音声中の個人性情報制御法に関する研究

課題番号：07680388

平成7年度～平成9年度科学研究費補助金（基盤研究(C)(2)）
研究成果報告書

平成10年3月

研究代表者　　赤木　正人
（北陸先端科学技術大学院大学情報科学研究科）

目　　次

１．はしがき

研究組織

研究代表者：赤木正人
（北陸先端科学技術大学院大学情報科学研究科助教授）
研究分担者：飯島泰蔵（平成7年度～平成8年度）
（北陸先端科学技術大学院大学情報科学研究科教授）
研究分担者：岩城　護
（北陸先端科学技術大学院大学情報科学研究科助手）

研究経費

平成7年度	1,200 千円
平成8年度	600 千円
平成9年度	400 千円
計	2,200 千円

研究発表

(1) 学会誌等

・スペクトル包絡に含まれる個人性

[1] Kitamura, T. and Akagi, M. (1995). "Speaker individualities in speech spectral envelopes", J. Acoust. Soc. Jpn. (E), 16, 5, 283-289.
[2] 北村、赤木(1997). ”単母音の話者識別に寄与するスペクトル包絡成分”、日本音響学会誌、53, 3, 185-191.

・基本周波数に含まれる個人性

[1] Akagi, M. and Ienaga, T. (1997). "Speaker individuality in fundamental frequency contours and its control", J. Acoust. Soc. Jpn. (E), 18, 2 73-80.

(2) 口頭発表

・スペクトル包絡に含まれる個人性

[1] 北村、赤木 (1994). ”音声のスペクトル包絡に含まれる個人性について”、電子情報通信学会技術報告、SP93-146
[2] 北村、赤木 (1994). ”スペクトル包絡における個人情報に関する検討”、平成６年春季音響学会講演論文、3-4-10
[3] Kitamura, T. and Akagi, M. (1994). "Speaker Individualities in speechspectral envelopes", Proc. Int. Conf. Spoken Lang. Process. 94, 1183-1186.
[4] 北村、赤木 (1994). ”スペクトル包絡に含まれる個人性を利用した話者変換”、平成６年秋季音響学会講演論文、1-9-17
[5] 北村、赤木 (1995). ”スペクトル高域成分の変形と話者識別”、平成７年春季音響学

会講演論文、3-9-20
[6] 北村、赤木(1995). ”個人性情報を含む周波数帯域について”、電子情報通信学会技術報告、SP95-37
[7] 北村、赤木 (1995). ”スペクトル包絡と個人性判断の関係”、平成７年秋季音響学会講演論文、3-3-10.
[8] 北村、赤木(1996). ”話者識別に寄与するスペクトル包絡の成分について”、電子情報通信学会技術報告、SP95-144.
[9] 北村、赤木(1996). ”話者識別に寄与するスペクトル包絡の成分について”、平成８年春季音響学会講演論文、2-3-6.
[10] 北村、赤木(1996). ”単純類似度法による話者識別に適した周波数帯域の検討”、平成８年秋季音響学会講演論文、1-6-17.
[11] Kitamura, T. and Akagi, M. (1996). "Relationship between physical characteristics and speaker individualities in speech spectral envelopes", Proc ASA-ASJ Joint Meeting, 833-838.
[12] 北村、赤木(1996). ”連続音声中の母音に含まれる個人性について”、音響学会聴覚研究会資料、H-96-98
[13] 北村、赤木(1997). ”連続音声中の母音の話者識別におけるスペクトル包絡と基本周波数の役割”、平成９年春季音響学会講演論文、2-8-6.

・基本周波数に含まれる個人性
[1] 家永、赤木 (1995). ”音声のピッチ周波数の時間変化パターンに含まれる個人性とその制御”、電子情報通信学会技術報告、SP94-104
[2] Akagi, M. and Ienaga, T. (1995). "Speaker individualities in fundamental frequency contours and its control", Proc. EUROSPEECH95, 439-442.
[3] 大野、赤木(1998). “文音声中の基本周波数パターンに含まれる個人性の検討”、電子情報通信学会技術報告、SP
[4] 皆川、赤木(1998). “連続発話母音における基本周波数の変動とその知覚”、電子情報通信学会技術報告、SP

・異常構音
[1] 高木、北村、赤木、鈴木、藤田、道(1996). ”側音化構音の音響特性について”、平成８年春季音響学会講演論文、1-3-3.
[2] 赤木正人、高木直子、北村達也、鈴木規子、藤田幸弘、道健一(1996). ”側音化構音の知覚と物理関連量”、電子情報通信学会技術報告、SP96-34.
[3] Akagi, M., Kitamura, T., Suzuki, N. and Michi, K. (1996). "Perception of lateral misarticulation and its physical correlates", Proc ASA-ASJ Joint Meeting, 933-936.

２．概要

2.1　研究の位置づけ

音声に個人性が含まれるということは周知の事実であるが、ではどのような物理量が音声の個人性を決定する要因になっているのかという問に対する答えは未だ完全ではない。音声は、声道情報を反映したスペクトルの包絡特性および声帯情報を反映したピッチ特性の二つの物理量で大まかには記述されると言われているが、個々の物理量に潜む個人性関連量についての細かい議論はあまりなされていないのが現状である。また、個人性知覚に関連する物理量を積極的に制御し、音声認識・合成に応用する研究はほとんど行われていない。

そこで本研究では、人間の音声に関する知覚特性を考慮して、個人性の知覚と物理関連量およびこれらの制御法について検討を行った。具体的には、

(1) 音声における人間の個人性知覚特性から物理関連量を推定することを目的として、人間が個人差を知覚する時に用いる物理量が最も個人性を表す物理量であるという仮定の下に、個人差の知覚と様々な物理量の関係を心理物理実験を通して明らかにする。

(2) 音声認識・合成への応用を目的として、個々の物理量の制御可能性を考察する。

(3) 個人性知覚の特徴の一つであると考えられる音声の音色についての研究に関連して、歪みを持った音声（異常構音）と正常音声の比較から、歪み音の知覚とこれを生起させる物理関連量について調査する。

ことを目的として、研究を遂行した。

個人性に関連する個々の物理量を抽出しこれを制御する方法を見つけ出すことは、音声の個人性を議論する基礎的な研究に貢献するのみでなく、応用面としての音声認識・合成技術において非常に重要な要素となる。たとえば、個人性制御モデルが構築されれば、個人性を音韻性を損なわない範囲で除去することが可能となり多数話者音声認識の認識率向上が期待できる。また、合成音声に個人性を付与できることとなり多様な合成音声が得られることとなる。さらに、個人性情報を抽出できるようになり話者認識・照合のための特徴量として使用できるようになる。

2.2　成果の概要

2.2.1　母音スペクトル包絡に含まれる個人性

母音スペクトル包絡のどの部分にどのように個人性が含まれているか。結論を先に述べれば、個人性は2 kHz付近に存在するピーク以上の帯域に主に含まれ、この帯域のピーク位置とピークの高さが個人性に関連する。この結論は、次のような工学的および心理物理的手法によって明らかになった。

単独発話母音のどのスペクトル帯域に個人差が主に含まれているかを調べるために、平均スペクトルからの分散を計算した。結果を図１に示す。図の上段は各母音（５母音）ごとに話者間で平均をとり、この平均からの分散を表示してある。また、下段は話者ごとに母音間で平均をとり、この平均からの分散を示してある。図から、22 ERB rate （約2.2 kHz）を境として低域は母音の分散が大きく、高域は話者間の分散が大きいことがわかる。この結果から、スペクトルの高域に個人差が含まれていると推測される。なお、ERB rateは人間の聴覚特性を考慮した周波数軸である。

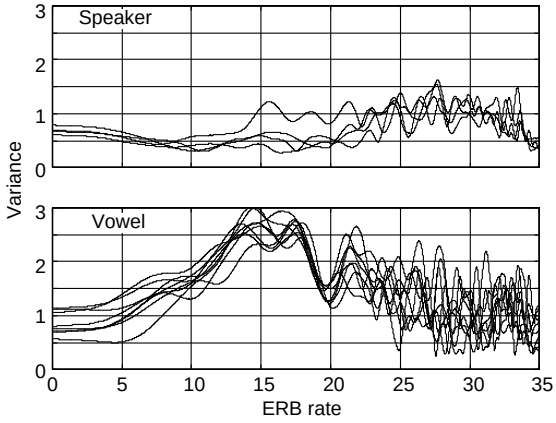

図１　各母音における話者間の分散（上段）と各話者における母音間の分散

22 ERB rate以上の帯域が本当に個人性に関連するかどうかを調べるために、図２に示すように、原形のスペクトル(A)に対して22 ERB rate以上の帯域を回帰直線で反転させたスペクトル(C)、回帰直線で置き換えたスペクトル(D)を用意し、これをもとに基本周波数を同じにして合成した母音を被験者に聞かせて個人性判断を行わせた。結果を図３に示す。図３中の(B)は、12 - 22 ERB rate帯域を回帰直線で反転させた母音である。図からわかるように、高域の変形は音韻性には影響を与えないが、個人性判断には影響を与える。また、12 - 22 ERB rate帯域は音韻性を担う帯域である。これらの結果は、被験者が個人性を判断する場合に用いている特徴は高域に存在することを示している。

次に、スペクトルの高域に存在するどのような特徴が個人性判断に関係するかを調べるために、高域に存在するスペクトルのピークとディップを図４のように変形し、これから合成した母音を被験者に呈示した。図５は、刺激ORGに対しての話者識別率からの誤り増加量である。ピークのみ残したスペクトルでは誤りはあまり増えないが、回帰直線で置き換えたスペクトルでは、誤りは増えている。誤りはORG < PEAK < DIP < REGとなり、この結果から、母音によって個人を特定するためには、スペクトルピークの位置と大きさが重

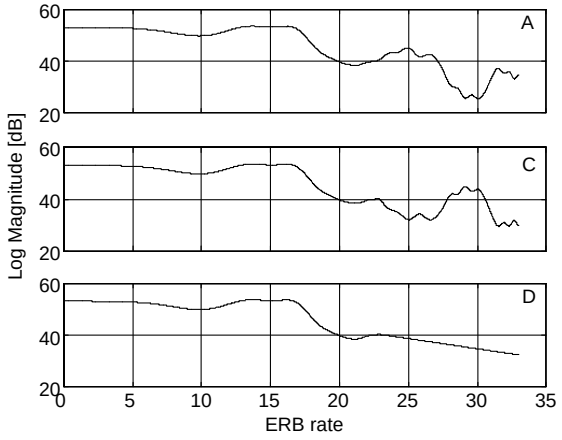

図２　スペクトル包絡を変形させた母音のスペクトル包絡。(A) 変形なし、(C) 回帰直線に対して反転、(D) 回帰直線で置換。

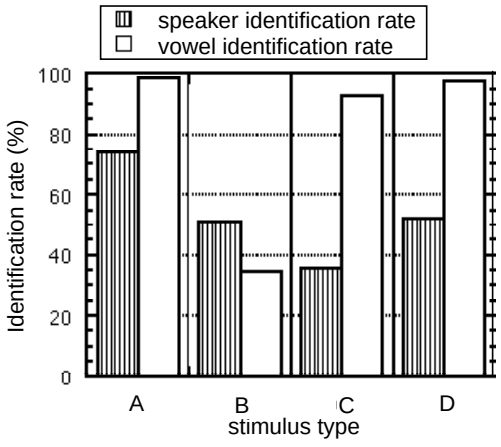

図３　話者識別率と音韻識別率

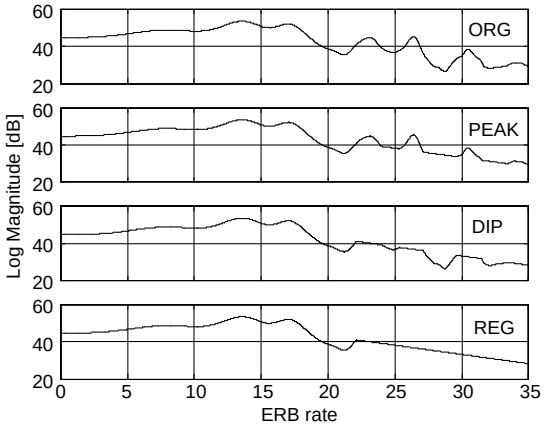

図４　刺激音ORG（原スペクトル）、PEAK（回帰直線より上のピークを保存）、DIP（回帰直線より下のディップを保存）、REG（回帰直線で置換）のスペクトル包絡。

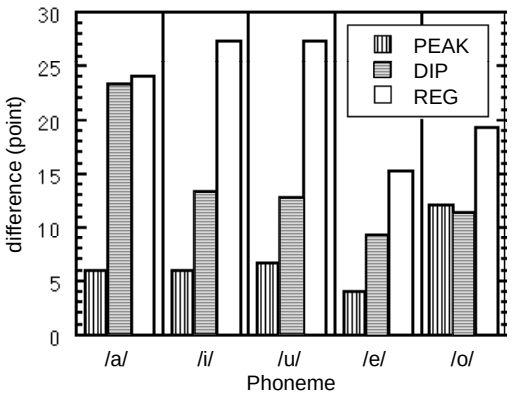

図５　刺激音PEAK, DIP, REGの誤り増加率。

要であることがわかる。これらの結果は、連続音声中の母音においても同様である。

さらに、スペクトル包絡の全周波数帯域を使うのではなく、スペクトル包絡の高域を用いて話者認識を行えば、高い弁別性能が得られることも明らかになった。

これらの結果から、

(1) 個人性はスペクトル包絡全体に現れるが、高域により多く現れる。

(2) 話者識別にはスペクトル包絡のディップよりもピークが重要な意味を持っている。

(3) 個人性はスペクトル包絡の 20 ERB rate (1740 Hz) 付近のピーク以上の帯域に顕著に現れ、この帯域を利用して声質変換が可能である。

(4) スペクトル包絡における個人性は基本周波数における個人性よりも話者識別に寄与する。

(5) 本研究により個人性を表すことが明らかになったスペクトル包絡の高域を音声合成に応用することが可能である。

(6) この帯域における個人性を利用して話者正規化や話者適応を行う技術を開発することにより、不特定話者音声認識の性能向上が期待できる。

ことが明らかとなった。

2.2.2　基本周波数に含まれる個人性

基本周波数パターンに含まれる個人性について、多数話者が発話した単語および文章の分析（工学的手法）と基本周波数パターンを変形して合成した単語および文章の聴取実験（心理物理学的手法）により検討を行った。結論を先に述べれば、単語においては基本周波数包絡変化の大きさに関係するパラメータが個人性に主に関係しており、このパラメータを変形することにより個人性が制御できることが明らかになった。また、文章においては、基本周波数包絡変化の大きさに関係するパラメータだけでなく、アクセントなどのタイミングに関係するパラメータも個人性に関係することが明らかとなった。

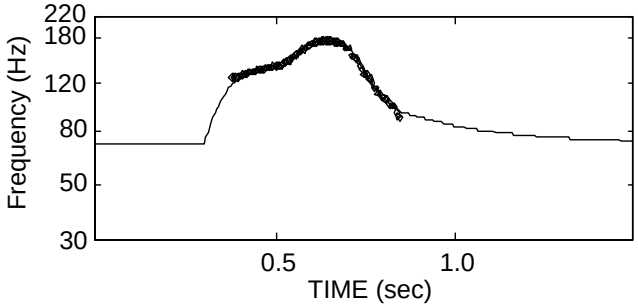

図６　抽出された基本周波数包絡（太線）と藤崎モデルによる
近似結果（細線）。音声は単語「青い」。

基本周波数パターンは、まず図６に示すように抽出され、そして藤崎モデルによってパラメータ化される。このパラメータは、パターンの変化の大きさを表すものとアクセントなどのタイミングを表すものに大別される。それぞれのパラメータの話者による分散をＦ比によって調べた結果、単語、文章双方ともパターンの変化の大きさを表すパラメータの分散が大きくなっていた。これは、パターンの変化の大きさを表すパラメータが個人性を多く含むことを示唆するものである。

この仮説を証明するために、単語音声においては、図７に示すように、タイミングを表すパラメータ $\Delta T_i \ (i = 0,1,2,3)$はそのまま用いて、基本周波数パターンの変化の大きさを表すパラメータ F_b, A_p, A_aを他の話者と入れ換えた音声を合成する。なお、スペクトル包絡は話者間で平均してある。これを被験者に呈示し個人性判断を行わせた。結果は図８の通りである。変形しない音声(A)に対して話者識別率は約90%、変形した音声(B)では約90%基本周波数パターンの変化の大きさを表すパラメータを用いた話者として判断している。この結果は、パターンの変化の大きさを表すパラメータが基本周波数パターンに含まれる個人性を表す重要な物理量であることを示している。

また文章においては、タイミングを表すパラメータおよび基本周波数パターンの変化の大きさを表す3パラメータの計4パラメータを組み合わせて、話者間で入れ換え文章音声を合成する。このとき、スペクトル包絡は基本周波数パターンを用いない第三者のものを用いた。これらの合成音声を被験者に呈示し比較判断を行わせた。この結果から、個人性は変化パターンのダイナミックスにも含まれるが、それ以上に時間情報（アクセントのタイミングなど）に含まれることが明らかとなった。

これらの結果を用いれば、音声合成において基本周波数パターンを用いた話者変換が可能となる。

さらに、声帯振動周期の揺れをelectroglotograph (EGG)を用いて系統的に調べたところ、話者によって比較的遅い変化（約10 Hz）を持つ者、比較的速い変化（30-60 Hz）を持つ者、変化があまりない者の３グループに分けられることがわかった。また、聴取実験の結果、これらをお互いに知覚分離できることが明らかとなった。これより、声帯振動の揺れも個人性と関連する一つの物理量であることが言える。

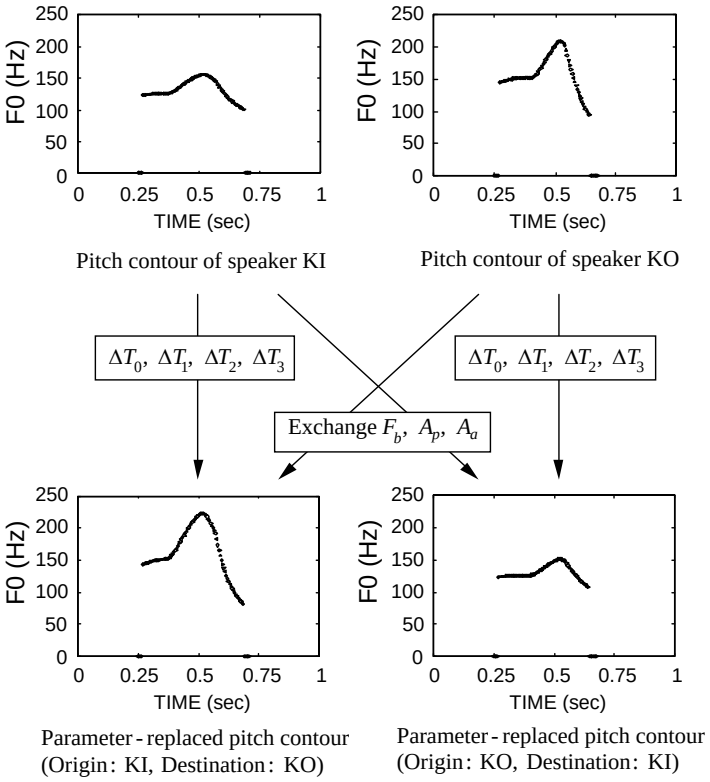

図７　パラメータ置換の概略図。音声は単語「青い」。

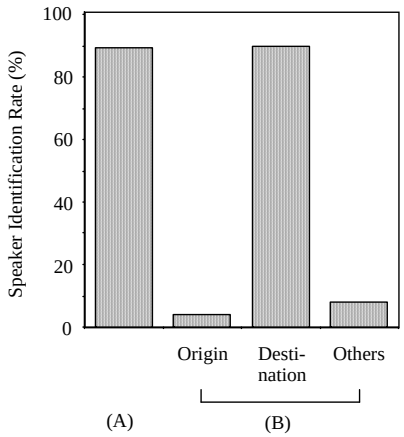

図８　原音声(A)と基本周波数変形音声(B)の話者認識率。
Origin（タイミングを表すパラメータの話者と認識した割合）、
Destination（包絡の変化の大きさを表すパラメータの話者と認識した割合）。

2.2.3　異常構音の音色

言語障害による異常構音についても、音声の個人性の一分野である。本研究では異常構音の一種である側音化構音を扱っている。

側音化構音とは、舌、顎などに形態的障害がないにもかかわらず、子音/sh/, /ch/などを発話する場合、舌を口蓋中央に接触させるために、呼気が臼後部より口腔前庭の側方より出ることにより音が歪むものである。正常構音の場合は呼気は口腔正中より出るが、側音化

構音の場合は口腔の側方より出るために独特の歪み音を呈する。構音時に口唇が側方へ変移することもある。舌、顎などの運動機能が未熟な小学校低学年以下の子供でよく見受けられる。

診断は、ステンレス板による呼気の流出部位検査、エレクトロ・パラトグラフィーによる舌の接触様式の測定、舌造影側方頭部X線規格写真法（X線セファログラフ）を用いて客観的に行なわれなければならないが、膨大な時間がかかるため、現在は言語臨床家による臨床診断に頼っているのが現状である。そこで、言語臨床家は異常構音のどの特徴をとらえて診断を下しているのかを、工学的手法および心理物理学的手法を用いて特定し、側音化構音診断の自動化を試みる。

側音化構音の典型的な分析例（子音/sh/）を図１０に示す。側音化構音のスペクトル包絡においては、5 kHz以上の帯域におけるパワーが少なく、3kHz付近に存在するスペクトルのピーク位置と大きさがほぼ周期的に変化していることが確認された。一方、正常音ではこのピークは見られない。

そこで、これらの特徴が言語臨床家の聴覚判断に影響を与えているかどうかを調べるために、側音化構音のスペクトルに典型的なピークが存在する25 ERB rate付近の5 ERBの帯域と正常話者の同じ位置の帯域をバンドパスフィルタとノッチフィルタを用いて入れ換え（図１１）、言語臨床家に呈示した。結果は、正常話者のスペクトルの25 ERB rate付近の5 ERBの帯域を入れ換えただけで、側音化構音の聴覚印象が著しく増大した。これは、側音化構音の特徴が25 ERB rate付近のピークとその時間変化にあることを示唆するものである。

この音響的特徴が声道形状のどの部分に起因しているのかを調べるために、X線セファロトレースグラフから声道形状を抽出し、これを用いて声道の伝達特性を推定した。また、正常者の声道形状を計算機上で変化させることにより、側音化構音と同様の音響特性の実現を試みた。その結果、側音化構音に特徴的なスペクトル包絡は、声道のせばめの長さとその位置に関係することが明らかとなった（図１２）。

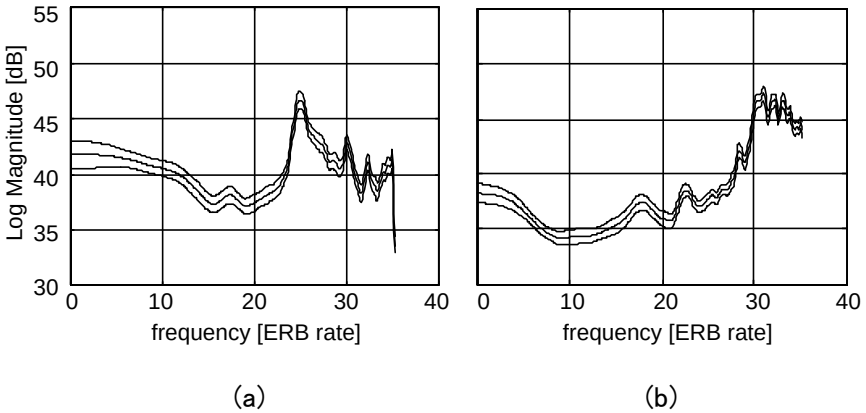

図１０　側音化構音音声のスペクトル(a)と正常音声のスペクトル(b)。

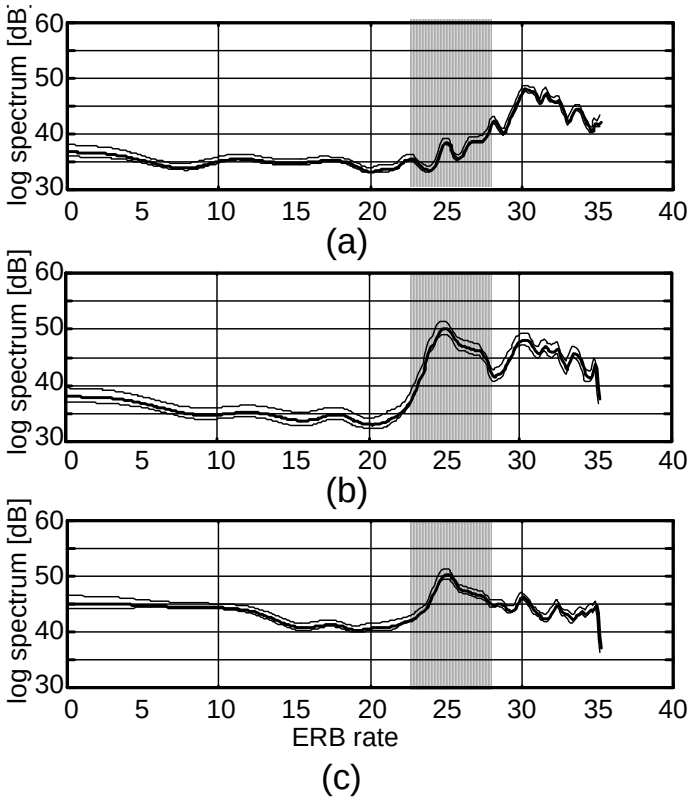

図１１　呈示音声のスペクトル。(a) 正常音声、(b) 正常音声の25 ERB rate付近の5 ERBの帯域を側音化構音と置き換えた音声、(c) 側音化構音音声。

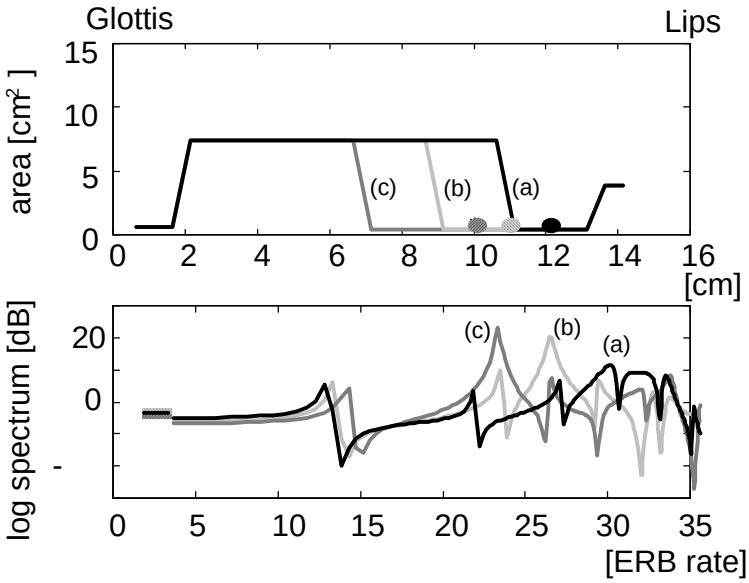

図１２　声道における狭めの長さおよび位置と声道伝達関数の関係、(a) 正常。(b), (c)と狭めが長くなるに従い、ピーク周波数は下降し、スペクトル包絡形状は側音化構音音声に近くなる。

３．研究成果