

Title	異常構音の音色知覚と音響特性に関する研究
Author(s)	林, 勝己
Citation	
Issue Date	1997-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1009
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

修 士 論 文

側音化構音の音色知覚と音響特性に関する研究

指導教官 赤木 正人 助教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻
音情報処理学講座

510100 林 勝己

1997 年 2 月 14 日

目 次

1	緒論	1
1.1	はじめに—側音化構音とは—	1
1.2	側音化構音に関する研究動向	1
1.3	本研究の意義と目的	4
2	正常構音と側音化構音の比較分析	6
2.1	目的	6
2.2	分析手法	6
2.2.1	音声データ	6
2.2.2	音声の録音	7
2.2.3	ケプストラム分析	7
2.3	分析結果	8
2.3.1	単一話者での比較分析結果	16
2.4	考察	16
2.4.1	静的音響特性	17
2.4.2	動的音響特性	17
3	スペクトル包絡変形音の聴取実験	19
3.1	実験の目的	19
3.2	実験方法	19
3.2.1	スペクトル包絡変形音の作成方法	19
3.2.2	使用した原音声	20
3.2.3	スペクトル変形パターン	20

3.2.4	対応する特徴量	22
3.2.5	聴覚印象の評価方法	22
3.3	実験結果	24
3.4	考察	24
3.4.1	総合結果からの考察	28
3.4.2	各話者毎の傾向	31
4	声道モデルによる側音化構音のモデル化	33
4.1	目的	33
4.2	モデルの概要	34
4.2.1	音声の生成過程	34
4.2.2	音響管の等価電気回路	34
4.2.3	駆動音源	36
4.3	計算機での実装	36
4.4	シミュレーション (1)	37
4.4.1	結果と考察	37
4.5	シミュレーション (2)	41
4.5.1	声道形状の測定	41
4.5.2	シミュレーション結果と考察	42
4.5.3	狭めの長さと共振周波数の関係	45
5	結論	46
5.1	本研究の成果	46
5.2	今後の課題	46
6	謝辞	48
A		51

目 次

1.1	発声器官の概略図	2
1.2	EPG のシステム	2
1.3	側音化構音に対するアプローチ	4
2.1	話者 ur の分析結果 (側音化構音)	10
2.2	話者 fj の分析結果 (側音化構音)	11
2.3	話者 hr の分析結果 (側音化構音)	12
2.4	話者 FT の分析結果 (正常構音)	13
2.5	話者 SN の分析結果 (正常構音)	14
2.6	話者 in の平均スペクトル (細線:正常構音時, 太線:側音化構音時)	16
3.1	変形したスペクトル包絡	23
3.2	話者 ur の得点結果	25
3.3	話者 fj の得点結果	25
3.4	話者 hr の得点結果	26
3.5	話者 fk の得点結果	26
3.6	4 話者の平均得点結果	27
3.7	話者 IN(in) の得点結果	27
3.8	話者 fk,TN(Normal) の平均スペクトルと標準偏差	31
4.1	音響管とその等価回路	34
4.2	放射インピーダンスの構成素子	35
4.3	母音/a/の声道断面積伝達関数	38
4.4	母音/a/の声道の周波数特性	38
4.5	母音/o/の声道断面積伝達関数	39

4.6	母音/o/の声道の周波数特性	39
4.7	母音/i/の声道断面積伝達関数	40
4.8	母音/i/の声道の周波数特性	40
4.9	3 セクションの音響管モデル	42
4.10	正常構音話者の声道の周波数特性	43
4.11	話者 ur の声道の周波数特性	43
4.12	話者 fj の声道の周波数特性	43
4.13	話者 hr の声道の周波数特性	44
4.14	話者 fk の声道の周波数特性	44
4.15	狭めの長さと共振周波数	45

表 目 次

2.1	録音条件	7
2.2	正常な /f/ の周波数特性 [ERB rate]	15
2.3	側音化した /f/ の周波数特性 [ERB rate]	15
3.1	スペクトル変形パターン	21
3.2	各変形音声の特徴	22
3.3	聴覚印象の点数化方法	24
3.4	音響特性と対応した平均得点 (4 話者)	28
4.1	母音 /a/ の声道共振周波数と Formant 周波数	38
4.2	母音 /o/ の声道共振周波数と Formant 周波数	39
4.3	母音 /i/ の声道共振周波数と Formant 周波数	40
4.4	各話者の声道パラメータ	41

第 1 章

緒論

1.1 はじめに—側音化構音とは—

側音化構音とは、口腔形成及び機能に明らかな異常の見られない機能性構音障害のひとつであり、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。但し、口蓋裂などの原因により構音時に側音化する場合は、口蓋化構音と呼ばれるが、これも側音化構音の一種である

図 1.1に発声器官の概略図 [1] を示す。

正常構音の場合、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から出ることにより独特の歪音を呈する。構音時に口唇が側方に偏位することもある。

側音化構音は軽度の障害と考えられているが、/ʃ i/ (し) と発音しても /hi/ (ひ) と聞きとられ、日常生活で支障を生じていることが少なくないのが現状である。

1.2 側音化構音に関する研究動向

異常構音話者の発音に関する研究は、医学方面で盛んに行われている。

まず、構音形態を客観的に測定する手法の例を、以下に示す。

EPG[5] electropalatograph の略。電極を口蓋にあて、構音時の口蓋と舌の接触状態を調べる。図 1.2に EPG の人工口蓋と出力結果の例を示す。

呼気流出検査 スレンレス板を用いて、構音時に呼気がどの部位から流出しているかを検



図 1.1: 発声器官の概略図

査する.

Cephalogram 舌造影側方頭部 X 線規格写真法. 構音時の声道形状を撮影する.

MRI[2] 磁気共鳴映像法 (Magnetic Resonance Imaging). 異常構音での測定報告はないが, X 線と比較して安全性が高く, 任意斜断像の撮影が容易である. 近年, 音声生成過程の観測や声道形状の計測に応用されている.

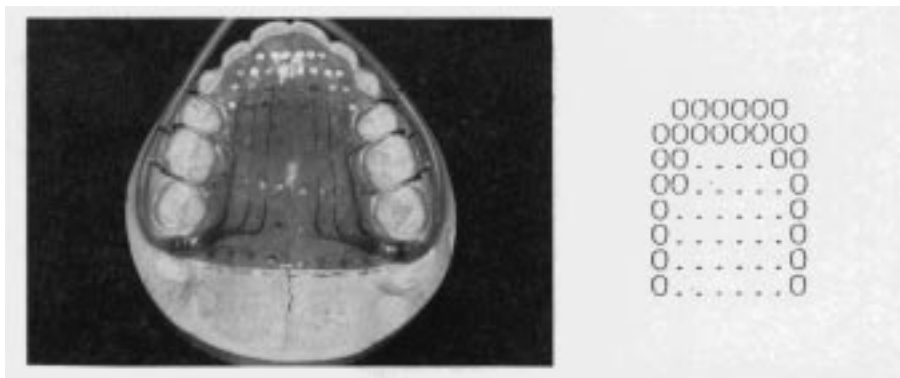


図 1.2: EPG のシステム

これらの手法を用いて, 構音時の舌と口蓋の接触の位置などの形態が明らかにされ, その構音方法により, 独特な音色が生成されていることがわかった [3].

ところで、一般的な母音や子音の音声生成過程では様々な計測法が提案されている [4] のに対し、異常構音の場合、厳密な音源（雑音源）の位置や、発声される音声の音響特性との対応関係など解明されていない点も多い。また、これらの測定は時間的、金銭的成本が非常に高く汎用的ではないという欠点も抱えている。

そうした現状から、側音化構音の診断は、主に、言語治療家の聴覚印象による構音検査や、発話明瞭度検査といった音声言語による聴覚判定に依存している。

一方で、工学的な立場から異常構音の音響特性を調査した報告例を以下に示す。

Suzuki & Wakumoto[5] は日本語とイギリス英語の /s/ の側音化構音を比較し、その結果、日本人とイギリス人の構音はよく似ているということがわかった。また、両言語の構音は、正常構音よりも低域にエネルギーのピークがあるという特徴を示した。

また、高橋 [6] は、側音化構音の /f i/ , /t f i/ についてソナグラフを用いた分析を行なった。その結果、正常構音では約 4kHz 以上の帯域に周波数成分が大きく存在するのに対し、側音化構音ではスペクトルのパワーレベルは周波数軸上でほぼ一様であることが分かった。

異常構音の音響特性と音色との関係を合成音を用いた実験により明らかにした例もある。

荒井と岡崎 [7] は、線形予測分析によって /s/ の口蓋化構音の音響的特徴を確認した後、単音節 /sa/ , /su/ , /se/ の合成音を作成し、その聴取実験を行なうことにより、口蓋化構音の特徴を特定した。その結果、極の周波数が 1800Hz 前後に配置されているとき、口蓋化構音と聴取される傾向が特に強いことが分かった。このことは、言語治療家はスペクトル包絡のピークを聞いて口蓋化構音を判断していることを示している。

また、高木ら [8] の研究では、側音化構音のスペクトル包絡には、25 ERB rate(3.14 kHz) 付近にスペクトルピークがあり、そのピークの存在と時間変動が、聴覚印象に影響を与えていることが示唆されている。

しかし、いずれの研究も、対象となる異常構音話者数が少なく、患者によって音響特性の傾向が異なるため、発音障害の程度や構音方法との関係を定量的に評価することができていない。

1.3 本研究の意義と目的

異常構音の一種である側音化構音や口蓋化構音に伴う歪音の診断は、主として言語治療家の聴覚的な判断に頼っているのが現状であり、客観的に診断できる支援システムの開発が強く望まれている。

そのためには、医学分野で別々に行なわれている [聴覚印象による診断 (評価)] と [構音形態の客観評価] とを [音響特性] を介して結びつけ、異常構音の定量的評価方法を見出す必要がある。図 1.3 に示す対応関係について、具体的に以下の 2 点について解明しなければならない。

- 異常構音の音色に対応する特徴量が、音響特性のどの部分に含まれているか。
- 異常化構音の構音形態と音響特性の関係。

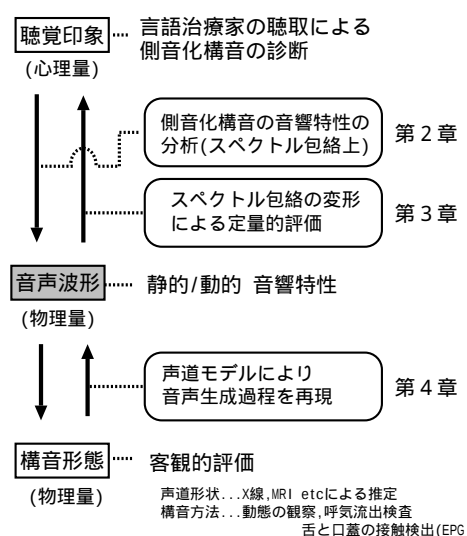


図 1.3: 側音化構音に対するアプローチ

ここでいう音色とは、音の印象を総称したもので、識別的側面や感覚的側面を持っている。識別的側面を持つ音色として、音声の持つ個性や異常構音による歪音などがあるが、その中で異常構音による歪音を取り扱った。

本研究は、異常構音の特徴的な音色が音響特性のどの部分に含まれるかを明らかにするとともに、歪音の生成過程をモデル化することによって、異常構音の定量的評価方法を開発することを目的としている。

研究の手法として、以下の3点に着目し、それぞれ対応する章で評価、検討をおこなっている。図 1.3 中で示す章も同じである。

- 分析する話者数を増やして、正常話者との比較分析を行った。ここで、特に音響特性の時間的变化に着目して分析を行った (第 2 章)。
- 音響特性と聴覚印象との関係性を評価する手法として、スペクトル変形による定量評価を行った (第 3 章)。
- 音声生成過程をモデル化し、構音方法によって音響特性がどのように変化するかシミュレーションを行った (第 4 章)。

第 2 章

正常構音と側音化構音の比較分析

2.1 目的

現在, 側音化構音に特徴的な音響特性 (スペクトル包絡) として I: 25 ERB rate(3.14 kHz) 付近に現れるピーク, II: 25 ERB rate 付近に現れるピークの周期的ゆらぎの発生, III: 包絡の傾斜の違いなどが報告されている [6] [8] .

本研究では, 側音化構音の知覚と音響特性の関係を調べるために, スペクトル包絡上で側音化構音における歪音の物理的な構造を詳しく分析した. 以下に分析の方法と結果について述べる.

2.2 分析手法

2.2.1 音声データ

側音化構音の歪音が現れやすい連続持続音 / j / について周波数領域での音声分析を行った. 音声データとして, 昭和大学歯学部で異常構音と診断された患者 9 名による発話と, 正常話者 10 名による発話を使用した.

今回用いている側音化構音の話者のうち, 4 名 (ur, fj, hr, fk¹) の発話については, 言語臨床家による聴覚的判断に加えて EPG, 呼気流出部位検査, Cephalogram(図 1.2 参照) により構音方法が客観的に診断されている.

¹話者 fk の場合は, 厳密には口蓋裂による口蓋化構音に属する

2.2.2 音声の録音

側音化構音を有すると診断された話者 9 名の音声は，昭和大学歯学部で DAT に録音されたものを使用した．ただし，録音の際に空調による暗騒音が音声に付加されており，特に低周波領域でのパワーが変動する現象が起きている．

正常と認められる話者 10 名の音声は，本学の防音室で持続音 / f / を 3 回約 5s 発話し，DAT に録音した．

録音した音声は 48kHz サンプリングの DAT から DSP を用いて 20kHz ダウンサンプリングし，WS に保存している．

正常者の録音条件を表 2.1 に示す．

表 2.1: 録音条件

話者	女性 10 名
マイクロホン	SONY C-536P マイクロホンから約 15cm 離れたところにて発声
DAT レコーダ	SONY TCD-D10 PRO II
標本化周波数	録音時:48kHz , WS 保存時:20kHz
騒音レベル	22.7 dB(A) (リオン NL-04 で測定)

2.2.3 ケプストラム分析

パワーレベルを正規化した連続音 / f / の分析手法として，サンプリング周波数 20 kHz，フレーム長 25.6 ms，フレーム周期 6.4 ms で，改良ケプストラム法 [9] (加速係数 1.0，近似回数 3，分析次数 60 次) による分析を行い，周波数特性 (スペクトル包絡) を求めた．

音声信号 $x(n)$ を音源 $g(n)$ と共振系の特性 (具体的には声道のインパルス応答 $v(n)$) に分離して抽出する方法は，deconvolution と呼ばれ，その方法の一つがケプストラム (cepstrum) 分析である．その処理の概略を以下に示す．

$x(n)$ は $g(n)$ と $v(n)$ の畳み込みであるから，

$$x(n) = g(n) * v(n) \quad ([*] \text{ means convolution.}) \quad (2.1)$$

と表すことが出来る．これをフーリエ変換すると

$$X(k) = G(k) \cdot V(k) \quad (2.2)$$

この振幅対数を取れば,

$$\log |X(k)| = \log |G(k)| + \log |V(k)| \quad (2.3)$$

となる．これを逆フーリエ変換すると

$$\hat{x}(n) = \hat{g}(n) + \hat{v}(n) \quad (2.4)$$

となり, $\hat{x}(n)$ は $\hat{g}(n)$ と $\hat{v}(n)$ の和で表せる． $\hat{x}(n)$ は周波数領域の IDFT であるから時間の次元を持つが, 途中で振幅対数を取っているため時間そのものではないという意味で, この逆変換次元のことをケフレンシ (quefrensy) といい, 変換された結果をケプストラム (cepstrum) という．

ここで, 音源の線スペクトルとスペクトル包絡をそれぞれ波形とみなしたとき, 前者は周波数が高く, 後者は低周波成分しか持たない．したがって, 逆変換された次元で, 低域成分は包絡に, 高域成分は音源パルス列に対応する．つまり, $\hat{g}(n)$ は $1/F_0$ (ピッチ周期) の近傍から高ケフレンシ領域の成分として, $\hat{v}(n)$ は低ケフレンシ領域の成分として表れるから, ケプストラムの $1/F_0$ の原点側を分離抽出するフィルタ (リフタ) をかける (リフタリングする) ことにより, インパルス応答成分を抽出できる．これを再び DFT することによってスペクトル包絡を求めることが出来る．

また, $1/F_0$ の近傍でのピーク値からピッチ周期が, 更にその成分の逆変換によって, 音源ピッチパルス列を再生することも可能である．

以上の方法は有声音では有効である．そこで摩擦音などの子音の場合にも適応したのが, 改良ケプストラム法である．詳しくは文献 [9] を参考にされたい．

2.3 分析結果

前述の分析手法により, 正常構音と側音化構音の $/f/$ のスペクトル包絡を比較した．

側音化構音の $/f/$ の分析結果を図 2.1 ~ 図 2.3 に, 正常構音の $/f/$ の分析結果を図 2.4 , 2.5 に示す．これらは, 側音化構音と正常構音の典型的なスペクトル包絡であり, ここに載

せなかった話者の分析結果は、いずれかの話者と似た傾向となった。

図表中で用いた周波数の単位として、人間の聴覚特性を考慮し [ERB rate] を使用した [10]。[Hz] から [ERB rate] への変換は式 2.5に示す非線形変換によって行なわれる。

$$[\text{ERB rate}] = 21.4 \log_{10}(4.37[\text{kHz}] + 1) \quad (2.5)$$

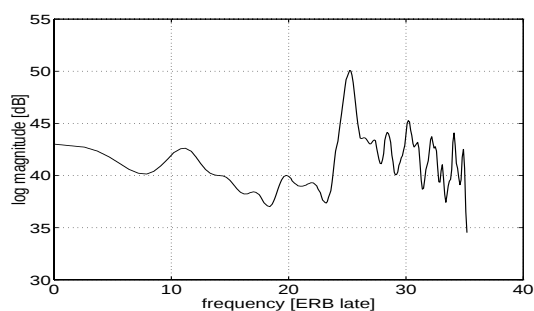
各図中の (a) ~ (d) は、定常部分の連続 4 フレーム分のスペクトル包絡であり、(e) は、それらのスペクトル包絡を 30 フレーム分重ねて表示したものである。

また、(f) には、各スペクトル系列の 150 フレーム分の時間平均スペクトルとパワーの標準偏差を示した。中央の太線が平均値であり、上下の線がその周波数での標準偏差を加減した値を示している。

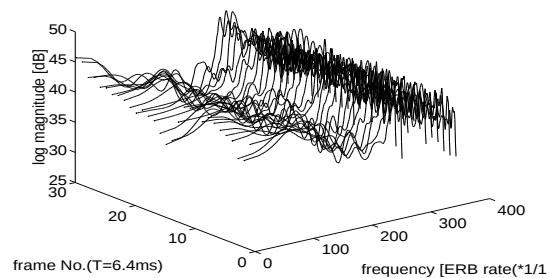
(g),(h) は、各スペクトル包絡の最大ピークと 25ERB rate 付近のピークについて、周波数とパワーレベルを示す。実線が 25ERB rate 付近のピークでの値であり、×点は、最大ピークでの値である。各表中に示した 4 つの値は、それらの平均値と分散値である。

また、表 2.2,2.3に全話者による /f/ のカットオフ周波数 (スペクトル包絡の盛り上がりはじめる周波数と定義) と最大ピーク周波数を示す。

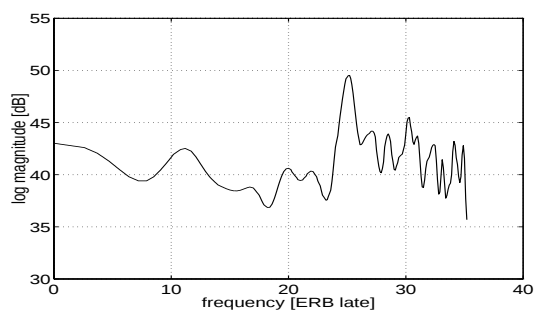
(注):話者の識別の際に、大文字で表しているのは正常構音話者であり、小文字は側音化構音話者である。



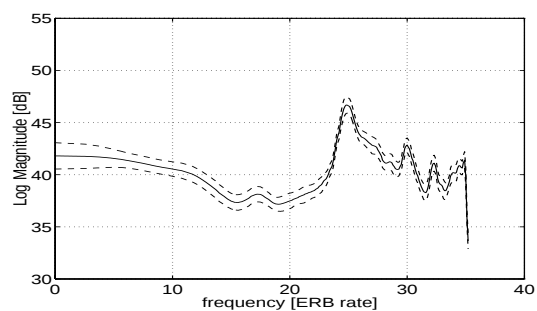
(a) フレーム 1



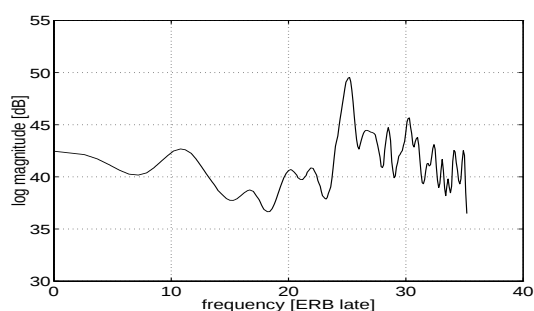
(e) 連続 30 フレームの系列



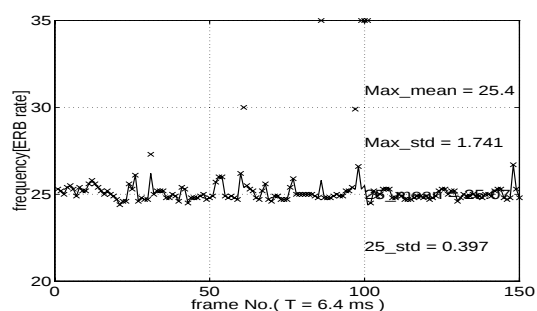
(b) フレーム 2



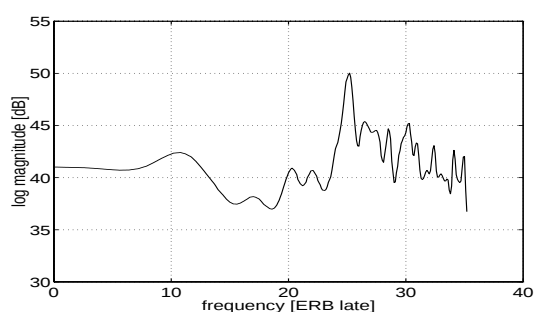
(f) 平均スペクトルと標準偏差



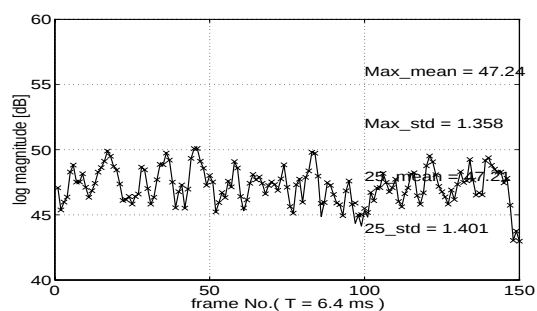
(c) フレーム 3



(g) 各フレームのピーク周波数

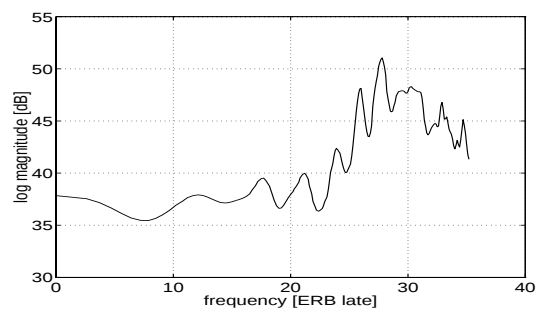


(d) フレーム 4

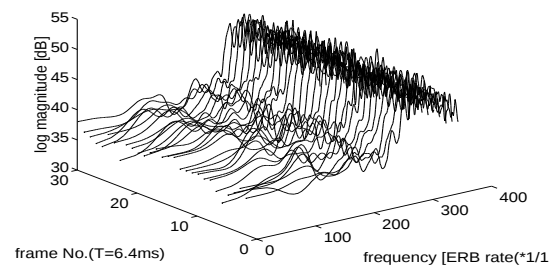


(h) ピーク周波数における
パワーレベル

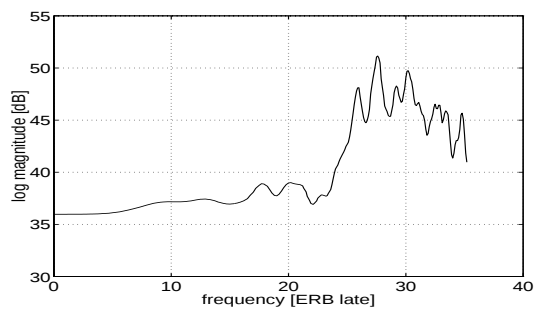
図 2.1: 話者 ur の分析結果 (側音化構音)



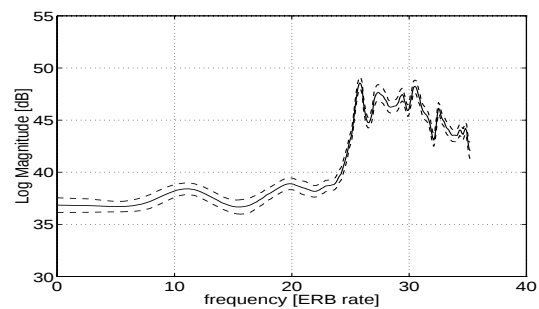
(a) フレーム 1



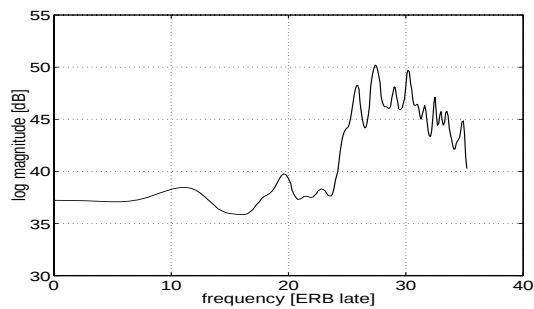
(e) 連続 30 フレームの系列



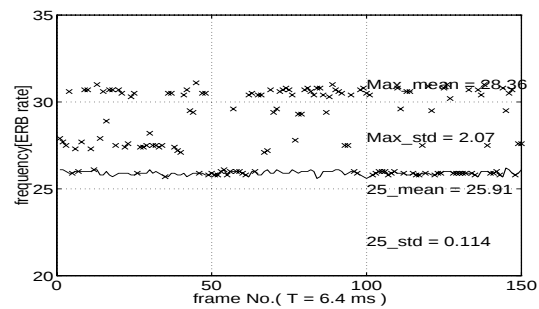
(b) フレーム 2



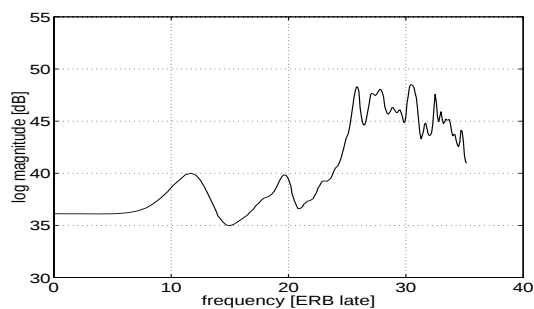
(f) 平均スペクトルと標準偏差



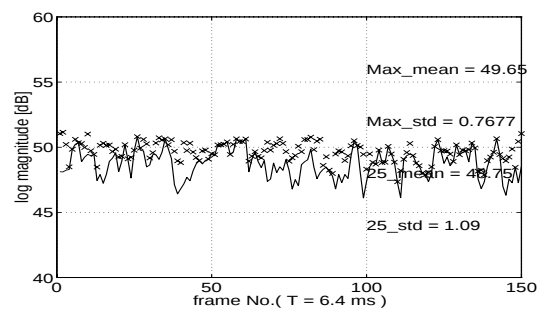
(c) フレーム 3



(g) 各フレームのピーク周波数

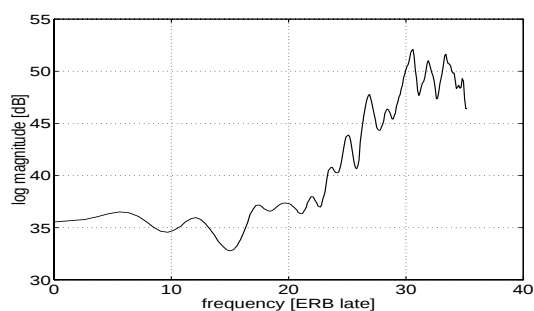


(d) フレーム 4

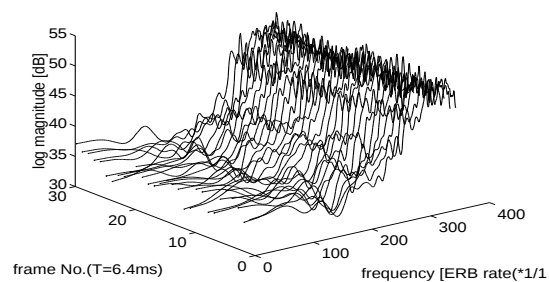


(h) (g) におけるパワーレベル

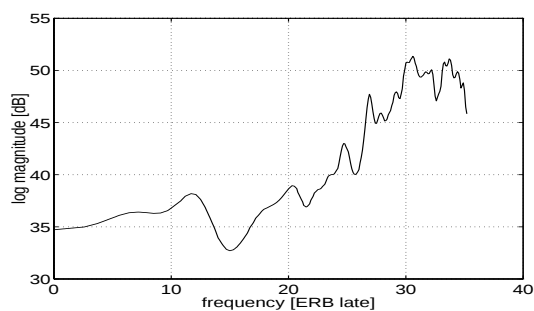
図 2.2: 話者 fj の分析結果 (側音化構音)



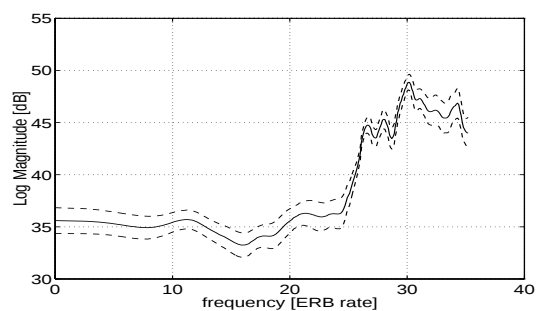
(a) フレーム 1



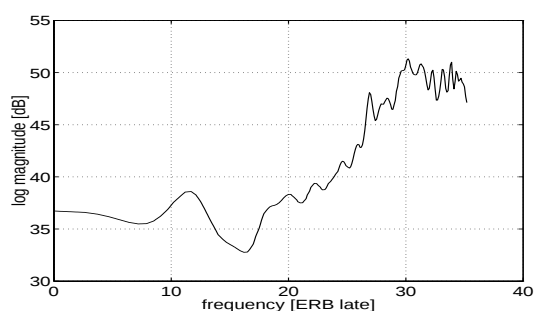
(e) 連続 30 フレームの系列



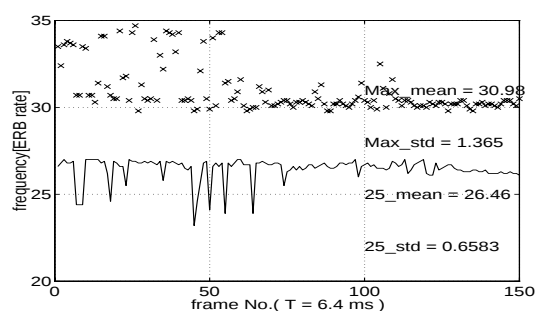
(b) フレーム 2



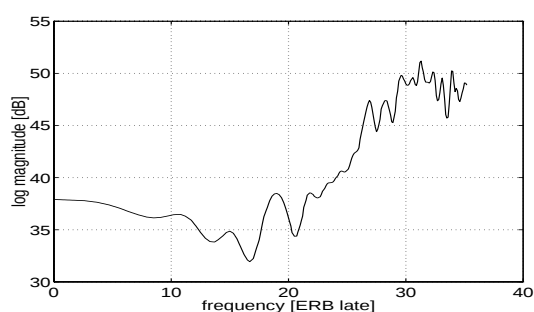
(f) 平均スペクトルと標準偏差



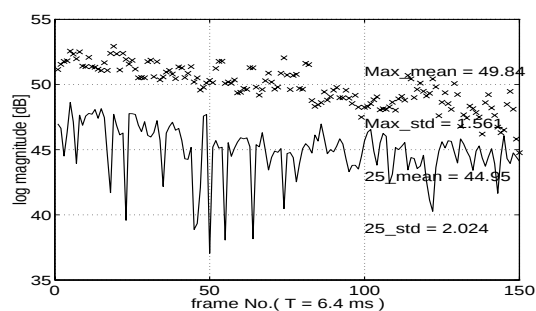
(c) フレーム 3



(g) 各フレームのピーク周波数

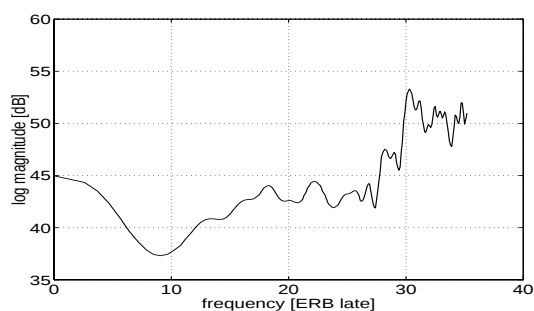


(d) フレーム 4

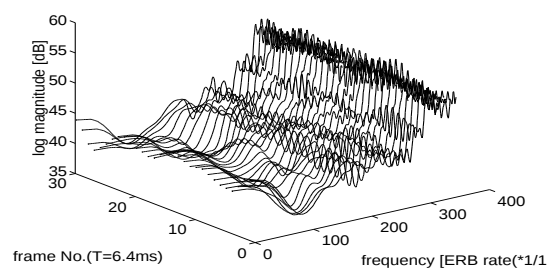


(h) ピーク周波数における
パワーレベル

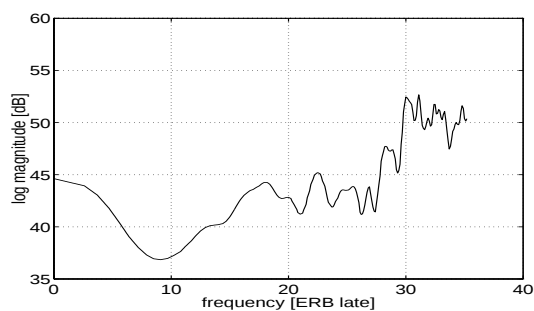
図 2.3: 話者 hr の分析結果 (側音化構音)



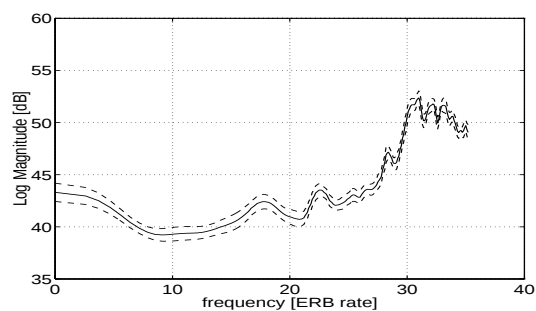
(a) フレーム 1



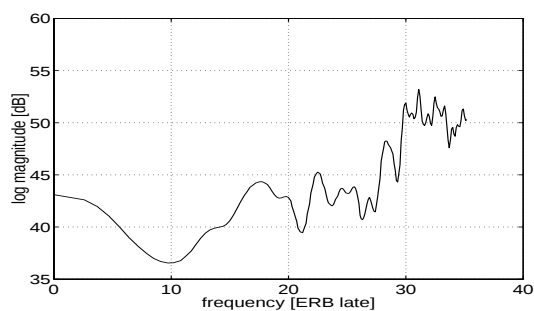
(e) 連続 30 フレームの系列



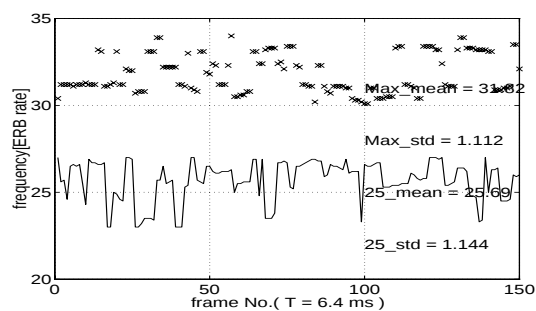
(b) フレーム 2



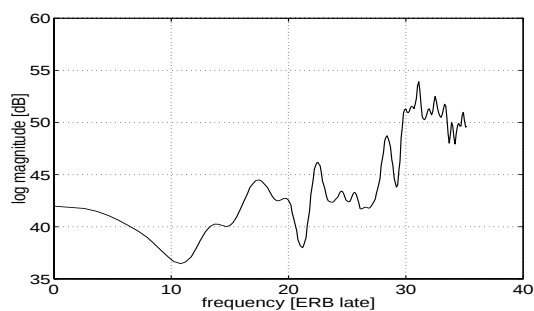
(f) 平均スペクトルと標準偏差



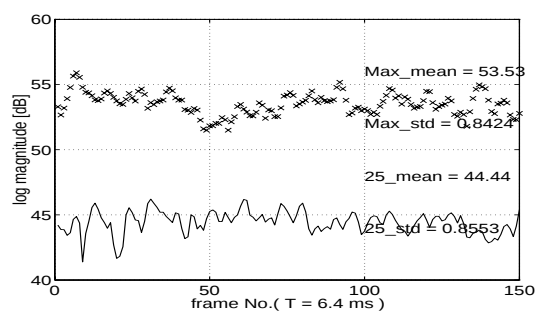
(c) フレーム 3



(g) 各フレームのピーク周波数

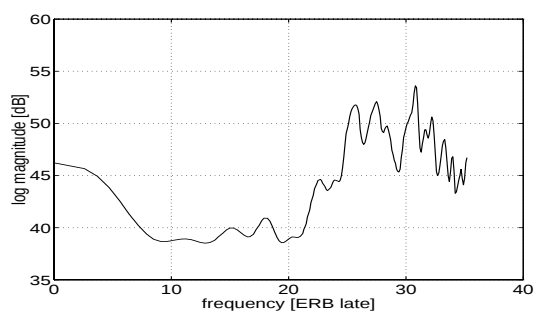


(d) フレーム 4

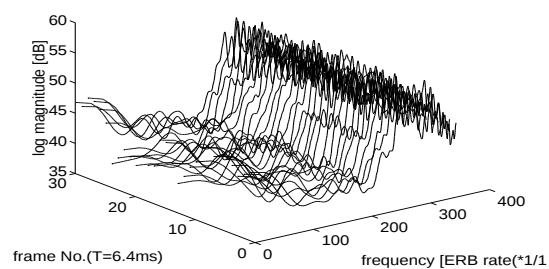


(h) ピーク周波数における
パワーレベル

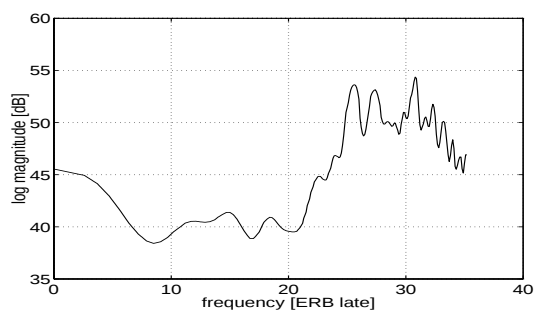
図 2.4: 話者 FT の分析結果 (正常構音)



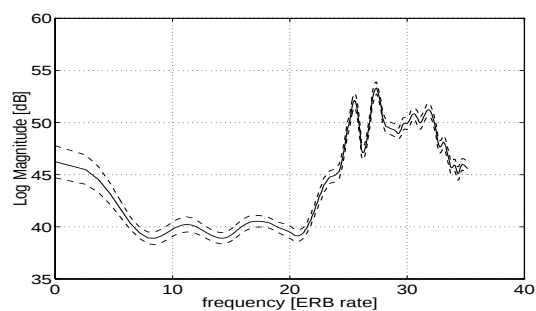
(a) フレーム 1



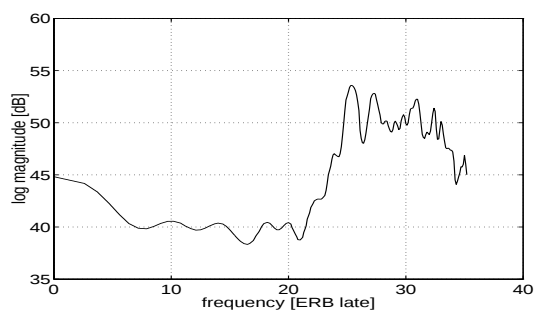
(e) 連続 30 フレームの系列



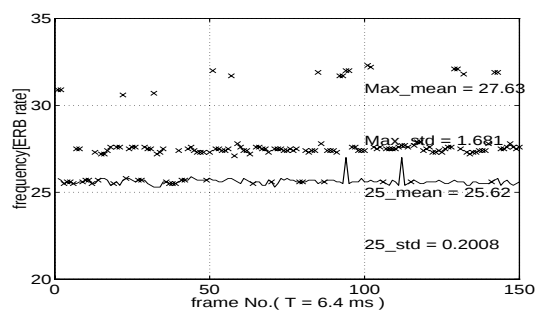
(b) フレーム 2



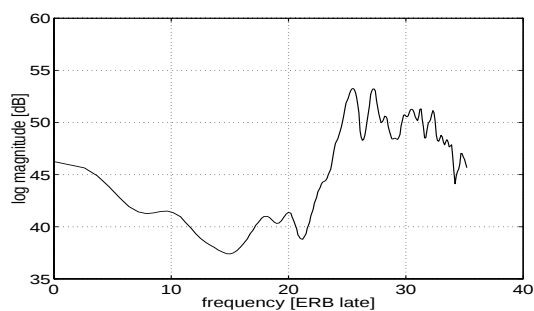
(f) 平均スペクトルと標準偏差



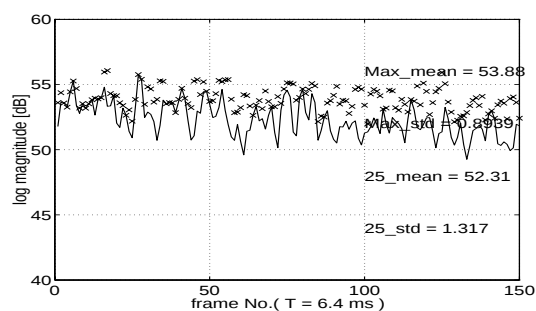
(c) フレーム 3



(g) 各フレームのピーク周波数



(d) フレーム 4



(h) ピーク周波数における
パワーレベル

図 2.5: 話者 SN の分析結果 (正常構音)

表 2.2: 正常な /f/ の周波数特性 [ERB rate]

カットオフ周波数									平均値 : 27.1
NM	NT	OT	SM	TY	YT	FT	KR	TN	SN
26	28	27	29	27	27	27	27	28	25

最大ピーク周波数									平均値 : 29.5
NM	NT	OT	SM	TY	YT	FT	KR	TN	SN
27.8	30.6	29.4	31.6	27.9	28.5	32.7	29.4	30.1	27.6

25 ERB rate 付近のピークの有無									
NM	NT	OT	SM	TY	YT	FT	KR	TN	SN
有	有	無	無	無	無	無	無	無	有

表 2.3: 側音化した /f/ の周波数特性 [ERB rate]

カットオフ周波数								平均値 : —
in	mi	sa	ued	uen	ur	fk	hr	fj
28	24	28	30	平坦	24	24	26	24

最大ピーク周波数								平均値 : —
in	mi	sa	ued	uen	ur	fk	hr	fj
30.9	25.2	30.7	31.3	23.8	25.2	25.5	30.2	25.9

25 ERB rate 付近のピークの有無								
in	mi	sa	ued	uen	ur	fk	hr	fj
有	有	有	有	無	有	有	有	有

2.3.1 単一話者での比較分析結果

話者 in は, 比較的症状の軽い側音化構音話者であり, 正常構音と聴取される場合と, 僅かに側音化と聴取される場合がある. 全く同じ条件下で前後して発話された, この2つの音声のスペクトル包絡の平均スペクトルを重ねて, 図 2.6に示した.

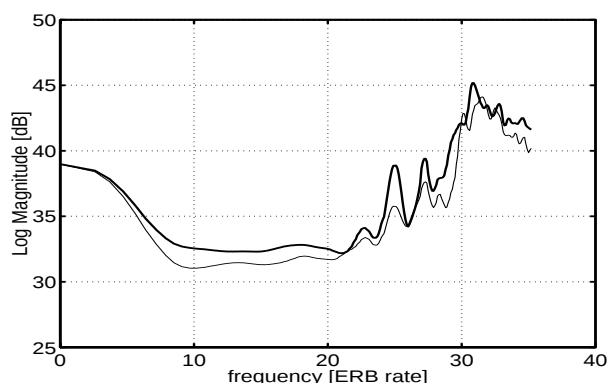


図 2.6: 話者 in の平均スペクトル (細線:正常構音時, 太線:側音化構音時)

この図から, 側音化構音と判断された音声には, 他の帯域がほとんど正常構音と同じなのであるのに対し, 25 ERB rate 付近に突出したピークが出現していることが分かり, このピークが言語臨床家が側音化構音の判断に用いている特徴量の1つであると考えられる.

2.4 考察

分析結果を見ると, 正常構音の最大ピークは 30 ERB rate (5.54kHz) 付近に見られ, 近傍の周波数帯域で十分なパワーがある. また, 側音化構音に特徴的な 25 ERB rate (3.14kHz) 付近でのピークは, 3 名の話者以外の正常構音には見られず, スペクトル包絡形状の話者間のばらつきが少ない.

一方, 側音化構音の場合, スペクトル包絡の形状に話者間でのばらつきが大きい, 共通して 25 ERB rate 付近にピークが発生している. その帯域でパワーが最大となる場合 (話者 mi, uen, ur, fj, fk) と, より高域で最大となる場合 (話者 in, sa, ued, hr) がある. また, 30 ERB rate (6.20 kHz) 以上の高域にはピークが無く, パワーが急激に落ち込んでいるか, 平坦になっている.

パワーレベルの時間変動について、正常構音の場合は各帯域でほぼ一定であるのに対し、側音化構音では、25 ERB rate 付近にみられるピークのパワーが他の帯域と比較して時間変動が多くなっている（図中、側音化構音の (f) から）。また、時間変動に周期性がみられるものもある（話者 ur(h) などから）。

ただし、側音化構音話者の低域パワーの分散（変動）が大きいのは、録音時に空調設備による騒音が加わったためと考えられ、今回の音響特性では、考慮に入れていない。

以下に側音化構音に特徴的な音響特性をまとめる。

2.4.1 静的音響特性

側音化構音を特徴づける静的音響特性として、以下の3点が挙げられる。

- 正常構音と比較して、話者間で包絡形状の違いが大きい。これは、話者間で構音方法の違いの大きさに起因する。
- 31 ERB rate (6.20 kHz) 以上の高域にはピークが無く、パワーが落ち込む。正常構音の場合、その帯域にピークがある（図 1(b)）。
- 25 ERB rate (3.14 kHz) 付近にピークが発生する。その帯域でパワーが最大となる場合と、より高域で最大となる場合がある。

ただし、各項目は単独で側音化構音と判定される条件ではなく、側音化構音は全ての条件を満たしている。その例として、25 ERB rate 付近にピークがあっても、31 ERB rate 以上の帯域に十分なパワーがある場合には、正常な音声と認識されている（正常構音話者 NM, NT, SN）。

2.4.2 動的音響特性

側音化構音を特徴づける動的音響特性として、以下の3点が挙げられる。

- 25 ERB rate 付近でのピークのパワーの変動が他の帯域や、正常構音のピークでの変動より大きい。
- パワーの変動に 30ms 程度の周期性があるものもある。

- パワーが最大となる周波数が変動しており、ピークが割れたり消滅することもある。

今回、正常構音と側音化構音それぞれについて、9 名と 10 名の分析を行ったが、正常な /*f*/ については、はっきりとした傾向を調べることができた。

一方、/ *f* / の異常構音時の口腔形状は、個人によって様々であるため、異常構音の症例ごとにスペクトル包絡の形状も大きく異なる。そこで、構音方法が客観的に診断されている 4 名の話者については、構音形態と音響特性の関係を 4 章で詳しく述べる。

しかしながら、9 名の話者では、障害の程度と直接結びつくような音響特性を定量的に評価することは困難であった。今後、更に似たような症例をもつ話者のデータを集め、構音形態と音響特性を対応付けて分類していく必要があろう。

第 3 章

スペクトル包絡変形音の聴取実験

3.1 実験の目的

側音化構音の音響特性を分析して得られた特徴的な物理量と、実際に人間が側音化構音の独特な音色を知覚するのに用いている物理量の関係を、定量的に評価することを目的として、スペクトル包絡変形音の聴取実験を行った。

スペクトル包絡を変形した合成音声を作成し、その合成音を昭和大学歯学部で側音化構音についての診断、治療を行なっている言語治療家に呈示し、側音化構音との音色の類似性を判断し評価してもらった。

3.2 実験方法

3.2.1 スペクトル包絡変形音の作成方法

スペクトル包絡を変型させる手法として、まず、正常ならびに側音化した持続音/*j*/に対し、ノッチフィルタまたはバンドパスフィルタをかけ、特定の周波数帯域のパワーレベルを変化させる。それらの刺激音を帯域毎に組み合わせてミキシングし、合成音声を作成した。

この変形方法をとれば、長時間平均スペクトルは同じ概形でも、正常構音と側音化構音のある周波数帯域を入れ換えることにより、パワーの変動が局所的に大きい音声と、正常構音のある帯域を強調／抑制することにより、変動が一様に少ない音声とを作成することが可能である。

また、時間領域での変形であるため、周波数領域で変形し逆変換する LMA 分析合成系 [11] [12] などと比較して、音質の劣化が少ないのも大きな利点である。

この合成音声の聴覚印象を比較評価することにより、静的特性と動的特性のどちらが大きな要因となっているかを定量的に評価した。

3.2.2 使用した原音声

5 名の発話による側音化構音と、2 名の正常構音 (内、1 名は側音化構音と同一話者) について、20kHz サンプリングで 1 秒間の音声区間を切り出した。

その後、パワーレベルを正規化し、スペクトル変形を行ない、音声区間の始端、終端にそれぞれ 50ms のテーパーを掛け、刺激音とした。

実験で使用した原音声は以下の通りである。単一話者内で、正常構音と側音化構音の音声を得られる in については、他の 4 名とわけて取り扱った。

(注):話者の識別の際に、大文字で表しているのは正常構音話者であり、小文字は側音化構音話者である。

- 正常構音 (Normal) .. 話者 TN の発話による連続持続音 /f/
- 側音化構音 (LM) .. 話者 ur,fj,hr,fk の発話による連続持続音 /f/
- 正常構音 & 側音化構音 .. 単一話者 IN(in) の発話による連続持続音 /f/

3.2.3 スペクトル変形パターン

刺激音の変形パターンは表 3.1 に示す 11 通りである。

なお、表中の*注 4 にある、 X の帯域については、側音化構音の話者によってスペクトル包絡のピークの周波数が異なるため、話者毎に入れ換える帯域を変えている。具体的には、図 2.1(g) ~ 図 2.3(g) 中に示す 25 ERB rate 付近のピーク周波数 (平均値) を F とした時、 $X = F - 2 \sim F + 3$ ERB rate の帯域を入れ換える対象とした。

Normal については正常構音話者 2 名 [TN],[IN], [LM0] ~ [LM9] については側音化構音話者 5 名 [ur,fj,hr,fk,in] 分の合計 52 個の刺激音を作成した。

表 3.1: スペクトル変形パターン

識別子	0-23	X (注:4)	28-31	31- [ERB rate]
LM0	側音	側音	側音	側音
LM1	正常	*注 1	正常	正常
LM2	正常	*注 2	正常	正常
LM3	正常	雑音	正常	正常
LM4	正常	側音	正常	正常
LM5	側音	側音	側音	正常
LM6	側音	側音	側音	*注 3
LM7	側音	正常	側音	側音
LM8	正常	*注 1	側音	側音
LM9	側音	雑音	側音	側音
Normal	正常	正常	正常	正常

* 注 1 : 正常構音のパワーを大きくしたもの
側音化構音のピークのパワーレベルにあわせた.(約 1.5 倍)

注 2 : 正常構音のパワーを大きくしたもの
側音化構音のピークのパワーレベルより大きい.(約 2 倍)

注 3 : 側音化構音のパワーを大きくしたもの
正常構音の高域のパワーレベルにあわせた.(約 2 倍)

表 3.2: 各変形音声の特徴

				高域のパワー	
				あり	なし
25	なし			Normal.	LM7.
ERB	あり	変動なし		LM1,LM2.	LM8.
rate		変動あり	周期性なし	LM3.	LM9.
ピーク			周期性あり	LM4,LM5,LM6.	LM0.

3.2.4 対応する特徴量

表 3.1の変形パターンに対応した音響特性は表 3.2に示すようになっている.

2 章での分析結果から,32 ERB rate 以上の高域のパワーの有無と 25 ERB rate 付近のピークに着目し,それらの特性を変化させて LM1 ~ LM9 を作成し,側音化構音の聴覚印象に影響の大きい特徴量を比較できるようにした.

正常構音の原音声は [Normal] であり,側音化構音の原音声は [LM0] である.

図 3.1に正常構音の原音声 [TN], 側音化構音の原音声 [ur0] とその変形音声のスペクトル包絡を示す. [ur1] は,[ur2] とほぼ同形であるので省略した.

3.2.5 聴覚印象の評価方法

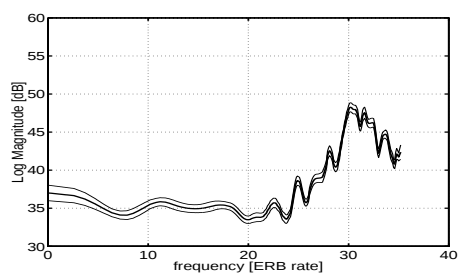
被験者は昭和大学歯学部において日頃,側音化構音の診断,治療を行なっている 4 人の言語臨床家である.スペクトル変形した各発話をランダムに並びかえて提示し,側音化構音との類似性を評価してもらった(1 人につき順序を変えて 3 セット).評価方法として,各刺激音に対し以下の 3 段階を階層的に評価してもらった.

1. 各発話 共通 — [(A) 極めて正常な /f/ ⇔ (B) 正常な /f/ ⇔ 歪音]

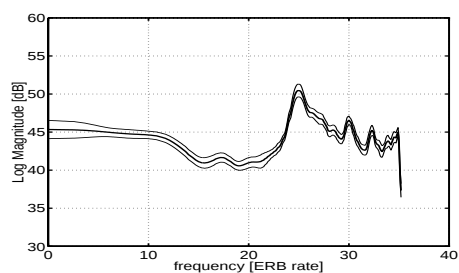
2. 歪み音と判断した場合 — 側音化構音との類似性を評価

[側音化構音による歪音 ⇔ (F) それ以外の歪音 (雑音)]

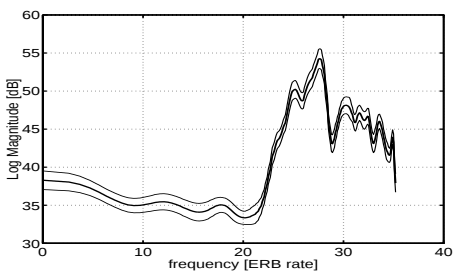
3. 側音化構音と判断した場合 — 側音化構音の障害の程度を評価



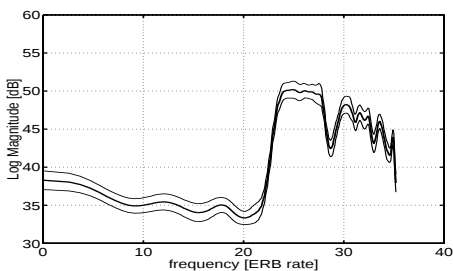
正常構音原音声 [TN]



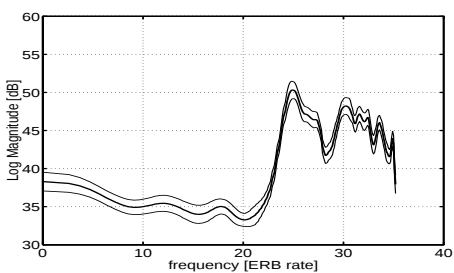
側音化構音原音声 [ur0]



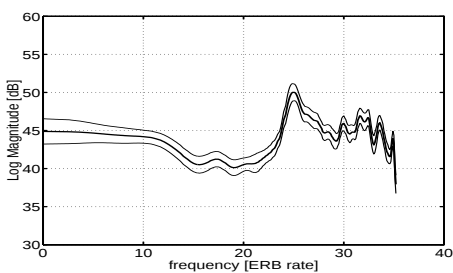
変形音声 [ur2]



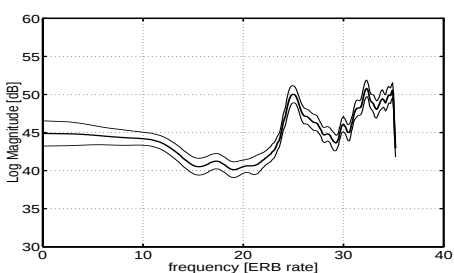
変形音声 [ur3]



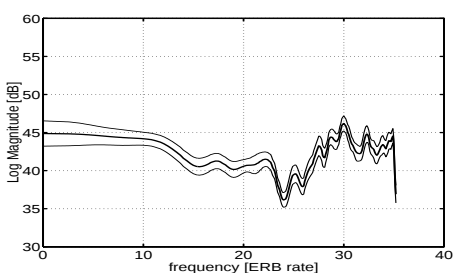
変形音声 [ur4]



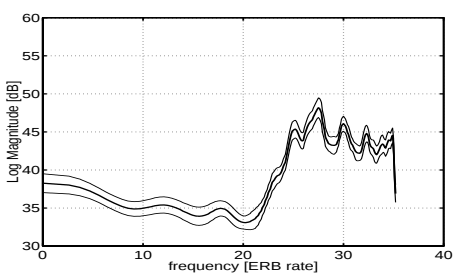
変形音声 [ur5]



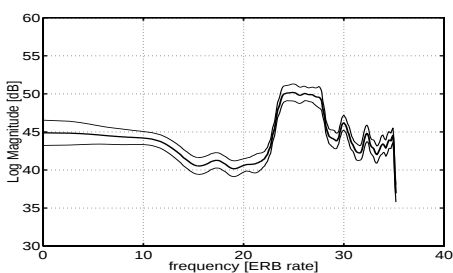
変形音声 [ur6]



変形音声 [ur7]



変形音声 [ur8]



変形音声 [ur9]

[(C) やや側音化⇔ (D) かなり側音化⇔ (E) 側音化]

この3段階評価とは別に、各音声の印象(構音方法のイメージ等)を自由にコメントしてもらっている。詳しい実験条件や、回答用紙は付録Aに示す。

また、実際に評価を行なうときには、表3.3に示す、2つの項目を用いて点数化した。

表 3.3: 聴覚印象の点数化方法

評価項目	(A)	(B)	(C)	(D)	(E)	(F)
側音化構音との類似性 (LM Point)	0	1	2	3	4	-
雑音との類似性 (Noise Point)	-	-	-	-	-	1

3.3 実験結果

側音化構音の4名の話者についての実験結果を図3.2から図3.5に、話者4名の平均と正常構音(TN)の結果を、図3.6と表3.4に示す。

表3.4は、前述の表3.2にLM Point(Noise Point)の結果を書き込んだものである。

また、図3.7に、話者IN(in)の実験結果を示す。

LM Pointが高い場合、側音化構音と判断された割合が高いことを示し、低い場合は正常構音と判断された割合が高いことを示している。

また、Noise Pointが高い場合は、側音化構音以外の歪音(雑音)と判断された割合が高いことを示している。この中には、[雑音に聞こえる]評価や、[/s/や/sw/に近く聞こえる]評価が含まれている。

各話者の得点結果中のLM Pointの最大値は40であり、Noise Pointの最大値は10である。絶対値に意味はないので、他の刺激音との相対評価を行なう。

3.4 考察

前述した通り、話者間で評価にバラツキが大きいので、総合して当てはまる傾向、ある話者には当てあまる傾向がみられた。そこで、総合結果からの考察と、各話者別の考察を行

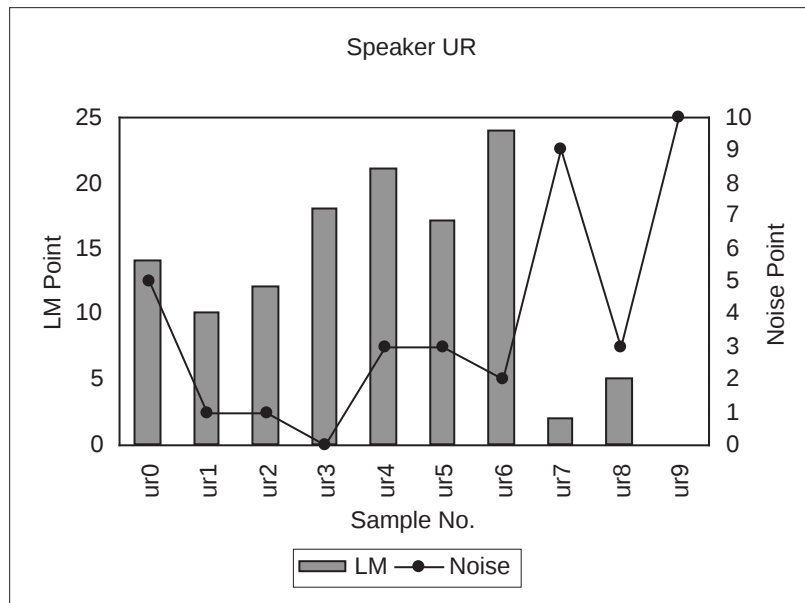


図 3.2: 話者 ur の得点結果

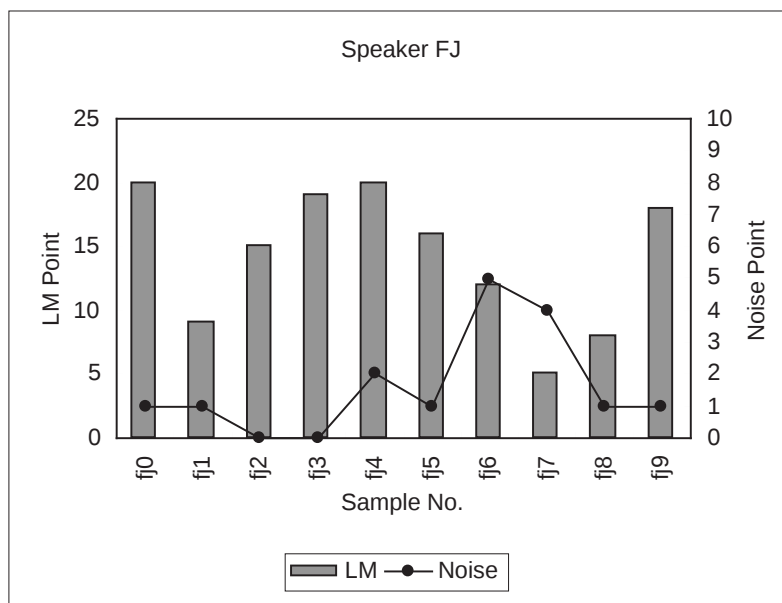


図 3.3: 話者 fj の得点結果

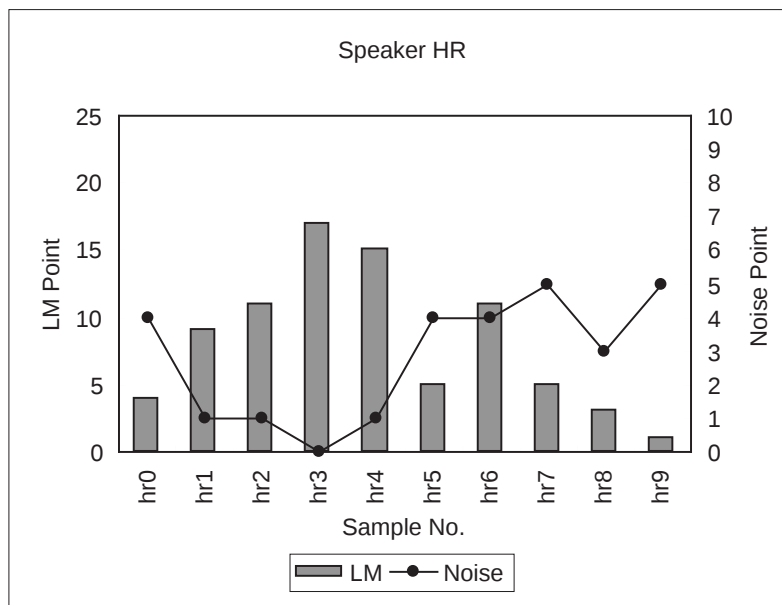


図 3.4: 話者 hr の得点結果

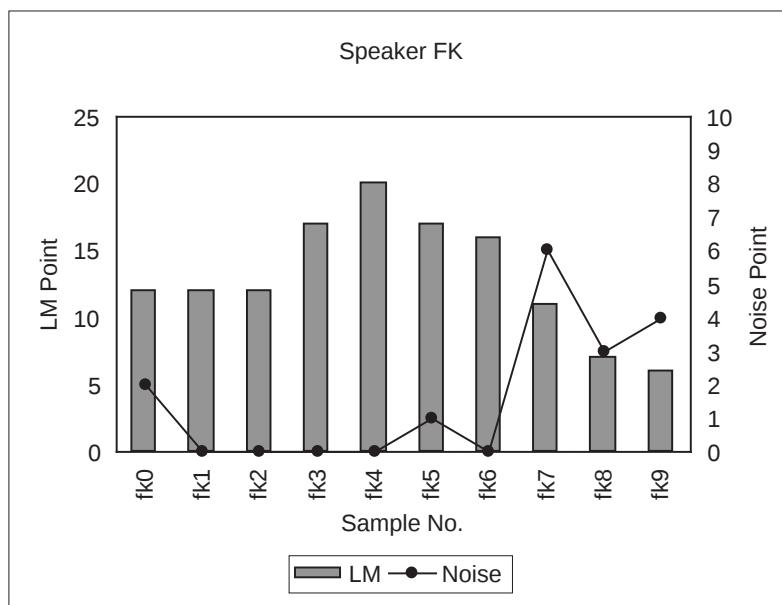


図 3.5: 話者 fk の得点結果

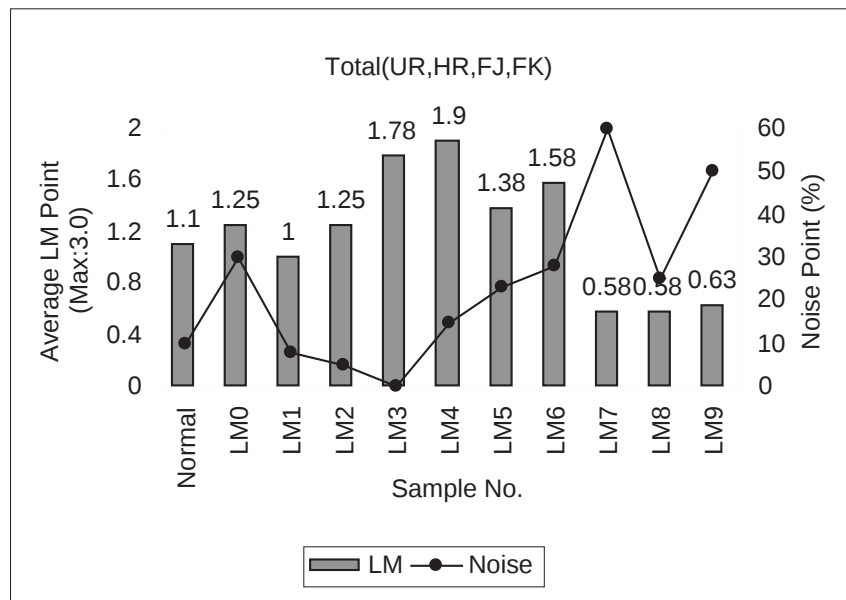


図 3.6: 4 話者の平均得点結果

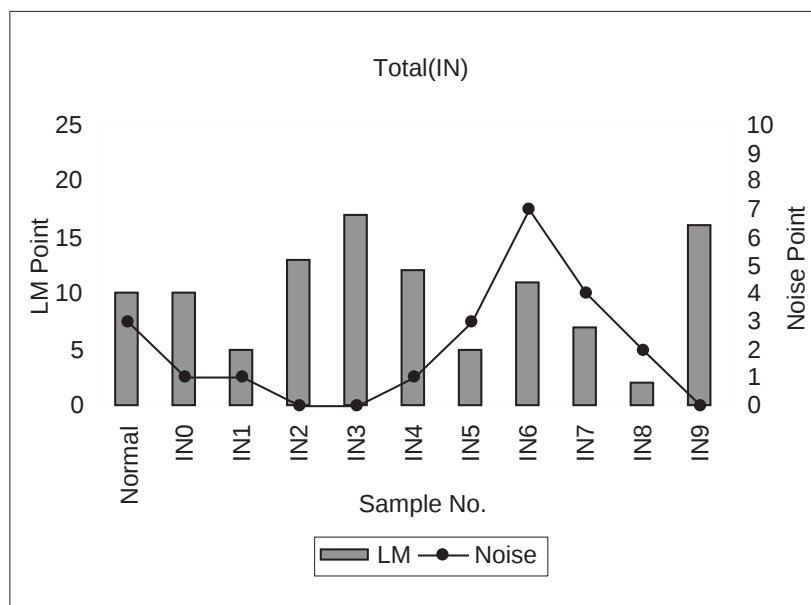


図 3.7: 話者 IN(in) の得点結果

表 3.4: 音響特性と対応した平均得点 (4 話者)

				高域のパワー	
				あり	なし
25	なし			1.10(10)	0.58(60)
ERB	あり	変動なし		1.00(8),1.25(5).	0.58(25).
rate		変動あり	周期性なし	1.78(0).	0.68(50).
ピーク			周期性あり	1.90(15),1.38(23),1.58(28).	1.25(30).

なった。

3.4.1 総合結果からの考察

まず、側音化構音の独特な音色を形成している帯域について、以下の4点から検討した。

- 25 ERB rate 付近のピークの影響 — 25 ERB rate 付近のピークがある場合とな
い場合の聴覚印象を比較し、25 ERB rate 付近のピークが側音化構音と判断する重要
な特徴量となっているかを検討する。

結果 — 高域、25 ERB rate 付近ともにピークの無い音声 [LM7] は、側音化構音ではな
い歪み音と判断された割合が非常に高かった。内訳は [雑音], [/s/に近い], [/sw/に近い] の
3つの評価がほぼ同数であった。/f/の音響特性が失われ、/s/と/f/の音韻境界のような特性
になったためであろう。

同様に、ピークのない正常構音の原音声 [Normal] については、本来、[極めて正常な/f/]
と判断されるべきであるが、今回の実験結果では、[正常な/f/(LM Point =1)] と判断され
た場合が多くなった。このため、側音化構音との類似性 (LM Point) では [LM7], [LM9] の
ほうが低い得点になっている。しかし、これは [LM7], [LM9] がほとんどが側音化構音以
外の歪み音 (雑音) と評価されているために生じたもので、正常構音と評価したものではな
い。自由にコメントしてもらった回答からも、[Normal], [LM1], [LM8] の3つの音声と同
程度で最も正常構音に近いと判断されている。一方、それ以外の 25ERB rate 付近にピー
クがある音声では、側音化構音と判断される割合が高くなっている。

また,[LM1] と [LM2] とを比較すると, 他の帯域は同じであるのに, 25ERB rate 付近のパワーレベルを大きくした [LM2] の方が, 側音化構音と判断された割合が多い。

以上の結果から, 側音化構音の重要な特徴は 25 ERB rate 付近のピークにあり, そのパワーレベルが大きい程, 側音化構音と判断されやすい事が分かる。

更に, 話者 hr のように, より高域の 30 ERB rate 付近にもピークがある場合に, 25 ERB rate 付近のピークを取り除けば, 側音化構音とは判断されなくなる ([LM7]) ことから, 25 ERB rate 付近のピークの存在が, 側音化構音と判断される絶対条件であることが分かる。

- 32 ERB rate 以上の帯域のパワーの必要性 — 正常と診断された /f/ にも, 25 ERB rate 付近にピークがある場合があった。32 ERB rate 以上の帯域のパワーが十分あれば, それ以下の帯域が, 側音化構音によるものであっても, 正常に聞こえるかを検討する。

結果 — 高域にパワーが無い音声 [LM7] では, 側音化構音ではない歪み音と判断される場合が多かった。原音声 [LM0] や, [fj9] では, 比較的多く側音化構音と判断されている。

一方, 低域は側音化構音を使い, 高域にパワーを持たせた音声 [LM5], [LM6] について, 原音声よりも正常構音との評価が高くなったのは, 話者 fj の場合のみで, 他の話者については, 逆に側音化構音と高く判断されている。これは, 原音声の聴覚印象はばらつきが大きく, コメントより僅差で雑音と判断した被験者が多く, 実際の印象よりも, 低い評価値として現れているのが原因であろう。

また,[LM1] と [LM2] では, 高域, 25 ERB rate 付近, とともにパワーがあるにもかかわらず, 正常構音と判断されている割合が高くなっているが, この点については, 動的要因の項目で述べる。

この結果から, 側音化構音の高域にパワーを付加しても, 正常には聞こえやすくないといえる。つまり, 低域側 (25 ERB rate 付近) の情報の方が, 側音化構音の判断に重要であることが分かる。

もうひとつの結論として, 高域にパワーがない場合には, 正常な /f/ ではないと判断されている。(特に, 雑音と判断されやすくなる。) このことから, 言語治療家が用いる聴覚印象の判断基準として, まず高域にパワーがなければ, 無条件に歪み音と判断し, それから 25 ERB rate 付近の情報を基に, 判断するという2段階での判断をしている可能性が示唆される。

- 25 ERB rate 付近のピークの変動の影響 (動的要因) — これまでに、側音化構音の特徴はスペクトル包絡の動的変化にもあることが示唆されている。

このことを、定量的に評価するために、25 ERB rate 付近のピークに (周期的/非周期的な) 変動をもたせ、その動的特性と聴覚印象の関係を検討する。

結果 — 25 ERB rate 付近にピークがある音声のうち、変動のない音声 [LM1],[LM2],[LM8] は、正常構音と判断されている割合が比較的高い。これは、正常と診断された /f/ にも、25 ERB rate 付近にピークがあるという分析結果を支持している。

また、雑音によってピークを持たせた音声 [LM3] と、側音化構音の原音声からピークを切り出した音声 [LM4] を比較すると、原音声を使ってピークに周期性をもたせた方が、わずかであるが、より側音化構音に聞こえるという結果が得られた。言語治療家が、その周期性を聞き分けているかは、この結果からでははっきりとは分らない。

一方、側音化構音の原音声 [LM0] と、雑音置換した音声 [LM9] の比較では、高域にパワーがないため、25 ERB rate のピークの変動の違いによる聴覚印象への影響が顕著に現れる。[LM0] のほうが、側音化構音と判断した割合がかなり高くなっている。

これらの結果から、側音化構音の聴覚印象を決めるのに、スペクトル包絡の動的変化が非常に大きな要因となっていることがわかる。また、動的変化にも周期性がある方が側音化構音と判断されやすいことがわかった。

- 帯域雑音を入れ換えた影響 —

25ERB rate 付近の帯域を雑音で入れ換えた [LM3] と、[LM9] の音声について検討する。高域にもパワーがある [LM3] については、側音化構音と判断されている。

一方、高域にパワーのない [LM9] では、雑音と判断される場合 ([ur9,hr9,fk9]) と側音化構音と判断される場合 ([fj9]) とで二分される結果となった。

この原因として、話者 ur,hr,fk の場合、25 ERB rate 付近の鋭いピークのパワーレベルに対し、同じパワーの雑音を前後 5 ERB の帯域に入れ換えたため、パワー多寡となった雑音成分が強調されてしまい、雑音と判断されてしまったものと思われる。

一方、話者 fj の場合、25 ERB rate 付近のピークは鋭くなく、高域側にも平均して大きなパワーがあるため、原音声とほぼ同じ評価になった。

3.4.2 各話者毎の傾向

各話者の原音声の音響特性の違いから、スペクトル変形をした時の聴覚印象は変わってくる。以下に、各原音声の音響特性と、他の話者と比較した傾向をまとめる。

時間平均したスペクトル包絡（標準偏差含む）は、話者 ur,fj,hr,in(IN) については、図 2.1(f)～図 2.3(f),2.6 (f) に示したとおりである。話者 fk,TN(Normal) については、以下の図 3.8に示す。

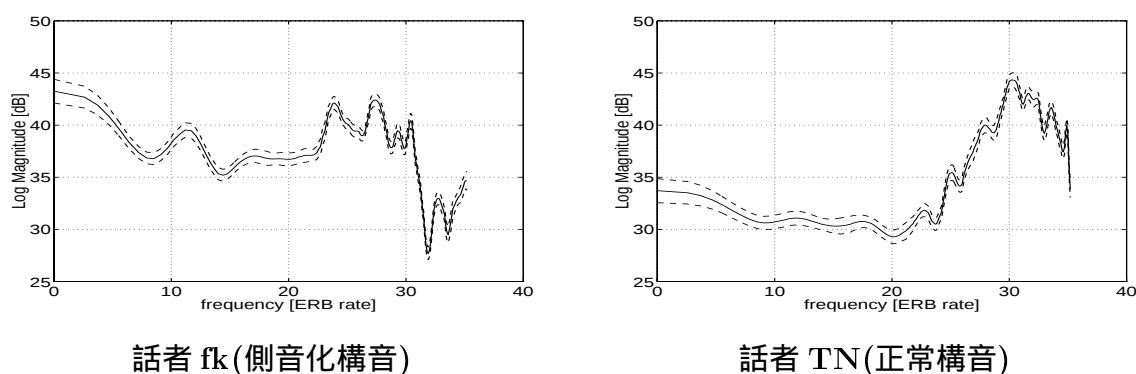


図 3.8: 話者 fk,TN(Normal) の平均スペクトルと標準偏差

話者 ur 25.0 ERB rate 付近に鋭いピークがあり、パワーレベルの変動が大きく、周期性も見られる。それ以上の高域にはピークが全くない。

— 25 ERB rate 付近のピークの情報がそのまま残っている [ur4] から [fj6] で側音化構音との評価が高くなった。同様にピークのある原音声 [ur0] では、低い評価値になっているが、コメントではかなり側音化に近い判断となっている。

話者 fj 25.9 ERB rate から 30.9 ERB rate の区間で平均して大きなパワーがあり、パワーレベルの変動が比較的小さい。

— [fj0],[fj4],[fj9] が、同程度で最も側音化構音との評価が高くなった。話者 ur とこの結果をあわせて、側音化構音の聴覚印象には 25 ERB rate 付近のピークの情報が非常に重要であることがわかる。

話者 hr 26.5 ERB rate 付近に突出したピークがみられ、パワーレベルの変動は非常に大きく、不規則である。それ以外の帯域は正常構音のスペクトル包絡に近く、側音化構

音の中でも, 高域のパワーが多い.

— 高域のパワーが比較的多いため, 全般的に正常構音と判断された割合が高い. また, 25 ERB rate 付近を雑音で入れ換えた [hr3] が最も側音化構音に近いと判断された. つまり, 高域側にパワーがある場合には, 25 ERB rate 付近の相対的なパワーレベルが大きい程, 側音化構音と判断されている.

話者 fk 25.0 ERB rate 付近に鋭いピークがあり, 高域でのパワーの落ち込みが最も激しい.

— 高域のパワーが最も少ないため, 全ての变形パターンで側音化構音と判断された割合が平均より高かった. つまり, 高域にパワーがない方が, 側音化構音と判断しやすくなる.

話者 in 側音化構音では, 25 ERB rate 付近に突出したピークがあり, 他の帯域は正常構音と同じ包絡である.

— 話者 hr と同様, 原音声よりも, 雑音で入れ換えた [in3], [in9] で側音化構音と判断される割合が高くなった.

第 4 章

声道モデルによる側音化構音のモデル化

4.1 目的

側音化構音の構音形態が正常構音と異なっていることは構音形態を客観的に測定することにより明らかになっており、構音形態が変わることによって、音響特性の違いが生じているものと思われる。

しかしながら、側音化構音に特徴的なスペクトルピークの変動等の音響特性が、声道形状や構音方法のどこに起因しているかについては、あまり解明されていない。

そこで、まず声道モデルを用いて音声生成メカニズムを工学的に実現する。次に、構音形態を変化させたときの音響特性をシミュレートし、スペクトル形状の違いが側音化構音の構音形態のどの部分に起因しているのか検討を行う。

これまでの研究 ([13] など) から、構音形態 (声道形状) を推定することにより、そこから発声される音声の音響特性 (声道伝達特性) を一意に求めることができるようになってきたが、逆に、音響特性から声道形状を一意に求めることは不可能であり、統計的な評価しかできないことを付記しておく。

4.2 モデルの概要

4.2.1 音声の生成過程

構音器官である声道では、発声器官から発せられた音源波を入力とし、出力として声道の共振特性をもった音声波が得られる。

共振系への入力となる音源について、母音、半母音、鼻音などの生成の場合では、声帯の振動が音源となるのに対し、摩擦音の生成の場合には、声道の狭め (constriction) 付近で発生する乱気流が雑音源となっている。

一方、声道は断面の形と断面積が連続的に変る一種の音響管である。形状が変ることによって管の伝達特性が変わり、同時に種々の音源が与えられることによって、種々の形のスペクトルや時間構造をもった言語音が口唇から放射されている。

4.2.2 音響管の等価電気回路

図 4.1 に示す音響管での音波の伝搬現象は、両端の体積速度 (U) と音圧 (P) の関係を電流 (I) と電圧 (V) の関係に置き換える事により、電気回路素子によって等価に表現する事が出来る [14]。

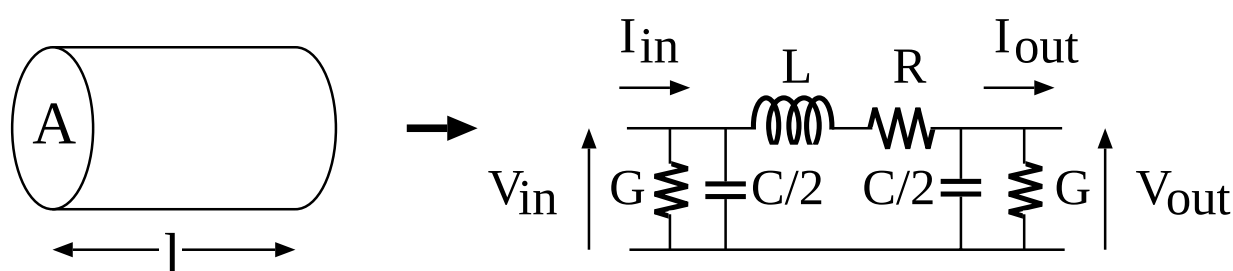


図 4.1: 音響管とその等価回路

等価回路の C と L は波動方程式を基に、セクション長 l と断面積 S をパラメータとして、以下のように求めることができる。ここで R, G は各セクションでの損失に相当する。

$$L = \rho l / k A. \quad C = k A l / \rho c^2. \quad (4.1)$$

$\rho = 1.14 \times 10^{-3} \text{gm/cm}^3$: 空気の密度 ,
 $c = 3.53 \times 10^4 \text{cm/sec}$: 音速 ,
 k : インピーダンスレベルによる定数.

声道断面積に対応した C と L の値を持つ単セクションを直列に接続することにより, 声道全体の共振特性等価な共振特性を持つ電気回路を設計することができる.

実際の電气的アナログ装置 (LEA) を用いて共振周波数の変化を観測した報告では, 各セクション長は 0.5cm と一定にして, 35 個 (声道長 17.5cm) 前後のセクションを直列接続している.

また, 口唇端には, 音声の放射特性を考慮に入れるために, 図 4.2 にしめす放射インピーダンスを負荷として与える. 詳しい値の導出方法については, 文献 [14] を参考されたい.

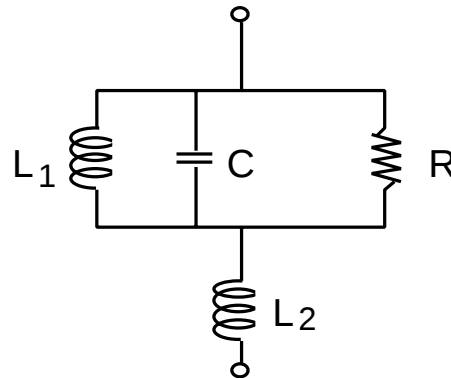


図 4.2: 放射インピーダンスの構成素子

$$\begin{aligned}
 L_1 &= 4.7 \times 10^{-3} / A^{\frac{1}{2}} \\
 R &= 510 / A \\
 C &= 0.0093 \times 10^{-6} A^{\frac{1}{2}} \\
 L_2 &= 0.79 \times 10^{-3} / A^{\frac{1}{2}}
 \end{aligned}$$

1 つのセクションの伝達関数 $H_i(\omega)$ は, 声帯側の端での体積速度を U_i , 口唇側の端を

U_{i-1} とすると,

$$H_i(\omega) = \frac{U_{i-1}(\omega)}{U_i(\omega)}$$

と定義できる. また, 声道全体の伝達関数 $H(\omega)$ は,

$$H(\omega) = \frac{U_{out}(\omega)}{U_{in}(\omega)} \left(= \prod_{i=1}^K H_i(\omega) : \text{声帯音源の場合} \right)$$

と記述できる. ここで, $U_{out}(\omega)$ は放射部 (唇) での体積速度, $U_{in}(\omega)$ は音源での体積速度, K はセクション数を表す.

実際の電気回路では, 入力音源での電流 (I_{in}) と放射部での出力電流 (I_{out}) から, 声道伝達関数を求めることができる.

4.2.3 駆動音源

回路を駆動する音源として, 以下の2種類を仮定している. 離散化した伝達関数への入力も, 同様な波形を使用した.

母音の場合 喉頭部での声帯振動のかわりに, 直列回路の入力端に周期インパルス電流源を並列に付加する.

摩擦音の場合 狭めの調音位置に対応する2つのセクション間に, 雑音の電圧源を直列に付加する.

4.3 計算機での実装

母音の場合, 以下の手順で電気回路を等価なデジタルフィルタに変換する.

1. 声道断面積関数から, 各セクションでの C, L の値を求め, そのセクションでの F パラメータを求める. それらを縦続接続することにより, 声道全体の回路の F パラメータを求める.
2. F パラメータの値と, 放射インピーダンスの値を使ってアナログでの伝達関数 $H(\omega)$ を求める.

3. $H(\omega)$ をラプラス変換して $H(s)$ を求め, 式 3 に示す双二次変換により, 離散的に等価な伝達関数 (デジタルフィルタ) $H(z)$ を求める. (F_s はサンプリング周波数である.)

$$H(z) = H(s) \Big|_{s=2f_s \frac{(z-1)}{(z+1)}}$$

4. $H(z)$ から, フィルタの振幅特性や, 入力音源波形を加えた時の出力音声を求める.

子音の場合には, 音源の位置が声道の端にはないので, F パラメータだけでは, 表現できない. そこで音源から声帯側のセクションは, インピーダンスを直接計算し, 音源から口唇側の F パラメータに接続した. 手順 (2) 以降は母音と同様である.

4.4 シミュレーション (1)

まず, この声道モデルの有用性を検証するために日本語 3 母音 /a/, /o/, /i/ の生成過程をシミュレーションした.

声道断面積関数は, 文献 [15] を参考に, 図 4.3, 図 4.5, 図 4.7 に示す値を用いた.

セクション長は, 声道断面積の変化が緩やかな区間では長くとり, その区間で断面積が一定とみなした. 図中のグラフでは, 高さが同じ区間を単一セクションとして扱っている. 声帯から口唇までのセクション数は約 10 個である.

4.4.1 結果と考察

各声道断面積から求めた, 声道の周波数特性を図 4.4, 図 4.6, 図 4.8 に示す.

また, 表 4.1 ~ 表 4.3 は, 理想的な Formant 周波数と実際に得られた共振周波数を示す.

音声スペクトルから求められる Formant 周波数は音源のピッチなどにより値が変化するが, Formant 周波数とほぼ同じ値の声道の共振周波数が, シミュレーションの結果から得られた.

また, 共振周波数における帯域幅もかなり現実の音声に近いものが得られた.

母音 /a/ と /o/ は, 声帯に近い部分の断面積はほぼ同じであり, /o/ のほうが, 口唇部の断面積が小さくなることによって, F_1, F_2 が /a/ よりも低くなるが, この現象も本モデルから確認された.

以上の点から, 本声道モデルの妥当性が確認された. しかし, 合成された音声の音質は, あまりよくない結果となった.

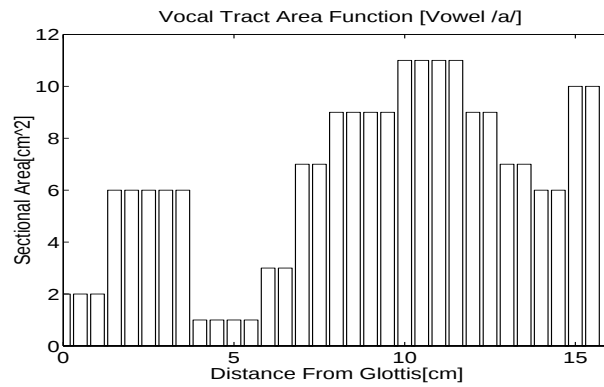


図 4.3: 母音/a/の声道断面積伝達関数

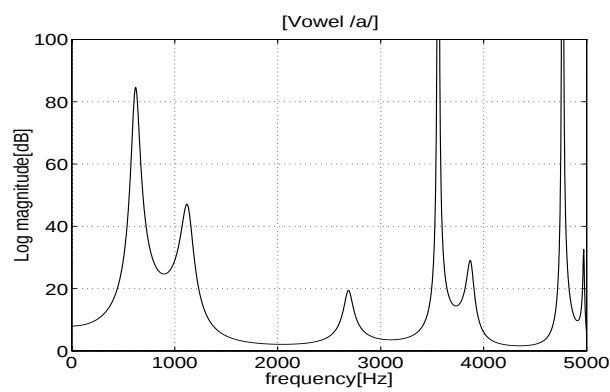


図 4.4: 母音/a/の声道の周波数特性

表 4.1: 母音/a/の声道共振周波数と Formant 周波数

	F1	F2	F3
希望する Formant 周波数 [Hz]	730	1090	2440
得られた共振周波数 [Hz]	622	1121	2690

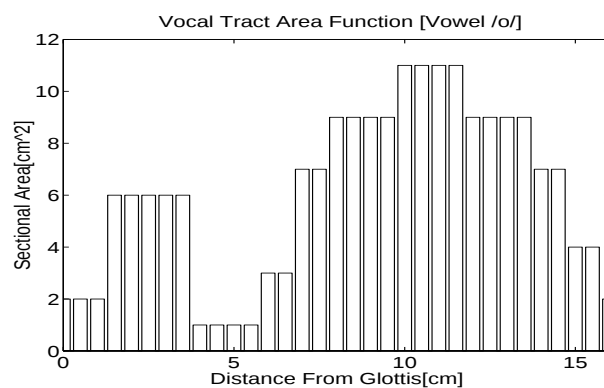


図 4.5: 母音/o/の声道断面積伝達関数

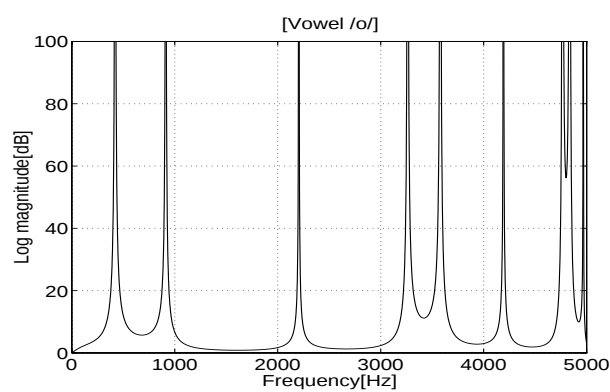


図 4.6: 母音/o/の声道の周波数特性

表 4.2: 母音/o/の声道共振周波数と Formant 周波数

	F1	F2	F3
希望する Formant 周波数 [Hz]	570	840	2410
得られた共振周波数 [Hz]	454	930	2207

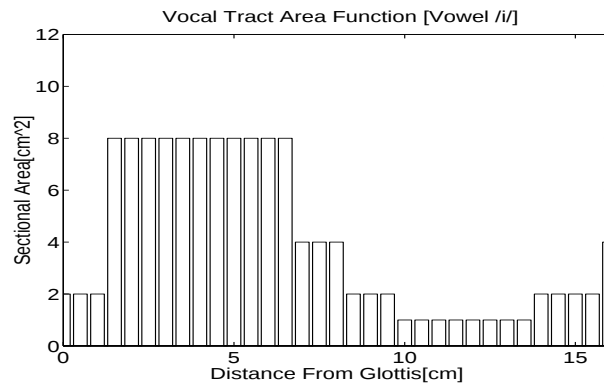


図 4.7: 母音 /i/ の声道断面積伝達関数

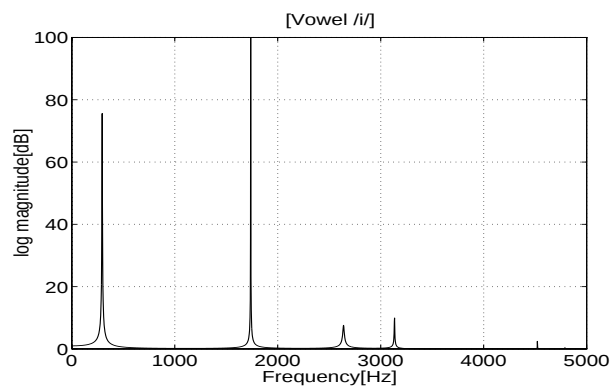


図 4.8: 母音 /i/ の声道の周波数特性

表 4.3: 母音 /i/ の声道共振周波数と Formant 周波数

	F1	F2	F3
希望する Formant 周波数 [Hz]	730	1090	2440
得られた共振周波数 [Hz]	622	1121	2690

4.5 シミュレーション (2)

母音の場合, モデルがうまく機能していることが分かった. 続いて, 側音化構音と正常構音による, 連続摩擦音 $/f/$ の生成過程をシミュレーションした.

4.5.1 声道形状の測定

第3章の実験でも使用した, 側音化構音の患者4名 (ur, fj, hr, fk) については, 昭和大学歯学部にて X 線で撮影されたセファロトレースグラフを用いた [8], セファロトレースグラフから, 声道の部分抜きだし, 声道を3つのセクションに分け, 垂線の中点を結ぶ線を声道の中心線としてその長さを計測した. 2次元画像であるため, 声道断面積は厳密に求めることが出来ないが, 声道の各セクションが円柱であると仮定し, 直径から声道断面積を推定した. 正常構音話者については, 文献 [16] を参考に推定した声道長と声道断面積を用いた.

声道のセクション長と断面積を表 4.4 に示す. 図 4.9 に, 対応した音響管モデルを示す.

表 4.4: 各話者の声道パラメータ

Speaker	Back cavity		Tongue construction		Front cavity	
	Δl_b	A_b	Δl_c	A_c	Δl_f	A_f
ur	6.6	5.3	5.7	1	2.7	1
fj	5.9	12.1	5.2	1	1.7	3
hr	7.4	7.2	4.4	1	2.0	3
fk	7.5	12.8	4.4	1	2.6	2
Normal	10.0	8	2.5	1	2.0	4.5

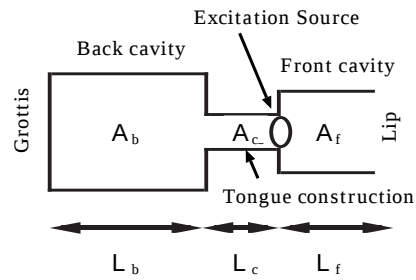


図 4.9: 3 セクシ ョンの音響管モデル

4.5.2 シミュレーション結果と考察

正常構音話者の結果を図 4.10に, 4 名の側音化構音話者 (ur,fj,hr,fk) の結果を図 4.11 ~ 図 4.14に示す. 声道のセクシ ョンは 1cm 単位で区切って周波数特性を計算した.

周波数特性は, 実際の音声のものとはかなり, 異なっている. この原因として, まず, 声道断面積を求める時の誤差が考えられる. 音響管モデルによる声道の近似は, 簡略化のためかなりの制約を加えているから, 実際の声道の特性とはやはりくいちがってくる.

しかし, 正常構音の場合に, 25 ERB rate 付近の帯域にはパワーが少なくなる現象は見られた.

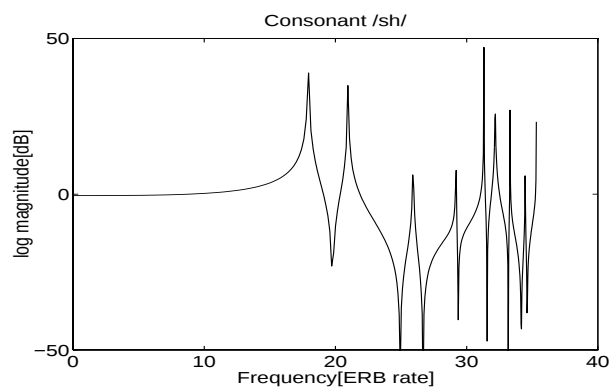


図 4.10: 正常構音話者の声道の周波数特性

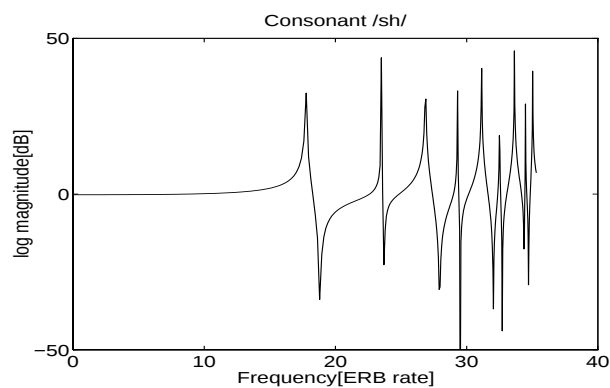


図 4.11: 話者 ur の声道の周波数特性

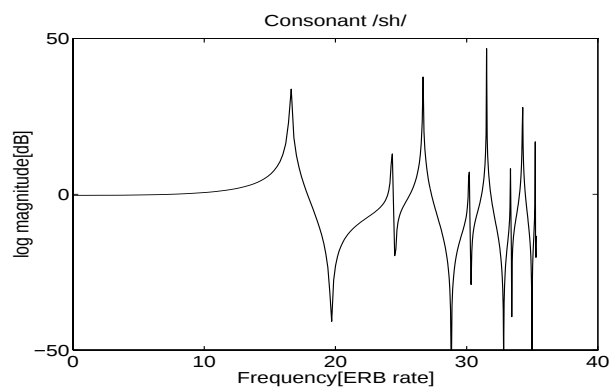


図 4.12: 話者 fj の声道の周波数特性

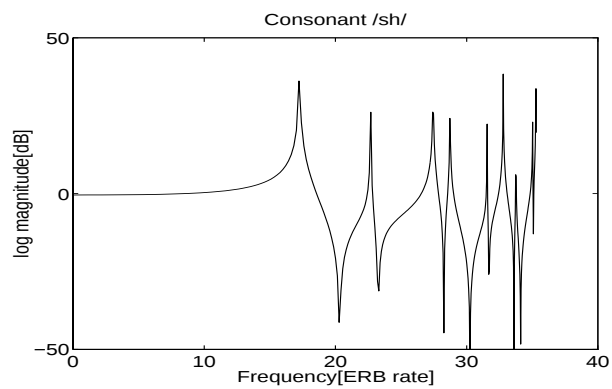


図 4.13: 話者 hr の声道の周波数特性

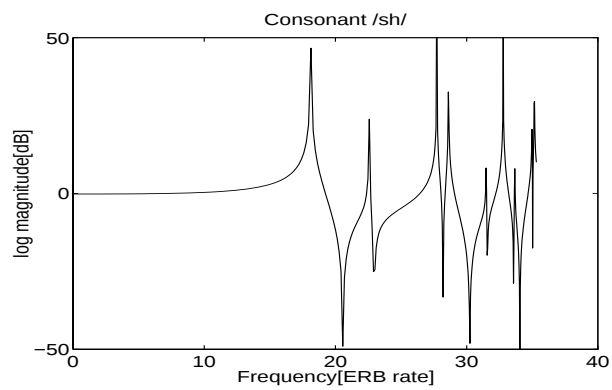


図 4.14: 話者 fk の声道の周波数特性

4.5.3 狭めの長さとの共振周波数の関係

動態の客観的観測から、側音化構音の話者により舌が口蓋に付着している面積が違ふことが報告されている [17]. 口蓋への付着面積が大きくなると、呼気は、大きく両側に回り込むようになるので、狭めが長くなり、る。そのときに、声道の周波数特性がどのように変化するかシミュレートした。

$$\Delta l_b + \Delta l_c = \text{constant}.$$

として、 Δl_c を大きくしていった時の、共振周波数の変化を図 4.15にしめす。簡略化のため、セクション数は3つで計算を行なった。

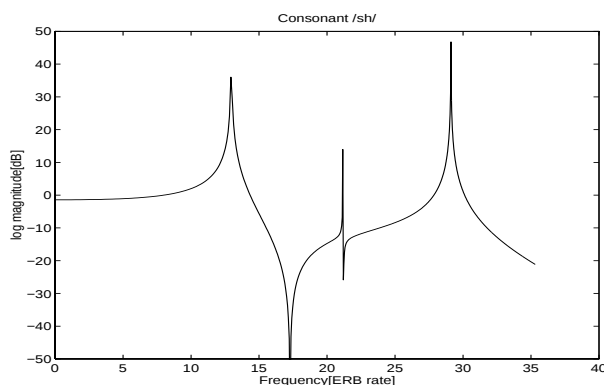


図 4.15: 狭めの長さとの共振周波数

この図から、30ERB rate 付近のピーク周波数は変わらないのに対し、20ERB rate 付近のピークの周波数とそのパワーレベルが大きく変化していることがわかる。

第 5 章

結論

5.1 本研究の成果

本研究では、側音化構音と正常構音の持続音 /f/ を用いて、聴覚印象と、音響特性、構音形態の 3 つの関係について検討を行った。その結果、以下のことが明らかになった。

1. 音響特性について。側音化構音と診断された持続音 /f/ のスペクトル包絡には、25 ERB rate 付近に 時間変動の激しいピークがある。また、32 ERB rate 以上の高域にはパワーが少ない。
2. 25 ERB rate 付近のピークの周期的な変動が、側音化構音の聴覚印象を決める最も大きな要因となっている。
3. 声道の狭めの長さや断面積によって、声道伝達関数のピーク周波数や振幅値が変化する。すなわち、側音化構音のスペクトル包絡のピークは、狭めの位置と長さによって決まる。

5.2 今後の課題

本研究では、側音化構音には共通して 25 ERB rate 付近にピークの変動があり、その帯域の情報が、側音化構音の判断する時に最も重要な情報となっていると結論づけたが、言語治療家がそれ以外の音響特性をもとに判断している可能性もある。実際に、スペクトル

包絡の傾きにも、正常話者とは異なる特徴があるという報告があるが、この傾きが聴覚印象にどのくらい影響をおよぼすか、といった検討の余地は残っている。

一方、側音化構音と一概に分類される音声でも、障害の程度や構音形態は話者毎に多種多用であり、それによる音響特性もさまざまである。今後、音響特性を利用した診断支援システムの構築を目指すには、構音形態ごとに側音化構音を細分化して、音響特性や聴覚印象の違いなどを検討する必要がある。（しかし、患者数が少なく、最初に述べたように構音形態を客観的に評価するには非常にコストがかかるから、実現はかなり困難である。）

また、側音化構音の構音形態は厳密には解明されていないから、声道モデルを時可変にするなどの高精度化を更に行えば、その手助けも可能である。

第 6 章

謝辞

常日頃から数多くの有益な御助言、御指導を頂きました赤木正人助教授、飯島泰蔵教授、岩城護先生、ならびに本学の教官の皆様に深く感謝いたします。

様々な点で、御助言、御指導くださいました飯島、赤木両研究室の皆様に深く感謝いたします。

また、本研究を進めるにあたって、側音化構音の音声を提供して下さり、貴重な助言もして頂きました昭和大学歯学部 鈴木規子助教授ならびに博士課程 1 年の斉藤浩人氏に感謝いたします。

音声を録音させていただいた本学の 9 名の皆様、聴取実験に協力していただいた昭和大学歯学部他の 5 名の言語臨床家の皆様にも、厚くお礼申し上げます。

最後に、2 年間の研究生生活を支えて下さった、家族、友人、ほか全ての皆様に厚く感謝いたします。

参考文献

- [1] 新版 聴覚と音声 pp.243 社団法人 電子情報通信学会 (1980)
- [2] 楊 長盛, 粕谷 英樹, 加納 滋, 佐藤 俊彦: “MRI による声道形状の精密計測法の検討”, 信学論 (A), **J77-A**, 10, pp.1327-1335(1994).
- [3] 鈴木 規子, 山下 夕香里, 道 健一: “軟口蓋ダイナミックパラトグラフィーによる舌運動様式の観察”, 音声研究会資料, H-82-15(1982-2)
- [4] 今泉 敏: “音声分野における測定法”, 音響学誌, **47**, No.5, pp.357(1991)
- [5] Suzuki, N., et al.: “Crosslinguistic study of lateral misarticulation using electropalatography”, European J. Disorders of Comm., 30, 237-245(1995).
- [6] Takahashi, K.: “Acoustical evaluation of Japanese lateral misarticulation”, J. Japanese Cleft Palate Association, 11, 2, pp.178-193, (in Japanese)(1986).
- [7] 荒井隆行, 岡崎恵子, 今富摂子: “摩擦音「s」の口蓋化構音に対する分析と合成”, 信学技報, **SP94-100** (1995-3).
- [8] 高木 直子: “側音化構音の知覚と音響特性に関する関係の基礎的研究”, 北陸先端科学技術大学院大学修士論文, (1995).
- [9] 今井聖, 阿部芳春: “改良ケプストラム法によるスペクトル包絡の抽出”, 信学論, **J62-A**, 4, pp.217-223 (1979)
- [10] Glasberg, B.R. and Moore, B.C.J.: “Derivation of auditory filter shapes from notched-noise data”, Hearing Research, **47**, pp.103-138.

- [11] 今井聖, 北村正 : “対数振幅特性近似フィルタを用いた音声の分析合成系” , 信学論 **J61-A**,No.6(1978-6).
- [12] 今井聖, 北村正 : “対数振幅特性近似 (LMA) フィルタ” , 信学論 **J63-A**,No.12(1982-12).
- [13] 榊原 宏, 中井 孝芳, 鈴木 久喜 : “鼻腔の音響伝達特性のシュミレーションモデルによる検討” 信学技報 ,**SP95-112**,pp.25-32(1996-1)
- [14] Fant G. ,et al.: “An Electrical Analog of the Vocal Tract” JASA vol.25,No4,pp.734-742(1953)
- [15] Chiba,T. and Kajiyama,M. : “The Vowel its Nature and Stracture” , Tokyo-Kaiseikan Pub.Co.,Ltd.(1942)
- [16] Victor W. Zue.:“Acoustic Theory of Speech Production” , Tech. Report of Depasrtment of Electrical Engineering & Computer Science, MIT(1989)
- [17] Suzuki,N.,et al.:“Articulatory and Acoustical Characteristics of Lateral Misarticulation in Japanese” ASA and ASJ Third Joint Meeting ,**3aSC23**, pp.919-922(1996)

付録 A

スペクトル包絡変形音の聴取実験での、被験者の実験条件（被験者に実験前に渡したもの）と解答用紙.