が原因で、50%以上が BS=0 に誤判定してしまうためであると考えられる。このことより、BS=0.5 の精度の向上がこれからの課題として挙げられる。

BS		SVMによる推定値			合計
		0	0.5	1	
専門家	0	67.3%(409)	24.0%(147)	8.7%(54)	100%(610)
	0.5	38.5%(619)	32.6%(528)	28.9%(466)	100%(1613)
	1	3.9%(30)	15.5%(63)	80.6%(324)	100%(403)

平均正解率	平均適合率	F値
60.2%	49.5%	54.3

図 6.1: SVM0-$_{BS}$ の正解率（データ総数）など

BS		SVMによる推定値			合計
		0	0.5	1	
専門家	0	82.3%(502)	9.1%(55)	8.6%(53)	100%(610)
	0.5	53.4%(860)	15.6%(253)	31.0%(500)	100%(1613)
	1	7.3%(30)	8.9%(35)	83.8%(338)	100%(403)

平均正解率	平均適合率	F値
60.6%	49.3%	54.4

図 6.2: SVM1-$_{BS}$ の正解率（データ総数）など

6.3　まとめ

本章では、BS の予測について YUBA メソッドの初期段階に限定しない学習データセットから構築した SVM（SVM0-$_{BS}$）と、YUBA メソッドの初期段階に限定した学習データセットから構築した SVM （SVM1-$_{BS}$）を用いる場合の判別制度の比較を行った。その結果としては、SVM1-$_{BS}$ の方が、息漏れのない歌声と息漏れのある歌声共に高い精度を得ることができたものの、平均正解

率、平均適合率、F 値をみると、値にほとんど差はなく SVM に用いる音声を限定することの効果は得られないと判断される。しかし、従来の研究の平均正解率は 55%であったため、データベースを再構築したことにより少しではあるが精度は改善されている。

課題として、BS=0.5 の音声の判別精度が良くなかったため、その点を改善することが重要である。

第7章　専門家の再評価精度に関する検討

本章では、専門家による FMR と BS の再評価と第5〜6章で示したSVMの評価とを FMR、BS それぞれで比較・検討する。

7.1　専門家の評価精度

第5〜6章で述べたようにSVMでの FMR の平均正解率は7割、BS の平均正解率は6割程度であり、学習・評価に用いる音声データを限定しても精度の改善は認められなかった。そこで、これらの正解率の値が持つ意味を確認するために専門家の判断の精度（安定性）を調べることを考えた。そのために、データベース作成時に協力してもらった同じ専門家1名に再度 FMR と BS の評価を依頼した。評価に使用した音声データは第3章で構築したデータベースから、FMR と BS の各値を母音/a/と/o/のバランスが良くなるように10個ずつ取り出した合計150個の音声データである。これらを1音1音ランダムに再生し、図3.1の評価イメージに基づき改めてラベリングを行ってもらった。その再評価の結果がデータベース収録の専門家の判断と比べ、どのくらい一致しているのかを確かめる。

7.2　FMR についての専門家の再評価結果

FMR について専門家の再評価の正解率を導いた。これはデータベース収録の専門家の判断を正解としたものである。結果を図7.1に表す。また、専門家に再評価してもらった音声を第5章で述べたYUBAメソッドの初期段階に限定しない学習データセットから構築したSVM（SVM0-$_{FMR}$）とYUBAメソッドの初期段階に限定した学習データセットから構築したSVM（SVM1-$_{FMR}$）を用いて評価し、専門家の再評価結果との比較を行った。その評価結果を図7.2、7.3に示す。それぞれの結果をみると、専門家は FMR=0.5の評価がばらついてしまい、それがSVMの評価にも同じような傾向として現れていることがわかる。表声（FMR $<$ 0.5）では専門家、SVMの評価が

共に 8 割を超える高い精度で予測できているが、裏声（FMR ＞ 0.5）では SVM0-FMR での評価が 58.3%と非常に低くなっている。

次に全体の平均正解率と平均適合率と F 値を比較する。平均正解率は専門家が 67.8%、SVM0-$_{FMR}$ が 70.0%、SVM1-$_{FMR}$ が 73.3%で、平均適合率は専門家が 68.2%、SVM0-$_{FMR}$ が 69.7%、SVM1-$_{FMR}$ が 72.5%で、F 値は専門家が 68.0、SVM0-$_{FMR}$ が 69.8、SVM1-$_{FMR}$ が 72.9 であり、どの値をみても、それぞれに大きな差はなく、SVM による評価が専門家の評価とほぼ同等であることがわかった。

7.2.1　FMR についての専門家の再評価結果と SVM の評価結果の関係

専門家の再評価がデータベース収録の専門家の判断と一致している音声（110 個）に限定して、SVM0-$_{FMR}$ と SVM1-$_{FMR}$ の異なる 2 つの SVM による予測の一致率を改めて算出した。その結果を図 7.4 に示す。結果をみると SVM0-$_{FMR}$ は裏声の一致した割合が 51.2%と低かった。しかし SVM1-$_{FMR}$ は表声、裏声の両方に関して一致した割合が 9 割近くあり、専門家と同じ評価ができているという結果になった。また FMR=0.5 の歌声に関してはどちらも 50%程度と低い結果になった。また、全体の一致率をみると、SVM1-$_{FMR}$ が 85.5%と高いことから、専門家と同等の評価が行われており、YUBA メソッドの初期段階トレーニングに使用できる見込みがあるといえる。

7.3　BS についての専門家の再評価結果

本節では BS について専門家の再評価の正解率を導いた。こちらも FMR 同様にデータベース収録の専門家の判断を正解としている。結果を図 7.5 に示す。また、専門家に再評価してもらった音声を第 6 章で述べた YUBA メソッドの初期段階に限定しない学習データセットから構築した SVM（SVM0-$_{BS}$）と YUBA メソッドの初期段階に限定した学習データセットから構築した SVM（SVM1-$_{BS}$）を用いて評価し、専門家の再評価結果との比較を行った。それぞれの結果を図 7.6、7.7 に示す。それぞれの結果をみると、息漏れのある歌声（BS=1）に関しては SVM の方が高い精度で正解している。一方、BS=0、BS=0.5 のときの正解率が専門家の方が高い精度を確保している。また、全体の平均正解率は専門家が 45.3%、SVM0-$_{BS}$ が 44.0%、SVM1-$_{BS}$ が 42.0%となり専門家と SVM でほとんど差がないという結果になった。平均適合率をみると専門家が 50.4%、SVM0-$_{BS}$ が 46.5%、SVM1-$_{BS}$ が 41.3%であり、専門家の評価の方が有効であるという結果になっ

FMR		専門家の再評価			合計
		<0.5	=0.5	>0.5	
専門家	>0.5	0.0%(0)	18.3%(11)	81.7%(49)	100%(60)
	=0.5	23.3%(7)	40.0%(12)	36.7%(11)	100%(30)
	<0.5	81.7%(49)	18.3%(11)	0.0%(0)	100%(60)

平均正解率	平均適合率	F値
67.8%	68.2%	68.0

図 7.1: FMR について専門家の再評価の正解率（データ総数）など

FMR		SVMによる推定値			合計
		<0.5	=0.5	>0.5	
専門家	>0.5	1.7%(1)	40.0%(24)	58.3%(35)	100%(60)
	=0.5	20.0%(6)	56.7%(17)	23.3%(7)	100%(30)
	<0.5	95.0%(57)	1.7%(1)	3.3%(2)	100%(60)

平均正解率	平均適合率	F値
70.0%	69.7%	69.8

図 7.2: SVM0-$_{FMR}$ の正解率（データ総数）など

FMR		SVMによる推定値			合計
		<0.5	=0.5	>0.5	
専門家	>0.5	0.0%(0)	15.0%(9)	85.0%(51)	100%(60)
	=0.5	10.0%(3)	43.3%(13)	46.7%(14)	100%(30)
	<0.5	91.7%(55)	5.0%(3)	3.3%(2)	100%(60)

平均正解率	平均適合率	F値
73.3%	72.5%	72.9

図 7.3: SVM1-$_{FMR}$ の正解率（データ総数）など

た。これは SVM が BS=0 の評価を BS=1 に真逆に判定することが多い点と、BS=0.5 の評価精度が悪いことが原因といえる。また、F 値については専門家が 47.7、SVM0-$_{BS}$ が 45.2、SVM1-$_{BS}$ が 41.6 であり、SVM1-$_{BS}$ が少しだが劣るという結果になった。しかし全体的にそこまで大きな差はないことから、SVM で専門家と同等の評価が可能であると考える。

SVM0-$_{FMR}$	<0.5	0.5	>0.5	全体
一致率 (SVM/専門家)	93.9% (46/49)	58.3% (7/12)	51.2% (29/49)	74.5% (82/110)

SVM1-$_{FMR}$	<0.5	0.5	>0.5	全体
一致率 (SVM/専門家)	89.8% (44/49)	50.0% (6/12)	89.8% (44/49)	85.5% (94/110)

図 7.4: FMR について専門家の再評価と SVM の評価の比較

7.3.1　BS についての専門家の再評価結果と SVM の評価結果の関係

専門家の再評価がデータベース収録の専門家の判断と一致している音声（68 個）に限定して、SVM0-$_{BS}$ と SVM1-$_{BS}$ の異なる 2 つの SVM による予測の一致率を改めて算出した。その結果を図 7.8 に示す。結果をみるとどちらも BS=1 の歌声に関しては 90%程の高い割合で一致している。また、SVM0-$_{BS}$ は BS=0 の一致した割合が 50.0%と低かった。それに対して SVM1-$_{BS}$ は BS=0 の一致した割合が 7 割近くあり、専門家と同じような評価ができているという結果になった。また BS=0.5 の歌声に関してはどちらも 20%程度と低い結果になった。全体の一致率はどちらも 51.5%と低い予測精度となった。そのため、BS の予測精度はまだ改善が必要なレベルであり、YUBA メソッドに使用することは現段階では難しいと考えられる。

7.4　まとめ

本章では、専門家の再評価と SVM による FMR と BS の予測結果の比較・検討を行った。まず、データベース収録の専門家の判断を正解とした場合の、FMR については専門家の再評価と、SVM0-$_{FMR}$ と SVM1-$_{FMR}$ の異なる 2 つの SVM による予測結果の平均正解率、平均適合率、F 値を比較した。専門家と SVM0-$_{FMR}$、SVM1-$_{FMR}$ で平均正解率が 70%程度でほとんど差はなく、平均適合率、F 値も共に同じような値であったことから、専門家とほぼ同等の評価といえる。また、BS についても専門家の再評価と、SVM0-$_{BS}$ と SVM1-$_{BS}$ の異なる 2 つの SVM による予測結果の平均正解率、平均適合率、F 値を比較した。専門家と SVM0-$_{BS}$、SVM1-$_{BS}$ で平均正解率が 45%程度で差はなく、平均適合率、F 値も共に同じような値であったことから、FMR 同様 BS の予測に差はなく、こちらも専門家とほぼ同等の評価といえる。

BS		専門家の再評価			合計
		0	0.5	1	
専門家	0	44.0%(22)	46.0%(23)	10.0%(5)	100%(50)
	0.5	26.0%(13)	58.0%(29)	16.0%(8)	100%(50)
	1	4.0%(2)	62.0%(31)	34.0%(17)	100%(50)

平均正解率	平均適合率	F値
45.3%	50.4%	47.7

図 7.5: BS について専門家の再評価の正解率（データ総数）など

BS		SVMによる推定値			合計
		0	0.5	1	
専門家	0	28.0%(14)	20.0%(10)	52.0%(26)	100%(50)
	0.5	18.0%(9)	32.0%(16)	50.0%(25)	100%(50)
	1	4.0%(2)	24.0%(12)	72.0%(36)	100%(50)

平均正解率	平均適合率	F値
44.0%	46.5%	45.2

図 7.6: SVM0-$_{BS}$ の正解率（データ総数）など

BS		SVMによる推定値			合計
		0	0.5	1	
専門家	0	38.0%(19)	14.0%(7)	48.0%(24)	100%(50)
	0.5	34.0%(17)	18.0%(9)	48.0%(24)	100%(50)
	1	14.0%(7)	16.0%(8)	70.0%(35)	100%(50)

平均正解率	平均適合率	F値
42.0%	41.3%	41.6

図 7.7: SVM1-$_{BS}$ の正解率（データ総数）など

また、専門家の評価がばらついてしまっているが、これは再評価するまでに 2 年間という時間が空いてしまったことから、評価基準が少しだが変わった可能性がある。それに加えて評価方法も違っており、1 回目の評価では収録音声をフレーズ毎に聞いて判断していたが、再評価では 1 音毎に聞いて評価していた。専門家はフレーズ毎に聞いた方が判定しやすい可能性もある。このことから、評価がばらついてしまったと判断するが、今後このことを確かめるために、専門家の再評価回数、再評価データ数を増やすことや、専門家の増員による評価の平均化によって教師デー

SVM0-$_{BS}$	0	0.5	1	全体
一致率 (SVM/専門家)	50.0% (11/22)	27.6% (8/29)	94.1% (16/17)	51.5% (35/68)

SVM1-$_{BS}$	0	0.5	1	全体
一致率 (SVM/専門家)	68.2% (15/22)	17.2% (5/29)	88.2% (15/17)	51.5% (35/68)

図 7.8: BS について専門家の再評価と SVM の評価の比較

タを見直すことも重要になってくると考えられる。

次に、専門家の再評価がデータベース収録の専門家の判断と一致している音声に限定して、FMR では SVM0-$_{FMR}$ と SVM1-$_{FMR}$ の BS では SVM0-$_{BS}$ と SVM1-$_{BS}$ のそれぞれ異なる 2 つの SVM による一致率を算出した。FMR については、SVM1-$_{FMR}$ の予測結果との一致率が、85.5%と高い精度を示すことから、YUBA メソッドの初期段階に限定したトレーニングには使用できる見込みがあることがわかった。一方、BS は一致率が SVM0-$_{BS}$、SVM1-$_{BS}$ 共に 51.5%と低い予測精度で、まだ精度の改善が必要であるという結果になった。

これより、専門家の再評価の結果が特に BS で低いことから、データベースに収録されている判定結果を見直す必要があると考えられる。そのため、専門家の再評価データ数や専門家の数や評価回数を増やすことが課題として挙げられる。

第8章　総括

本研究はYUBAメソッドに基づく個人による歌唱トレーニングの効率化を目的に、音声の物理的パラメータを入力とする機械学習モデルを、発声音の声質判定に利用しようという一連の研究に属するものである。本研究では声質の評価対象を特にYUBAメソッドの初期段階に限定してSVMによる判定精度の向上を目指すと共に、ヒトによる判定結果との精度の比較も行った。

以下に各章の内容を要約する。

第1章では、本研究の背景・目的について説明を加えた。

第2章では、ヒトの音声の特徴とともに研究の遂行に必要となる表声/裏声の発声メカニズムと歌唱トレーニング法『YUBAメソッド』について概説した。

第3章では、新たに再構築したデータベースの作成手順と収録されている音声のパラメータなどについて概説した。収録されている10,295件の音声サンプルの中に発声が安定しておらず音程が不安定なサンプルが含まれていたため、そのような音声サンプルを取り除き、再度VoiceSauceという解析ソフトで音声の物理パラメータを算出し、データベースを再構築した。その結果、有効な音声サンプルは9,329件となった。

第4章では、表声/裏声（FMR）と息漏れ度合（BS）を予測と評価をするための学習・評価データの作成法、そのためのSVMの構成について概説した。本研究では、データベース収録のデータを10個のデータセットに分け、SVMによる判別精度を評価するために10分割交差検証を行った。この概要について説明した。

第5章では、FMRの判別を行う以下の2つのSVMについて判別精度を比較・検討した。

- 発声音声をYUBAメソッド初期段階に限定しないデータセットで学習したSVM
  （SVM0-$_{FMR}$と表記）
- 発声音声をYUBAメソッド初期段階に限定したデータセットで学習したSVM
  （SVM1-$_{FMR}$と表記）

これらの SVM について FMR の判別結果の平均正解率、平均適合率、F 値を算出した。平均正解率は SVM0-$_{FMR}$ が 69.7%で SVM1-$_{FMR}$ が 70.3%、平均適合率は SVM0-$_{FMR}$ が 69.2%で SVM1-$_{FMR}$ が 67.7%、F 値は SVM0-$_{FMR}$ が 69.4 で SVM1-$_{FMR}$ が 69.0 となり、いずれも大きな差はなかった。詳細にみれば違いはあるものの、平均的には SVM の学習に用いる音声を限定することの効果は得られなかった。また FMR=0.5 の予測精度が低かったため、この点を改善することが重要になる。

第 6 章では、BS の判別を行う以下の SVM について判別精度を比較・検討した。

- 発声音声を YUBA メソッドの初期段階に限定しないデータセットで学習した SVM（SVM0-$_{BS}$ と表記）
- 発声音声を YUBA メソッド初期段階に限定したデータセットで学習した SVM（SVM1-$_{BS}$ と表記）

これらの SVM について平均正解率、平均適合率、F 値を算出した。平均正解率は SVM0-$_{BS}$ が 60.2%で SVM1-$_{BS}$ が 60.6%、平均適合率は SVM0-$_{BS}$ が 49.5%で SVM1-$_{BS}$ が 49.3%、F 値は SVM0-$_{BS}$ が 54.3 で SVM1-$_{BS}$ が 54.4 となり、いずれも大きな差はなかった。FMR と同様で詳細にみれば違いはあるものの、平均的には SVM の学習に用いる音声を限定することの効果は得られなかった。また FMR と同様に BS=0.5 の予測精度が低かったため、この点を改善することが重要になる。

第 7 章では、第 5〜6 章での FMR、BS の正解率の値が持つ意味を確認するために、専門家の判断の精度を調べた。そのために、FMR と BS の各値のバランスを考え、ランダムに抽出した 150 件の音声データについて専門家に再評価を依頼し、その再評価結果と SVM による評価結果とを比較した。具体的には、FMR では専門家の再評価と SVM0-$_{FMR}$、SVM1-$_{FMR}$ の予測結果を、データベース収録の専門家の判断を正解として、平均正解率や平均適合率、F 値で比較した。その結果、いずれの平均正解率も 70%程度であり、平均適合率、F 値もほぼ同等の評価であった。BS も FMR と同様に専門家の再評価と SVM0-$_{BS}$、SVM1-$_{BS}$ の予測結果を平均正解率や平均適合率、F 値で比較した。平均正解率は専門家を含めていずれも 45%程度であり、平均適合率、F 値もほぼ同等の評価であった。このことから FMR、BS 共に平均的には専門家と同等の評価を達成していることがわかった。

また、専門家の再評価とデータベースに登録されている判定結果が一致する音声に限定して、FMR、BS のそれぞれ異なる 2 つの SVM による予測結果の一致率を調べた。FMR については、

SVM1-$_{FMR}$ の予測結果との一致率が 85.5%と高い精度を示すことから、YUBA メソッドの初期段階に限定したトレーニングには使用できる見込みがあることがわかった。一方で BS の一致率は SVM0-$_{BS}$、SVM1-$_{BS}$ 共に 51.5%と低い予測精度で、まだ精度の改善が必要であることがわかった。

今後の課題としては、FMR=0.5、BS=0.5 の SVM による評価精度を上げ、平均正解率を向上させるために学習条件や入力要素の検討を引き続き行っていくことが挙げられる。また、専門家の再評価結果が特に BS では低かったため、専門家による再評価データ数や再評価回数を増やすこと、専門家の増員も課題として挙げられる。加えて、最終的には $FMR \cdot BS$ の分析結果をトレーニング中の訓練者にリアルタイムに視覚情報としてフィードバックすることを目標としているが、どのように訓練者に表示させるのかについても検討が必要である。

謝辞

本研究の遂行及び本論文作成に際し、終始多大なる御指導並びに御助言を賜った竹尾隆教授、野呂雄一准教授、歌唱トレーニング法についての御助言と御協力を賜った弓場徹教授に心より感謝の意を表します。また、本研究のために御協力下さった山本好弘三重大学工学部技術職員並びに院生、学部生諸氏、鈴鹿大学短期大学部助手大久保友加里氏に深く御礼申し上げます。

References

[1]テレビ番組『すイエんサー』ＮＨＫＥテレ（2012）

[2]テレビ番組『元気家族テレビ　となりのマエストロ』MBS 毎日放送 TBS 系列　情報バラエティ番組（2010）

[3]テレビ番組『DON!』日本テレビ（2010）

[4]テレビ番組『シルシルミシル』テレビ朝日（2010）

[5]換声点ショックの起こる音域（弓場の造語,2014）

[6]弓場徹『奇跡のボイストレーニング BOOK（CD 付）』（主婦の友,2004）

[7]弓場徹『奇跡のハイトーンボイストレーニング（プログラム CD 付）』（主婦の友,2006）

[8]弓場徹『CD をまねるだけ！歌のうまい子になる超簡単ボイストレーニング（CD 付）』（朝日新聞出版,2010）

[9]弓場徹, 他『合唱指導―悩みと疑問大解決』（音楽之友社）

[10]弓場徹, 他『JOHNS「音楽・音声・環境音－音の世界と耳鼻咽喉科」』（東京医学社）

[11]弓場徹, 他『耳鼻咽喉科・頭頸部外科クリニカルトレンド Part3』（中山書店）

[12]弓場徹『「歌う筋肉」 男声編』（ビクターエンタテインメント）

[13]弓場徹『「歌う筋肉」 女声編』（ビクターエンタテインメント）

[14]弓場徹『YUBA メソッド　初級ボイストレーニング編　あっという間に歌上手 1 』（フィークジャパン株式会社,2009）

[15]弓場徹『YUBA メソッド 中級ボイストレーニング編　驚異のカラオケ上達法 1 』（フィークジャパン株式会社,2009）

[16]弓場徹『「YUBA メソッド」による新発声指導法 1　歌う筋肉の秘密 』（ビクターエンタテインメント,2007）

[17]弓場徹『「YUBA メソッド」による新発声指導法 2　歌う筋肉強化トレーニング　変声前児童（小学生）＆女声用』（ビクターエンタテインメント,2007）

[18]弓場徹『YUBA メソッド」による新発声指導法 3　歌う筋肉強化トレーニング　混声用（中学校・高等学校）』（ビクターエンタテインメント,2007）

[19]講演『YUBA メソッドによるボイストレーニング法～音声障害にならないための声作り』日本吟剣詩舞振興会主催　東日本指導者研修会（2011）

[20]苅安誠, 城本修『改訂　音声障害』（建帛社,2012）

[21]福島結香, スペクトル構造に基づく表声/裏声判別指標に関する検討, 三重大学修士論文　(2013)

[22]浅野翔大, 裏声判別指標を用いたボイストレーニングソフトウェア, 三重大学修士論文　(2014)

[23]岩本享大, 機械学習に基づく歌唱音声の声質評価システムの構築, 三重大学修士論文　(2014)

[24]http://www.ee.ucla.edu/ spapl/voicesauce

[25]郡司隆男, 西垣内泰介,「ことばの科学ハンドブック」(研究社,2004)

[26]http://speechresearch.flw-web.net/112.html

[27]http://www.neuro.sfc.keio.ac.jp/ masato/study/SVM/index.htm

[28]http://www.cs.waikato.ac.nz/ml/weka/

[29]石井カルロス寿憲, 息漏れの自動検出における音響パラメータの提案,　信学技報　(2004)

[30]Akihito Sakai, Study on Quality Evaluation of Singing Voices using SVM -Optimization of input Elements for Beginners Stage of YUBA Method-, proceedings of IS^2EMU2016-C,(2016)