

論文 / 著書情報
Article / Book Information

論題(和文)	音声知覚におけるスペクトル変化情報の役割り
Title	
著者(和文)	古井貞熙
Author	SADAOKI FURUI
出典(和文)	騒音研究会資料(H85-01-6), , , pp. 1-8
Journal/Book name	, , , pp. 1-8
発行日 / Issue date	1985, 1

音声知覚におけるスペクトル変化情報の役割

ON THE ROLE OF SPECTRAL TRANSITION FOR SPEECH PERCEPTION

古井 貞熙

Sadaoki FURUI

日本電信電話公社武蔵野電気通信研究所

Musashino Electrical Communication Laboratory, N.T.T.

内容梗概：種々の時点で話頭又は話尾切断した音節の明瞭度と、切断時点でのスペクトルの動的特徴との関係を調べ、次の点を明らかにした。(1)話頭／話尾切断音節の明瞭度は、切断位置がスペクトル変化極大時点の前か後かで急激に変化し、この時点を含む10ms程度の区間が、最も重要な音韻性情報を担っていることを示している。(2)子音明瞭度と母音明瞭度の従属関係から、子音の知覚には母音の知覚が必要であるが、逆は成り立たないことがわかる。子音は基本的に音節を単位として、母音部へのスペクトルの遷移を主たる手掛かりとして知覚されていると言える。(3)多次元空間内でのスペクトルの動特性に関して、後続音韻ターゲットへの運動方向が定まり、その方向への変化速度が最大になった時点で、音韻が知覚されると推定される。

1. まえがき

音声はそのほとんどの部分において時間的に変化する特徴を有している。このため音声知覚において、スペクトルの静的特徴とともに動的特徴が重要な役割を果たしていることはよく知られている。

子音のスペクトルの動的特徴に関しては、これまでに主として合成音を用いた多くの研究があり、雑音成分の時間的及び周波数的パターン、VOT (Voice onset time)、及び母音立ち上がり時点でのホルマントの遷移によって、各子音の特徴が表現されている[1, 2]。これらのパターンは、着目している子音の前後の母音や、さらにこの前後の音韻が変わると、調音結合のために大きく変化するが、基本的には、調音方法による差は第1ホルマントの動特性に、調音位置による差は第2及び第3ホルマントの動特性に反映されると言われている。最近のスペクトル・リーディングの研究でも、子音を分類するプロトコルとして、定常的な雑音成分の存在等とともに、前後の母音を考慮したホルマントの動特性が用いられている[3]。FM音を用いたマスキング実験によれば、周波数変化時にマスキングパターンが大きなオーバーシュートを示し、FM音に対する人間の聴感度が高いことが確認されている[4]。

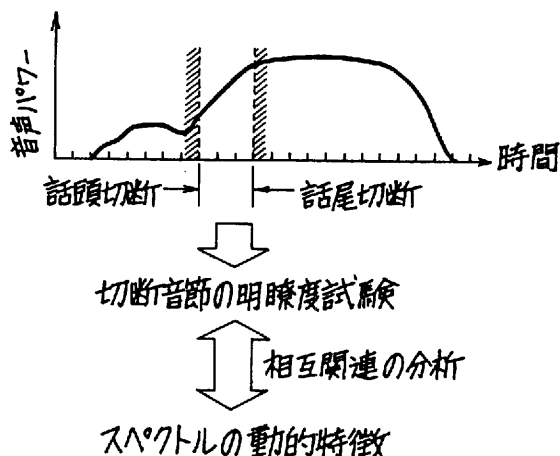
一方母音に関しても、連続音声の中では定常部分を殆ど持たずに隣接の音韻に向かってスペクトルが変化

し、その母音固有の目標値には到達しない。この際にスペクトルの動特性から目標値としての母音が知覚される一種の補償作用に関して、すでにいくつかの研究が行われている[5~8]。

生理学的には、聴覚神経系において、下位の神経系では持続形ニューロンがほとんどであるのに対し、上位の神経系では刺激の過渡的性質に対して応答するニューロンが多くなり、大脳皮質や内側膝状体では、持続形ニューロンはほとんど見られなくなると言われている。持続形ニューロンでも、刺激の冒頭ではパルスの頻度が高いが時間とともに急速にパルス数が減少する順応形ともいえるべきものが多い。これらはニューロン単体の性質ではなく、ニューロン間の相互接合、他のニューロンからの抑制作用によっている。猫の下丘や皮質においては、刺激の周波数変化の方向性（周波数上昇、下降）に応じるニューロンも見出されている[9, 10]。

しかし人間が実際の自然音声を知覚するときに、そのどのような動的特徴が最も重要な役割を果たし、音韻性を決定づける上で、静的情報と比較してどの程度の重要性を有しているのかは、必ずしも明確になっていない。音声知覚の基本である単音節の知覚においても、その比較的先頭に近い過渡部分が重要であることは言うまでもないが、どれだけの部分があれば知覚上十分なのかは、明らかにされていない。この目的の

ため、各音節の種々の時点から前の部分や後の部分を切断除去して明瞭度を測定する実験は、すでにいくつか行われているが[11~15]、切断位置を決める基準点が音声パワーの立ち上がり又は音声パワーの最大点に設定され、その基準点からの時間をパラメータとして、異なる音節間で明瞭度が平均化されて表現されているため、明瞭度特性がややあいまいな上、スペクトルの物理的性質との対応が明確でない。ここでは、各単音節の話頭又は話尾を種々の時点で切断除去した刺激を作成して、この明瞭度試験を行い、次にこの結果を音声スペクトルの各時点での動的特徴と対照して、スペクトルの動的特徴と切断音節の明瞭度との関係を調べた。



2. 話頭／話尾切断音節の明瞭度試験

明瞭度試験の実験条件は下記の通りである[16]。

- 100音節音声：女性アナウンサー2名がそれぞれ1回ずつ発声。
- 切断方法：図1に示すように、切断位置を音節の話頭から母音定常部までの広い範囲にわたって10ms間隔ですらし、その位置で振幅を10msの時間幅で線形に減衰させて、前の部分あるいは後の部分を除去する。
- 受聴条件：種々の刺激のうちほぼ同程度の明瞭度となる100音節をランダムに並べて受聴する。
- 受聴レベル：聞きやすいレベル(0-OTR)とし、雑音等は付加しない。
- 受聴者：試験員の女性4名。
- 受聴回数：各条件について4~5回受聴した結果を分析に用いた。

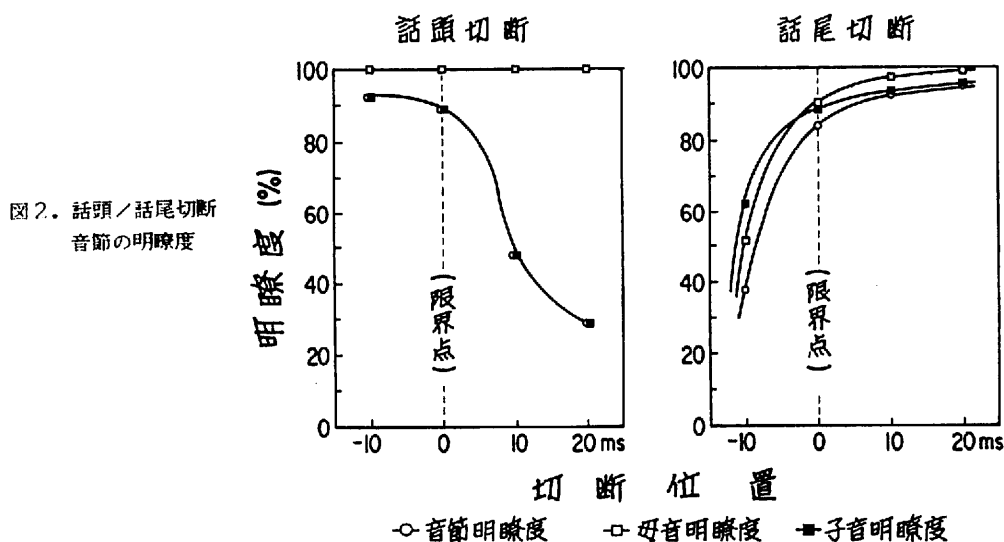
切断の結果として非常に時間長の短い刺激と長い

図1. 本研究の方法とねらい

刺激ができ、これらを混ぜて受聴するとその影響(短い刺激に対して注意が集中しにくくなる)を生ずる可能性があるため、予備実験の結果にもとづいて、ほぼ同程度の明瞭度となる100音節を組として、ランダムに並べて受聴した。また、この種のやや特殊な刺激に対しては明瞭度が安定するまでに時間がかかるので、本実験に先立って、2日間程度の予備実験を行った。

3. 明瞭度試験の結果

図2は、左に話頭切断、右に話尾切断の場合について、切断位置と子音、母音および音節明瞭度の関係を示したものである。明瞭度試験の結果として、各音節について、切断位置を10ms間隔ですらしたときの明瞭度特性が得られるので、それぞれ音節明瞭度が



80~90%となる時点を「限界点」と定め、すべての音節の明瞭度特性を、限界点をあわせて平均したのがこの結果である。

これからわかるように、話頭切断の場合は切断位置を先頭から後部へずらしていくと、子音明瞭度が限界点を過ぎた時点で90%から50%へと急激に低下する。この場合、母音明瞭度はほぼ100%に保たれるので、音節明瞭度はほぼ子音明瞭度に一致する。一方話尾切断の場合は、切断位置を後部から前部へずらしていくと、やはり限界点を過ぎた時点で80~90%から40~60%へと明瞭度が急激に低下する。この場合は母音明瞭度と子音明瞭度はほぼ並行して低下する。ただし、ミクロ的には、切断位置が限界点の前の場合は母音明瞭度よりも子音明瞭度の方が高く、限界点の後ではこの傾向は逆転する(統計的に有意)。限界点の前後10msの区間での音節および子音明瞭度の変化特性は、話頭切断と話尾切断ではほぼ等しく、対称形となっている。

図3は、話尾切断の場合の切断位置に対する明瞭度特性について、音節明瞭度が急激に変化する区間、即ち限界点の直前10msの区間の明瞭度の変化量(傾斜)を音節毎に調べ、話頭の子音別にまとめて示したものである。5母音を除く26種類の音節群で、平均46.9%/10ms、標準偏差9.0%/10msであり、音節によるばらつきは比較的小さい。細かく見ると傾斜の値は、(p>t>k)、(b>d>g)、(o>u>a>e>i)の順に小さくなっており、子音では調音点があるものほど傾斜が大きい。即ち、調音点があるものほど、音節情報が短い区間に集中していることを示しており、破裂位置と母音立ち上がり時点の時間間隔や、スペクトル変化の速度に関係していると考えられる。母音では逆に調音点がないものほど傾斜が大きく、パワーの立ち上がりの早さに関係していると考えられる。

上述の結果から、話尾切断において、母音明瞭度と子音明瞭度に従属関係があることが示唆されたので、この性質を定量的に調べるため次の検討を行った。即ち、母音明瞭度と子音明瞭度が独立であれば両者の積はほぼ音節明瞭度に等しくなり、そうでなければ差を生ずるのである[12]、話頭切断と話尾切断音節の明瞭度に関して、この関係調べた。その結果図4に示すように、話尾切断においては、母音明瞭度と子音明瞭度の積と音節明瞭度の間には、かなり広い範囲にわたって10%程度の差がある。このことは、母音明瞭度と子音明瞭度の間に従属関係があることを示しており、上述の話頭切断と話尾切断の結果の対比から、次のことが導かれる。即ち、母音の知覚には必ずしも子音の知覚は必要ではないが、子音の知覚のためには

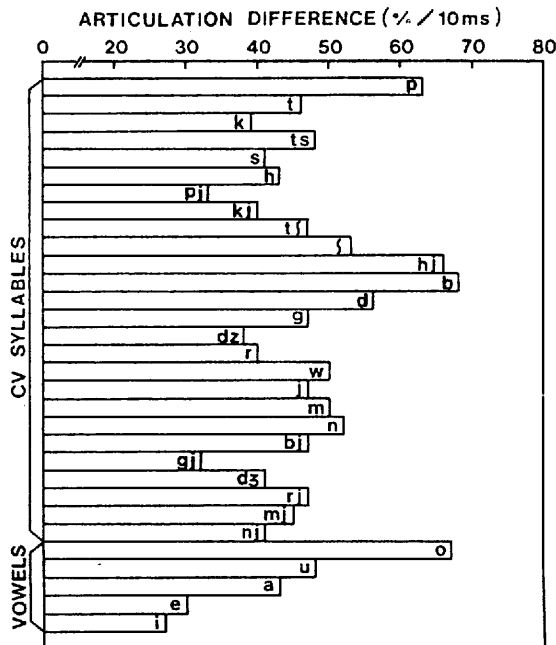


図3. 話尾切断における限界点の直前10msの区間の明瞭度の変化量(明瞭度の傾斜)

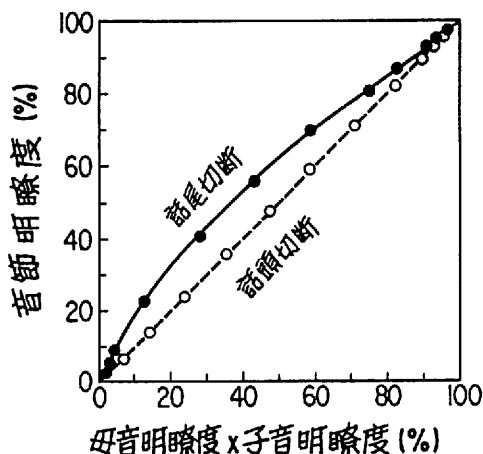


図4. 母音明瞭度と子音明瞭度の従属性

基本的に母音の知覚が必要とされる。ただし部分的には、子音単独で知覚できる機能も存在する。

話頭切断音節に特徴的な異聴には、/CjV/→/jV/ (C:子音、V:母音)、/p/→/h/、/ts/→/p/、/s/→/ts/→/t/、/pj/→/hj/、/kj/→/hj/、/ʃ/→/tʃ/、/d/→/r/等があり、話尾切断音節に特徴的な異聴には、/CjV/→/Ci/、/w/→/b.m/、/ts/→/tʃ/、/h/→/ʃ.s/、

/h,j/→/j/等があり、話尾切断の場合には、/p,t,k/相互間、/b,d,g/相互間、/m,n/相互間の異聴も多い。

4. スペクトル変化量

次に、音節の各時点におけるスペクトルの動的特徴と、その位置で切断した音節の明瞭度との関係を調べるため、音声スペクトルの動的特徴の一つの表現法として、次のようにして定義される音声スペクトルの各時点における変化量 $[17]$ を調べた。

$C_i(t)$: ケプストラム時系列 ($1 \leq i \leq p$: 次数)

$C_i(n)$: 時刻 t を原点にとり直したものの

($-n_0 \leq n \leq n_0$: 50ms 区間)

$\hat{C}_i(n) = a_i n + b_i$: 回帰直線

$$a_i = \left(\sum_{n=-n_0}^{n_0} C_i(n) \cdot n \right) / \left(\sum_{n=-n_0}^{n_0} n^2 \right)$$

$$D(t) = \left(\sum_{i=1}^p a_i \right) / p$$

$C_i(t)$ は i 次のケプストラム係数、即ち対数スペクトルの逆フーリエ変換のうちの低次の p 個の係数の時系列を示し、 p の値はここでは 10 に設定した。低次のケプストラム係数をフーリエ変換すると、平滑化された対数スペクトル包絡が得られるので、ケプストラム係数は平滑化された対数スペクトル包絡に直接対応している。 $C_i(n)$ は時刻 t を原点に取り直した時系列であり、 n の範囲は、時刻 t を中心とする 50 ms の区間とする。50 ms という長さは、我々が普通に連続音声を発声したときの平均音節長の 1/3 程度に相当し、音韻性を特徴づけるいわゆるホルマントの動特性をとらえるのに適当な長さとして、この値を選んだ。 $C_i(n)$ を図 5 に示すような回帰直線で近似し、この係数 a_i を求めると、この値は i 次のケプストラムで表現される対数スペクトル包絡の単位時間当りの変化量に対応する。この係数 a_i を、 p 次までの全ケプストラム時系列に関して二乗平均した値 $D(t)$ を計算すると、この値は p 次までのケプストラムで平滑化した対数スペクトル包絡の、単位時間当りの変化量に相当するので、これを時刻 t におけるスペクトル変化量と定義する。

5. 切断音節の明瞭度とスペクトル変化量の関係

各音節について各時点のスペクトル変化量と、その時点で話頭又は話尾切断した音節の明瞭度の関係を調べたところ、図 6 に示すような関係のあることがわかった。この図は音節 /n j o/ の場合であるが、図の上部に 5 ms 毎の各時点のスペクトル変化量と音声パワー、図の下部に各時点で切断した音節の明瞭度を示してある。これからわかるように、話頭/話尾切断

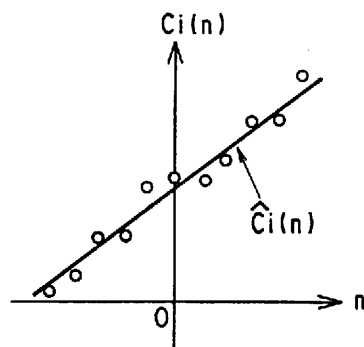


図 5. ケプストラムの線形回帰直線

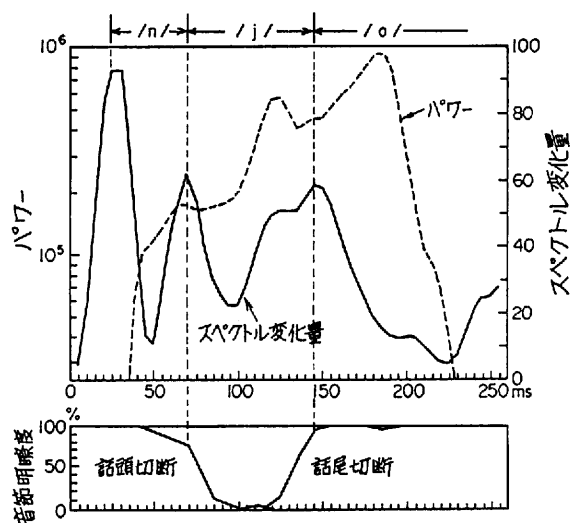


図 6. スペクトルの動的特徴と切断音節の明瞭度の関係

音節の明瞭度の限界点は、ほぼスペクトル変化量極大の時点に対応づけることができる。

すべての音節について、明瞭度の限界点とスペクトル変化極大点との時間関係を調べたところ、図 7 に示す結果が得られた。ただし /s, h, ʃ/ の定常的な無声雑音区間を有する音節は、10 ms の傾斜をつけてもその雑音区間の中間から話頭切断すると破裂性の子音への異聴を生じ、本質的に重要な音声区間を検出することが困難なため、これらの音節は話頭切断のデータからは除いている。図 7 で横軸は、明瞭度の限界点に対するスペクトル変化極大点の相対位置を示し、正の値は極大点が時間的に限界点の後にあること、負の値は極大点が限界点の前にあることを示す。破線は話頭切断の場合の出現頻度、実線は話尾切断の場合の

出現頻度を示し、両者ともほぼ同程度のばらつきがあるが、平均的には、話頭切断の限界点はスペクトル変化極大点の4.0ms手前に、話尾切断の限界点はスペクトル変化極大点の2.6ms後ろにあることがわかる。即ち、話頭／話尾切断された音節がスペクトル変化極大点を含むか否かによって、その明瞭度が大幅に変化する。このことは、スペクトルの動的特徴が音節の知覚に決定的な役割りを果たしており、子音が基本的に、後続する音韻（母音）へのスペクトルの遷移を手掛かりとして知覚されることを示している。なお図7で、分布の端付近に位置する音節には、拗音を含むものが多い。

音節を発声した音声において、比較的顕著なスペクトル変化極大点は、話頭と話尾を除いて、1～2個所存在する。拗音／j／を含む音節では、スペクトル変化極大点が通常2個所あり、上述の／n j o／の図で示したように、それぞれ話頭・話尾切断における限界点に対応する。その他の音節ではスペクトル変化が極大となる位置が、母音の立ち上がり部（母音への遷移部）の1個所となり、話頭・話尾切断における限界点がこの一つのスペクトル変化極大点に対応づけられる。

この様子を定量的に示したのが図8である。横軸は話尾切断の場合の明瞭度の限界点を基準（0）として示してある。拗音等を含まない音節では、この図の右に示すように、話頭・話尾切断音節の明瞭度の限界点はスペクトル変化極大点の両側に位置し、両者の明瞭度曲線は、スペクトル変化極大点の付近で交差する。拗音等を含む音節では、話頭／話尾切断音節の限界点が2個所の異なるスペクトル変化極大点に対応するの

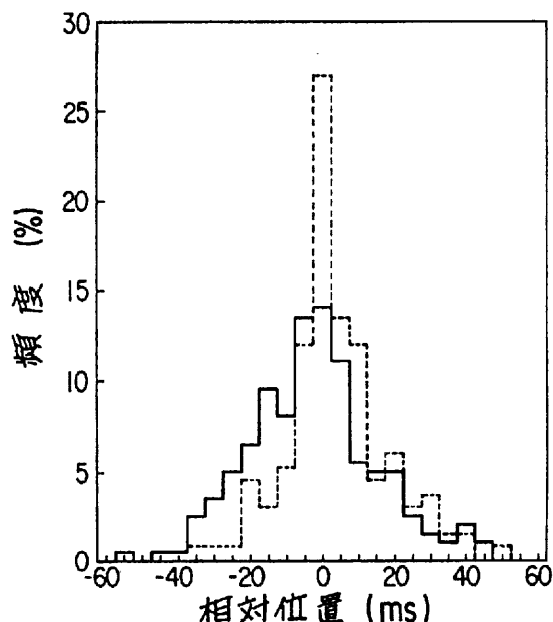


図7. 明瞭度の限界点に対するスペクトル変化極大点の相対位置

□ 話頭切断 ($\mu = 4.0\text{ms}$, $\sigma = 14.2\text{ms}$)

■ 話尾切断 ($\mu = -2.6\text{ms}$, $\sigma = 17.9\text{ms}$)

負の値は極大点が限界点の前にあることを示す

で、両者の限界点は平均で60ms程度離れている。以上の結果から、音節の主たる音韻性情報は、スペクトル変化極大点を含む10ms程度の区間内にあると推定される。

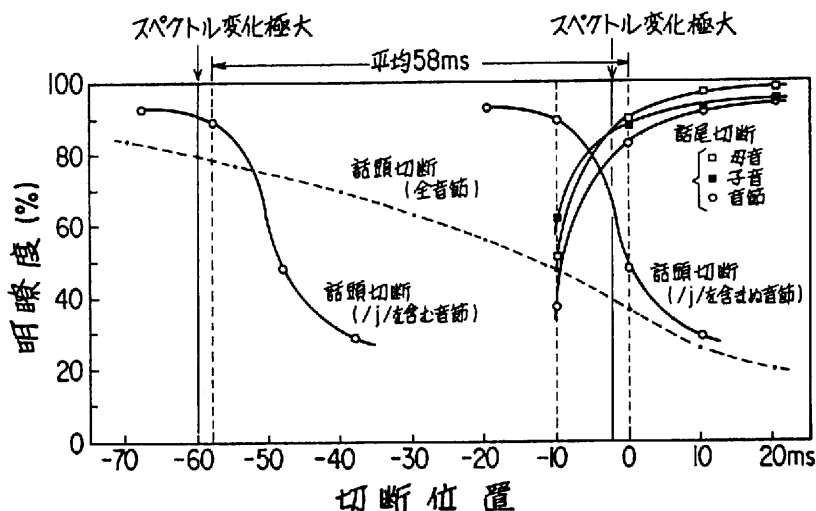


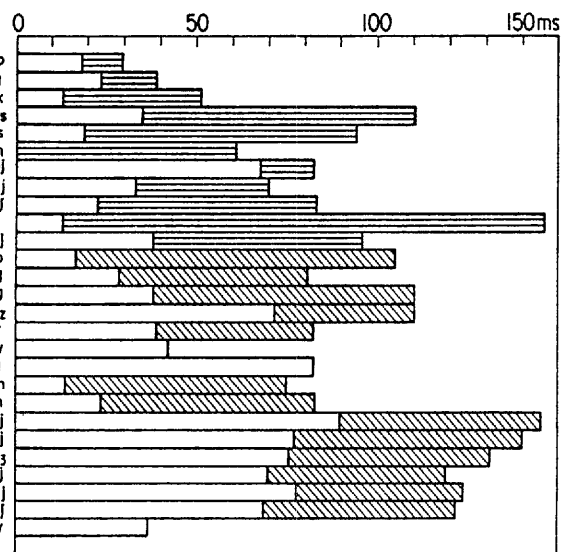
図8. 切断位置と切断音節の明瞭度の関係

6. 音節の知覚に必要な音声区間長

図9は、各音節について話頭から話尾切断の限界点までの長さを調べ、子音別に整理して示したものである。音節の先頭からどこまでの部分があれば、その音節が正しく聞きとれるかを示している。この話頭部分には、buzz-bar、buzz-barを除く有声音区間、及び無声音区間が含まれるので、視察によってこれらの境界を決定し、各区間を区別して示してある。これまでに説明してきた話頭切断の結果からわかるように、無声音区間やbuzz-barの区間は必ずしも音節の知覚に不可欠ではなく、話頭切断の限界点、即ちどこから後ろの部分が不可欠かの境界は、無声音を含む音節では平均的に子音と有声音区間の境界付近、有声音を含む音節では平均的にbuzz-barと有声音区間の境界の約10ms後ろにある。これが原因となって、この図での有声音区間の長さは、拗音を含む音節を除き、有声音を含む音節の方が無声音を含む音節よりも見かけ上10ms程度長くなっている（/p、t、k/平均18ms → /b、d、g/平均28ms）が、話頭切断の限界点と話尾切断の限界点の時間差、即ち音節の知覚に本質的に必要な有声音区間の長さには差がない。

この図からわかるその他の点としては、次のようなものがある。

- (1) /h/を含む音節の知覚には、有声音区間を必要としない。
- (2) 拗音を含む音節の知覚には、一般に拗音を含まない音節よりも長い有声音区間を必要とする。（例えば、/p j、k j、h j/平均46ms → /p、k、h/平均10ms）
- (3) /d z、j、b j、g j、d ʒ、r j、m j、n j/を含む音節の知覚には、ほとんど同じ長さの有



□ 有声音区間 ▨ 無声音区間 ▩ buzz bar

図9. 話頭から話尾切断の限界点までの長さ

声音区間を必要とし、平均で77msである。

- (4) /w/を含む音節と母音の知覚に必要な音声区間の長さは、平均40msである。この長さが、ほぼ、音節の知覚に必要な音声の最低の長さに対応していると考えられる。

7. スペクトルの動特性の視覚的検討

最後に、音声スペクトルの時間変化を視覚的にとらえやすい形式で表現し、スペクトル変化の極大点、あるいは明瞭度の限界点がどのような動的特徴を有しているかを調べた。このため、音声を多次元ケブスト

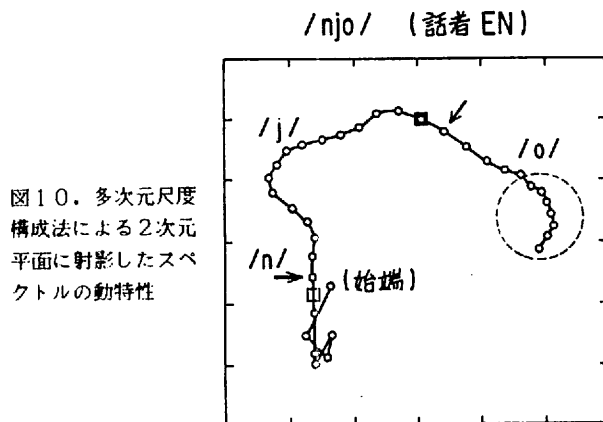
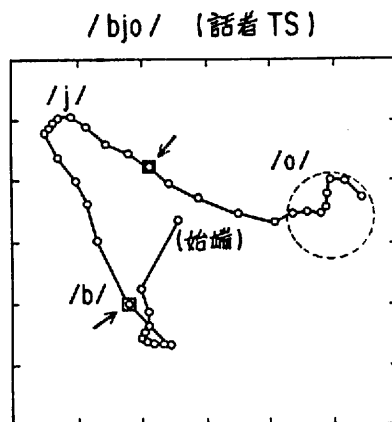


図10. 多次元尺度構成法による2次元平面に射影したスペクトルの動特性



□ 話頭切断の限界点 ● 話尾切断の限界点 ↑ スペクトル変化極大点

ラム空間内の点の動きとしてとらえ、音節の語頭から母音定常部までの区間のケプストラム時系列に多次元尺度構成法を適用して、点の動きを2次元平面に射影した。図10の左の図は話者ENの／n j o／であり、先程のスペクトル変化量と明瞭度の関係の例に用いた音声と同じである。右の図は話者TSの／b j o／であるが、両者の軸は必ずしも対応していない。2次元平面に射影したときのKruskalの適合度(ストレス)は、左の図で3.6%、右の図で3.3%であり、いずれもKruskalの言う"非常に適合している"レベルと、"よく適合している"レベルの中間にあり、スペクトルの動的特徴がよく表現されていると言える。

この図や他の各音節に関する同様の結果から、スペクトル変化極大点あるいは明瞭度の限界点は、多次元空間内でのスペクトルの動特性に関して、次の音韻(母音)ターゲットへの運動方向が定まり、その方向への変化速度が最大となった時点と解釈できることがわかった。人間はスペクトルのこのような動的特徴によって音韻を知覚していると推定される。

8. おわりに

以上、語頭/語尾切断した音節の明瞭度と、スペクトルの動的特徴の関係を実験的に調べた結果、次のことが明らかとなった。

- (1) 語頭/語尾切断音節の明瞭度は、共に切断位置がスペクトル変化極大時点の前か後かによって、急激に変化する。スペクトル変化極大時点を含む10ms程度の区間が、最も重要な音韻性情報を担っているといえる。
- (2) さらに子音と母音の明瞭度の従属関係から、子音の知覚には母音の知覚が必要であり、子音は基本的に音節を単位として、母音部へのスペクトルの遷移を手掛かりとして知覚されていると言える。
- (3) 多次元尺度構成法による2次元平面への射影の結果を含む種々の結果を総合して、多次元スペクトル空間での動的特徴、即ち音韻(母音)ターゲットに向かっての動特性が、音節(子音)の知覚に決定的役割を果たしていると言える。

スペクトル変化極大時点の前後の比較的短い区間には、音節の特徴が凝集しているということは、言い換えると、その区間の動特性から順時間方向の推定として母音を、逆時間方向の推定として子音を知覚することができることを意味しているとも言える。／s、h、j／のように定常的な雑音区間を有するものは、その部分も重要な知覚上の手掛かりの一つとなっていると考えられるが、人間の音声知覚は、特に連続音声に対しては、主としてスペクトルの動特性を手掛かりとし

て用いているのではないかと考えられる。ここで得られた結果が、母音性の強い日本語特有のものか、欧米語にも同様に成り立つものであるかは、今後さらに検討する必要がある。

今後は、音節の種類による知覚的キューの違いを含む詳細な分析や、異なる視点からの動特性の知覚実験[18]、動特性の表現法に関する研究等を進め、更に調音結合に対する知覚のメカニズムの解明へと研究を発展させていきたい。なお筆者らの別の研究により、音声の機械認識においても、スペクトルの動的特徴を静的特徴に組み合わせて用いる方法の有効性が高いことが確認されている[19,20]。

【謝辞】 日頃御指導頂く、情報通信基礎研究部第四研究室・寛室長、有益な示唆を頂いた工学院大学斎藤教授、プログラム及びデータを提供して頂いた東倉調査役、ならびに実験を担当されたクルー様他、研究室の諸氏に感謝する。

【文 献】

- [1] P. Delattre: "From Acoustic Cues to Distinctive Features", *Phonetica*, 18, pp. 198-230 (1968)
- [2] K.N. Stevens: "The Potential Role of Property Detectors in the Perception of Consonants", *Auditory Analysis and Perception of Speech*, Ed. by G. Fant and M.A.A. Tatham, Academic Press, London, pp. 304-330 (1975)
- [3] R.A. Cole, A.I. Rudnicky, V.W. Zue and D.R. Reddy: "Speech as Patterns on Paper", *Perception and Production of Fluent Speech*, Ed. by R.A. Cole, Lawrence Erlbaum Associates, Pub., Hillsdale, N.J., pp. 3-50 (1980)
- [4] 伊福部達: "FM音によるマスキング", 音響学会誌, 29, 11, pp. 679-687 (昭48)
- [5] T. Brady, A.S. House and K.N. Stevens: "Perception of Sounds Characterized by a Rapidly Changing Resonant Frequency", *J. Acoust. Soc. Amer.*, 33, 10, pp. 1357-1362 (1961)
- [6] 桑原、境: "連続音声中の母音連鎖における調音結合効果の正規化", 音響学会誌, 29, 2, pp. 91-99 (昭48)
- [7] 鈴木久喜: "母音・半母音・有声破裂音の知覚におけるホルマント遷移の変化量と変化速度との間の相補性およびその識別機構", 音響学会誌, 30, 3, pp. 169-180 (昭49)
- [8] 藤崎、樋口、二見、重野: "非定常音刺激の知覚とそのモデル", 音響学会音声研究, S81-35 (昭56)
- [9] 境、中山: "聴覚と音響心理", 日本音響学会編,

コロナ社 (昭53)

[10]三浦種敏監修:“新版・聴覚と音声”、電子通信学会 (昭55)

[11]R.Armed and R.Fatehchand:“Effect of Sample Duration on the Articulation of Sounds in Normal and Clipped Speech”.J.Acoust.Soc.Amer.,31,7,pp.1022-1029 (1959)

[12]斎藤取三:“日本語音韻の通話品質に関する基礎的研究”、名古屋大学学位論文 (昭36)

[13]W.A.Grimm:“Perception of Segments of English-Spoken Consonant-Vowel Syllables”.J.Acoust.Soc.Amer.,40,6,PP.1454-1461 (1966)

[14]桑原、境:“連続音声中の切り出し母音および音節の音韻知覚”、音響学会誌、28、5、pp.225-234 (昭47)

[15]W.Strange,J.J.Jenkins and T.L.Johnson:“Dynamic Specification of Coarticulated Vowels”,J.Acoust.Soc.Amer.,74,3,pp.695-705 (1983)

[16]古井貞熙:“音節知覚におけるスペクトルの動的特徴の役割について”、音響学会秋季講論、1-1-12 (昭59)

[17]嵯峨山、板倉:“音声の動的尺度に含まれる個人性情報”、音響学会春季講論、3-2-7 (昭54)

[18]東倉、片桐、古井:“時間軸伸縮による単音節の異聴現象”、音響学会秋季講論、1-1-11 (昭59)

[19]S.Furui:“Cepstral Analysis Technique for Automatic Speaker Verification”.IEEE Trans.Acoust.,Speech,Signal Processing,ASSP-29,2,pp.254-272 (1981)

[20]古井貞熙:“音声スペクトルの動的特徴を用いた単語音声認識”、音響学会音声研資、S84-65 (昭59)