

ボイストレーニングのための 声質改善計測ツール

相川清明[†] 上沼純子^{††} 秋竹朋子^{††}

ボイストレーニングによる声質改善を定量的に評価するためのシステムについて述べる。ボイストレーニングは呼吸、発声、調音、表情の矯正を行うが、それらの要因を同時かつ定量的に評価することは難しい。そこで、音源と声道の特徴を反映できる4種の音響パラメータを用いたリアルタイム声質計測ツールを作成した。音響パラメータとしては、倍音強度、スペクトルダイナミックレンジ、ホルマントQ、スペクトルの傾きを用いた。実験の結果、呼吸の改善が4種すべての音響パラメータの向上に重要であること、発声の質とスペクトルダイナミックレンジの関連が深いことが分かった。

A Tool for Measuring Voice Quality Improvement by Voice Training

Kiyoaki Aikawa[†], Junko Uenuma^{††} and Tomoko Akitake^{††}

This report describes a system for quantitatively evaluating voice quality improvement by voice training. Voice-training corrects respiration, phonation, articulation, and facial expression. It is difficult to simultaneously and quantitatively evaluate these four factors. A real time measuring system was developed based on acoustical feature parameters reflecting voice source and vocal tract characteristics. The acoustic parameters included harmonics intensity, spectral dynamic range, formant Q, and spectral slope. Experimental results indicated that abdominal respiration was most important for improving these four parameters. Spectral dynamic range was strongly correlated with phonation.

1. はじめに

ボイストレーニングでは、呼吸、発声、調音、表情などの訓練が行われる。従来では、声質の向上の評価はトレーニング教員による主観的な判断に委ねられていた。しかし、声質の複数の評価項目を同時にかつ定量的に評価することは難しい。

従来から声質は音声合成[1]、VoIPなどによる音声伝送[2]、医療診断の観測手段[3-5]などで研究されてきた。歌唱音声の音響的分析は多く報告されているが、主として基本周波数軌跡と歌唱ホルマントに関する研究である[6-13]。声質の音響分析についての報告もあるが[14-16]、ボイストレーニングの効果の系統的分析については報告されていない。

既に4種類の音響特徴が声質の向上を反映していることを示した[17-18]。本報告では、それらの音響特徴を用いた声質改善を計測するためのツールについて述べ、既に提案している音響特徴と声質の向上との関連を調査した結果について述べる。

2. ボイストレーニング

2.1 呼吸

ボイストレーニングでは腹式呼吸による呼気で強く発声するように指示される。このことは、音源波形、声帯振動の安定性などに関係するため、音源スペクトルの傾きや倍音構造の顕在性に特徴が表れると考えられる。呼気圧が高く、声帯が緊張すれば、短いパルス状の気流を発生するため、高次高調波を多く含む音源波形が生成されると考えられる。図1は声帯が開く時間による音源波形の高調波成分の違いを示している。声帯の開いている時間が短く、音源波形がインパルスに近づくほど高次倍音が強くなる。トランペットの鋭い立ち上がり部分において、高次倍音が明瞭にみられることから、高次倍音の強さについては、声帯の緊張によるものを含めた剛性、気管内部の圧力や壁面の状態が関係すると考えられる。

また、声帯の規則正しい振動も倍音構造を明瞭にする。声帯振動が規則的であれば、高調波のスペクトルはラインスペクトルに近づく。一方、振動周期が不規則であれば、スペクトルの幅が拡大し、各高調波の中心強度は低下する。

この他、声帯における空気の漏れは雑音成分を増加させ、倍音構造が雑音スペクトルに埋もれてしまう。

[†]東京工科大学メディア学部
Tokyo University of Technology, School of Media Science

^{††}ビジヴォ
Bijivo

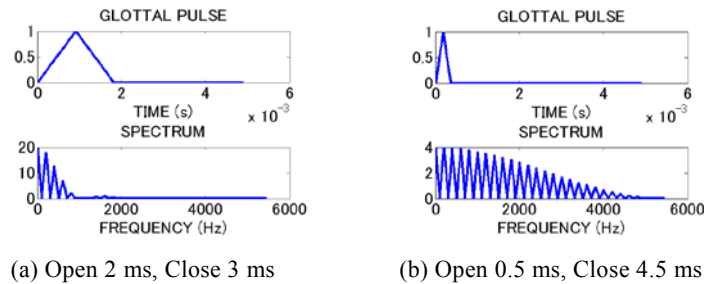


図 1 声帯開時間の長短による音源波形のスペクトルの違い. ($F_0=200$ Hz)

2.2 発声

声帯咽頭部の空間、筋肉の緊張の制御、軟口蓋組織の物理的性質などにより共振（共鳴）を向上させる。特に、歌唱においては口腔の制御だけでなく、鼻腔の共鳴も重視される。発声の特徴はホルマントの共振の強さ、あるいはスペクトル包絡の起伏の大きさに表れると考えられる。

2.3 調音

口腔、舌位置を制御して音素の特徴を生成することで、構音とも呼ばれる。音素の音響的特徴は口腔の共振に相当するホルマントの周波数配置により表わされる。音素の明瞭性には、これらのホルマントの共振が十分強いかどうかに関係する。共振の形成には、口腔内の形状だけでなく、口腔壁面の物性も関係する。また、頸部角度、顎の位置、舌の位置、口唇の形状が関係する。なお、口唇の開口面積は音声の放射特性とも関係する。調音の特徴は包括的にはスペクトル包絡のダイナミックレンジで表わされると考えられる。また、局所的にはスペクトルを声道フィルタと見た時のホルマントの Q で表わされると考えられる。

2.4 表情

ボイストレーニングにおいては、表情の制御も重要視される。表情には表情筋と呼ばれる複数の顔面の筋が関係する。表情筋の制御は口唇開口形状に影響し、音声の放射特性とスペクトルに関係する。口唇の開口は音素により異なるため、パワーの変動に影響が表れると考えられる。音声には鼻孔から放出される鼻音があるので、表情はこの成分との強度比とも関連がある。

3. 評価用音響特徴

3.1 倍音成分強度

図 1 に示したように、音源の性質は倍音構造の強さにより評価できる。倍音構造が明瞭に見られるということは、スペクトル上での基本周波数 F_0 の整数倍の周波数にピークが明瞭に見られるということである。そこで、基本周波数に対応するケプストラム係数の大きさにより、倍音成分の強さを計測することにした。予備実験によると、ベルカント唱法での発声では高域まで鋭い高調波が観測されている。フレーム n の倍音強度は以下の式で与えられる。

$$h(n) = \max_{k_{\min} \leq k \leq k_{\max}} c_k(n) \quad (1)$$

ここで、 $c_k(n)$ はフレーム n の k 次のケプストラム係数を表す。 k_{\min} と k_{\max} は、分析対象とする F_0 の最大値と最小値に対応するケプストラム次数である。

3.2 パワー変動

口唇開口部の変化幅の大きさを計測するには、対数パワー時系列の分散が適当であると考えられる。

3.3 スペクトルの時間変化

デルタケプストラムはスペクトル変化の大きさを表す。このデルタケプストラムの 2 乗和で定義される動的尺度により、調音の明瞭さを計測することにした[19]。デルタケプストラムはケプストラム時系列の部分列の一次回帰直線の傾きで定義され、次数ごとに求められる。 k 次のデルタケプストラム $d_k(n)$ はケプストラム時系列 $c_k(n)$ から次式により求められる。 n はフレーム番号を表す。

$$d_k(n) = \frac{\sum_{l=-L}^L w(l) n c_k(n+l)}{\sum_{l=-L}^L l^2} \quad (2)$$

$$w(l) = \begin{cases} \frac{l+L}{L} & -L \leq l \leq 0 \\ \frac{L-l}{L} & 1 \leq l \leq L \end{cases}$$

ここで、 $w(l)$ は三角窓を表す。窓幅は $2L+1$ となる。この窓はケプストラム時系列に対するデータ窓として機能する。

3.4 スペクトル包絡ダイナミックレンジ

ホルマントの共振が強ければ、結果的にスペクトル形状のダイナミックレンジが大きくなる。ここでは、線形予測分析に基づくスペクトル包絡を用いる。低次の線形予測分析により、倍音構造を除いたスペクトル概形を求めることができる。高域ホルマントの推定精度向上のため $1-0.98z^{-1}$ によりプリアンファシスを行う。また、対数スペクトル包絡の一次回帰直線成分を除去するという手段で個人性や音素に依存したスペクトルの傾きを除去する。傾きを除去した対数スペクトル包絡の2乗平均値すなわち分散により、ダイナミックレンジを表現する。

線形予測分析によるスペクトル包絡は、以下の式で表わされる。

$$S(z) = \frac{1}{\sum_{k=0}^p a_k z^{-k}} \quad (3)$$

ここで a_k は k 次の線形予測係数、 p は極数である。フレーム n のスペクトルの傾きはスペクトル包絡 $S(i, n)$ から以下の一次回帰直線の傾きにより求める。このスペクトルの傾きの値自身も音響特徴として利用する。

$$v(n) = \frac{\sum_{i=-1/2}^{1/2} S(i, n) i}{\sum_{i=-1/2}^{1/2} i^2} \quad (4)$$

ここで、 $i=-1/2$ と $i=1/2$ はそれぞれ最小、最高周波数に相当する。

3.5 ホルマントQ

k 次線形予測係数 a_k を用いるとスペクトル $S(z)$ は以下の式で表わされる。

$$S(z) = \frac{z^p}{\sum_{k=1}^p (z - u_k)} \quad (5)$$

ここで u_k は k 番目の極を表す。 k 番目の極の-3dB バンド幅 b_k と Q 値 Q_k は、極の半径 r_k と角度 θ_k を用いて以下のように表される。

$$u_k = r_k e^{j\theta_k} \quad (6)$$

$$b_k = 2 \arcsin\left(\frac{1-r_k}{2r_k}\right) \sqrt{-r_k^2 + 6r_k - 1} \quad (7)$$

$$Q_k = \theta_k / b_k \quad (8)$$

式(7)のバンド幅はバンド幅制御目的で用いられていた以下の関数に近い。

$$b_k = -2 \log r_k \quad (9)$$

4. 声質改善計測ツール

4.1 システム構成

前述の音響特徴を用いて声質を評価するためのシステムを作成した。図 2 に声質評価ツールの GUI を示す。システムは MATLAB を用いて作成され、PC 上で動作する。音声を取り込み、瞬時に分析結果を表示するという動作を繰り返す。

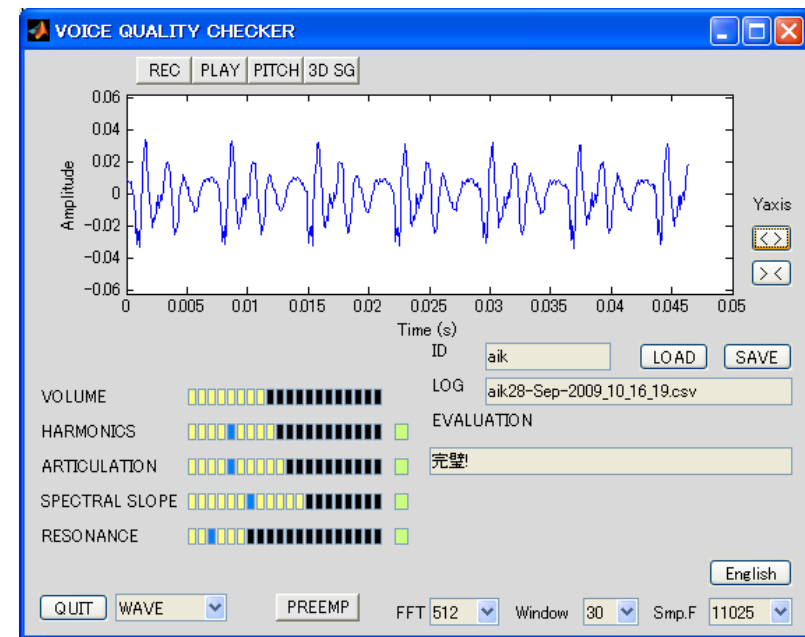
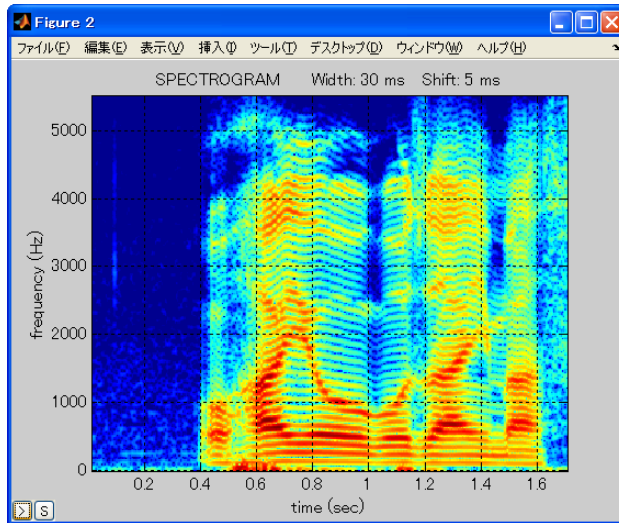
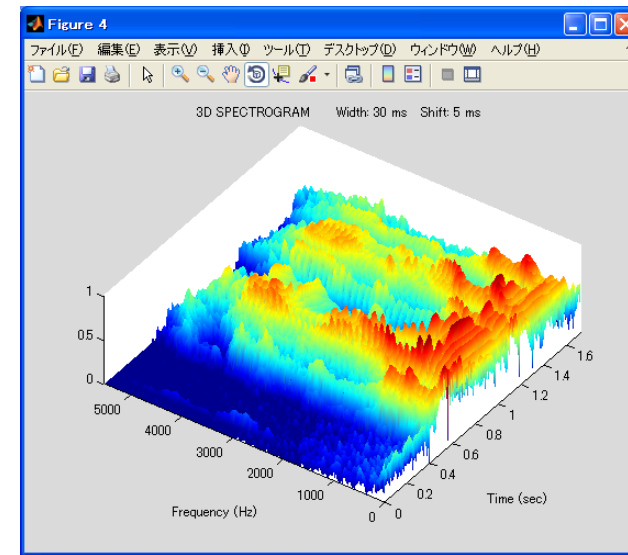


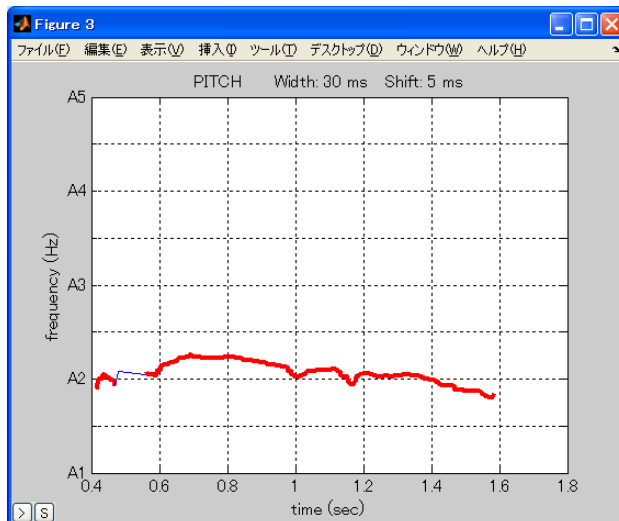
図 2 声質改善計測ツール(Voice Quality Checker)GUI.



(a) スペクトログラム



(c) 3D スペクトログラム



(b) F0 軌跡

図 3 音声切り出しモードで収録した音声「おはようございます」の分析。

4.1 ディスプレイ

上部の表示窓には、波形、FFT スペクトル、音源スペクトル、LPC スペクトル包絡、を表示できるようになっており、左下 QUIT ボタンの右のポップアップメニューにより選ぶことができる。メニューにはもう 1 つ音声自動切り出しモードが選択できる。

4.2 レベルメータ

左下にあるレベルメータは、上から、音量、倍音強度、スペクトルダイナミックレンジ、スペクトル傾き、ホルマント Q のレベルを表示する。それぞれは 20 のセグメントから成り、フルスケールで高品質音声の状態を示すようにレベルを調整している。セグメントは消灯時には黒、点灯時には黄色となる。

4.3 フレーズ分析機能

音声切り出しモードを選ぶと、単語や句単位の一連の音声を自動的に切り出し、図 3 (a) のスペクトログラムを表示する。上部のボタンにより、録音音声の確認、図 3 (b) の基本周波数軌跡、図 3 (c) の 3D スペクトログラムが表示できる。

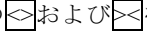

4.4 トレーニング前の値の記録と呼び出し

SAVE ボタンにより現在の4種の音響特徴の値がCSV形式のファイルに保存できる。このとき、IDで指定した識別子を先頭とし、それに保存年月日時分秒が続くファイル名となる。LOAD ボタンを押すとIDで指定された識別子を先頭とするファイル一覧が表示され、その中から適切な基準ファイルを選択できる。このファイルの情報はレベルメータ上で青色のマーカ―として表示される。現在の入力がこの基準を超えると、レベルメータのすぐ右側のランプが緑色に点灯する。右下の評価窓には、最もレベルメータの値が低い特徴を矯正するための指示が表示される。

4.5 分析条件

サンプリング周波数の初期設定値は11025 Hzである。サンプリング周波数としては22050 Hz, 44100 Hzも選択できるが、音声分析には11025 Hzが適切である。スペクトル分析はサンプル数256すなわち23.2 msを初期設定値とした。サンプル数としては512, 1024, 2048も選択できる。窓関数にはHanning窓を用いた。線形予測分析次数は14とした。F0推定範囲は80 Hzから400 Hzとした。ディスプレイの音源スペクトルは、リフタリングにより緩やかな変化を取り除いた対数スペクトルとして求めた。

4.6 その他の機能

右側のおよびを表示したボタンは波形表示選択時に縦軸を拡大または縮小するためのものである。下部のPREEMPボタンはスペクトルを平坦に近い形で表示するためのプリエンファシスのON/OFFを切り換える。右下の言語選択ボタンは、ボタンを押すたびに、コメント窓の記述言語が日本語、英語で切り替わる。

5. 実験

5.1 声質改善に関連のある音響特徴

20歳から54歳までの男性10名女性2名の被験者計12名のトレーニング前後の音声を比較した。発声内容は短文または孤立単語列である。音声はサンプリングレート8 kHzで収録され、AMR符号化により記録されたものである。雑音部は手動で除去した。分析は32 ms (256 サンプル) を1フレームとして10 msごとに行った。線形予測分析次数は10次である。本報告では、平均対数パワー以上を有声区間とみなし、この条件を満たすフレームを分析に使用した。すなわち、これらのフレームについて被験者ごとに求めた平均値をボイストレーニング前後で比較した。総フレーム数はトレーニング前が9341、トレーニング後9811である。

分析結果を表1に示す。ANOVAは被験者全体としてトレーニング前後の違いにつ

いて分散分析した結果である。また、被験者ごとにトレーニング前後での音響特徴量の差を取り、t検定を行った結果をT-TESTに示す。UPは値が増加した人数を表す。必ずしも全員の値が向上しているわけではないが、1名の被験者ですべての音響特徴の値が低下したのではない。基本周波数はトレーニング後上昇することが多いので、参考のため平均基本周波数についても有意差を分析した。

分散分析の結果によると、倍音強度、スペクトルダイナミックレンジ、ホルマントQにおいてトレーニング前後で有意差が見られる。t検定の結果では、上記の特徴に加えてスペクトルの傾きにおいて、分散分析よりさらに顕著な有意差が見られることが分かる。基本周波数はトレーニング前後で有意差はないので、音響特徴量の変化は基本周波数によるものではない。

表1 トレーニング前後での分散分析とt検定結果。ANOVA:トレーニング前全体と後全体の比較。UP:値が向上した人数(12名中)。T-TEST:被験者ごとの値の差の検定。

音響特徴量	ANOVA	UP	T-TEST
倍音強度	0.0084	11	0.0002
パワー分散	0.8288	7	0.3542
スペクトル変化	0.9628	5	0.4795
スペクトルダイナミックレンジ	0.0217	12	0.0005
ホルマントQ	0.0041	11	0.0008
スペクトル傾き	0.1217	11	0.0005
基本周波数	0.4754	8	0.1198

5.2 主観評価と音響特徴量の関係

26歳から56歳までの男性3名と女性2名の複数回のトレーニング総計9組のデータを用い、主観評価と音響特徴の関連を調べた。音声は44100 Hzで収録され、WMA符号化により記録されたものである。このデータを11025 Hzにダウンサンプルして用いた。分析は23.3 ms (256 サンプル) を1フレームとし、10 msごとに行った。線形予測分析次数は10次である。

表2にトレーニング前後で有意差が見られた音響特徴と主観的な評価との相関を示す。主観評価は、ボイストレーナ1名に呼吸、発声、調音、表情の改善を3段階評価してもらった。このうち呼吸については、すべての試行において十分な改善がみられるという主観評価結果となったので、相関分析からは除外している。この表によると、発声の改善とスペクトルダイナミックレンジの相関が最も強い0.81を示している。倍音強度は発声、調音との間に0.7以上の相関が得られている。呼吸がすべての被験

者において改善しているということは、呼吸はトレーニング前後で有意差が顕著であった倍音構造強度、スペクトルダイナミックレンジ、ホルマント Q、スペクトル傾きのすべてに関係すると言える。

表 2 音響特徴と主観評価の相関.

音響特徴量	発声	調音	表情
倍音強度	0.74	0.76	0.45
スペクトルダイナミックレンジ	0.81	0.61	0.45
ホルマント Q	0.45	0.69	0.40
スペクトル傾き	-0.29	-0.28	0.01

6. むすび

ボイストレーニングによる声質の向上を定量的に評価するための計測ツールについて述べた。用いた音響特徴は、被験者ごとのトレーニング前後での特徴量の差に関する t 検定において有意差が見られた、基本周波数の倍音成分の強度、スペクトル包絡のダイナミックレンジ、ホルマントの Q、スペクトルの傾きである。これらの特徴量と声質の向上の主観評価との関連を分析した結果、呼吸は全音響特徴に関係し、発声はスペクトルダイナミックレンジと関係が深いことが分かった。

今後はさらに、短文や句単位での評価用音響特徴量の導入が必要である。

謝辞 本研究の一部はオープンリサーチセンタ事業の援助を受けた。

参考文献

[1] I. Yanushevskaya, C. Gobl, A. Ni Chasaide: Voice quality and f0 cues for affect expression: implications for synthesis, In INTERSPEECH-2005, 1849-1852, 2005.
[2] H. Yamada and N. Higuchi: Voice quality evaluation of IP-based voice stream multiplexing schemes, Proc. of IEEE Conference on Local Computer Networks 2001, 356-364, 2001-11.
[3] K. Honda, H. Hirai, J. Estill, and Y. Tohkura: Contributions of vocal tract shape to voice quality: MRI data and articulatory modeling: Vocal Fold Physiology —Voice quality control —, Chapter 2, 23-38, 1995-1.
[4] A. Maier, M. Schuster, A. Batliner, E. Nöth, E. Nkenke: Automatic scoring of the intelligibility in patients with cancer of the oral cavity, INTERSPEECH2007, 1206-1209,

2007.
[5] V.J. Boucher: Acoustic correlates of laryngeal-muscle fatigue: findings for a phonometric prevention of acquired voice pathologies, INTERSPEECH2007, 1202-1205, 2007.
[6] 吉岡 典子, 長幡 大介, 柳田 益造, 中山 一郎: 能・狂言と洋楽歌唱における母音の相違, 信学技報, 音声, 100(595), pp.1-8, 2001-01.
[7] P. Lal: A comparison of singing evaluation algorithms, Proc. INTERSPEECH 2006, 2298-2301, 2006.
[8] D. Hoppe, M. Sadakata and P. Desain: Development of real-time visual feedback assistance in singing training: A review, J. Comput. Assist. Learn., 22, 308-316, 2006.
[9] H. Kawahara, O. Fujimura and Y. Konparu: Voice quality of artistic expression in Noh: An analysis-synthesis study on source-related parameters, ASA and ASJ joint meeting, Hawaii, 1pMU1 2006.
[10] T. Saitou, M. Goto, M. Unoki, and M. Akagi: Vocal conversion from speaking voice to singing voice using STRAIGHT, INTERSPEECH2007, 4005-4006, 2007.
[11] 吉田有里, 森勢将雅, 高橋徹, 河原英紀: ポップス系歌唱音声の STRAIGHT による分析とスペクトル変動の統計的性質について, 信学技報, vol.107, no.282, pp.31-36, 2007-10.
[12] T. Nakano, M. Goto and Y. Hiraga: MiruSinger: A singing skill visualization interface using real-time feedback and music CD recordings as referential data, ISM2007 Workshops, 75-76, 2008.
[13] 齋藤毅, 後藤真孝: 歌唱指導による音響特徴の変化とその歌唱力評価への影響, 信学技報, 応用音響, EA2009-18, pp.1-6, 2009-06.
[14] G. Klasmeyer: The perceptual importance of selected voice quality parameters, ICASSP'97, vol. 3, 1615, 1997.
[15] X. Sun: Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio, ICASSP2002, I, 333-336, 2002.
[16] P. Prasertvithyakarn, K. Iwano, and S. Furui: An automatic singing voice evaluation method for voice training, 音学講論, 911-912, 2008-03.
[17] 相川清明, 秋竹朋子: ボイストレーニングによる声質向上の音響分析, 信学技報, EA2009-50, pp.43-48, 2009-08.
[18] 相川清明, 秋竹朋子: ボイストレーニングの効果を評価するための音声特徴量, 日本音響学会講演論文集, pp.329-330, 2009-09.
[19] 嵯峨山茂樹, 板倉文忠: 音声の動的尺度に含まれる個人性情報, 音学講論, pp.589-590, 1979-06.