モーラ境界・シラブル境界における音響的分離度に関する定量的分析 *

◎ 朝川 智　峯松 信明　（東大・情報理工）　広瀬 啓吉　（東大・新領域）

1　はじめに

連続的な音声波形を音素、シラブル、モーラ、単語といった音声知覚における処理単位系列へと変換するためには、音声ストリームをセグメンテーションする必要がある。そのようなセグメンテーションの問題に対して、音響的な観点から分析を行っている。既に、連続音声中の音響的イベントとしてスペクトル遷移に着目し、**CV**音節の知覚にとって最も重要な情報は、子音から母音へのスペクトル遷移が最も著しい区間に存在することが報告されている [1]。

日本語と英語では単語境界の音素連鎖の様子は大きく異なる。日本語の場合基本的に **VV**, **VC** となるが、英語の場合は、**VV**, **VC**, **CV**, **CC** の可能性がある。本研究では、スペクトル遷移量の大小を音響的分離度の大小と考え、種々の形態の境界に対する音響的分離度の定量的分析を行った。

2　モーラ境界・シラブル境界における音響的分離度の分析

2.1　音響的分離度

スペクトル遷移の大小を表すものとしてケプストラムの動的特徴量 [2] (デルタケプストラム) を用いる。時刻 t におけるデルタケプストラムベクトル $\boldsymbol{a}(t)$ を

$$\boldsymbol{a}(t) = (a_1(t), a_2(t), \cdots, a_N(t)) \tag{1}$$

としたとき、スペクトルの遷移の尺度を $D(t)$ として、

$$D(t) = \sum_{n=0}^{N} \{a_n(t)\}^2 \tag{2}$$

と定義する。但し、N はデルタケプストラムベクトルの次元数である。この $D(t)$ をスペクトル遷移度と呼ぶことにする。スペクトル遷移度の時間変化の様子の一例として、“Where’s she from?” (/werzSifrAm/) におけるスペクトル遷移の様子を**図 1** に示す。音素と音素の境界近くにおいてピークを示しているのが確認される。

このスペクトル遷移量を用いて二種類の音響的分離度を定義する。一つは $D(t)$ を瞬時的な分離度と考え、与えられた音素遷移区間におけるこの $D(t)$ のピーク値で定義する (ピーク分離度)。もう一つは、音素遷移におけるスペクトルのトータルの遷移量に着眼し、音素遷移区間 t_1 から t_2 までのスペクトルのトータル遷移量を $T(t_1, t_2)$ として

$$T(t_1, t_2) = \sum_{n=0}^{N} \left\{ \sum_{t=t_1}^{t_2} a_n(t) \right\}^2 \tag{3}$$

と定義する (トータル分離度)。以上の二種類の音響的分離度についてその様子を分析する。

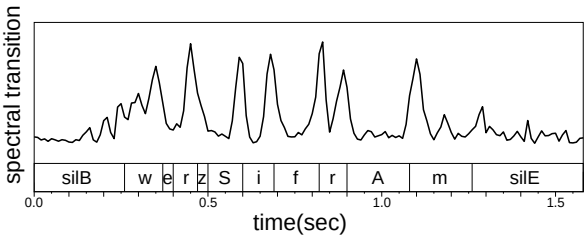

図 1. スペクトル遷移度の時間変化

2.2　モーラ境界とシラブル境界

英語音声中での単語境界においては、以下のような四種類の音素遷移が存在する。

(1)　$\mathbf{V} \rightarrow \mathbf{V}$　　(2)　$\mathbf{V} \rightarrow \mathbf{C}$

(3)　$\mathbf{C} \rightarrow \mathbf{V}$　　(4)　$\mathbf{C} \rightarrow \mathbf{C}$

このうち、(1),(2) の遷移がモーラ間の境界としても観測される遷移 (モーラ境界に相当する遷移) であり、これ以外の遷移 (3),(4) は、シラブル境界においてのみ見られる遷移である。

2.3　分析手法

米語母国語話者男性 1 名の米語音声発話に対して**表 1** に示される条件の下でデルタケプストラムパラメータを抽出した。

表 1. 分析条件

サンプリング	16bit / 16kHz
窓	窓長 25 ms, シフト長 10ms
パラメータ	ΔMFCC (0〜12 次元)
話者	米語母国語話者 男性 1 名
データ	746 文

以下の分析では、音声発声中の全ての異音素間境界のうち、単語間の音素遷移のみを対象として、(1)〜(4) のカテゴリに分類し分析を行うこととする。**表 2** に全音素セットを示す。このうち、**表 3** に示される有声音・無声音の組は同一音素と見なして、**CC**遷移とは扱わないこととした。さらに、(1),(2),(3) の遷移において、**V**に二重母音 (o, e, Y, W, O) が含まれるような遷移については分析対象から除外した。

図 2 に示されるように、分析対象となる音素遷移に対して、前音素の中心フレームから後音素の中心フレームまでを音素遷移区間とし、この遷移区間に

∗ Quantitative analysis of the degree of acoustic separation on mora and syllable boundaries
By Satoshi Asakawa, Nobuaki Minematsu and Keikichi Hirose (University of Tokyo)

表 2. 音素セット

母音	i,I,e,E,@,a,c,o,U,u,Y,O,W,R,A,x
子音	b,p,d,t,g,k,J,C,v,f,D,T,z,s,Z,S,m,n,G,l,r,w,y,h

表 3. 同一音素と見なす有声音・無声音の組

有声音	b	d	g	J	v	D	z	Z
無声音	p	t	k	C	f	T	s	S

おけるピーク分離度とトータル分離度を、当該音素遷移の音響的分離度として算出する。

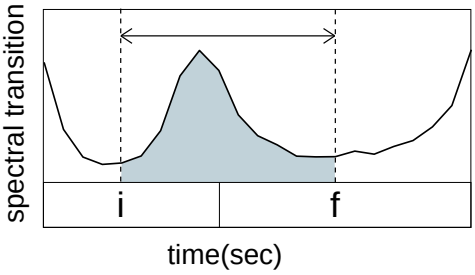

図 2. 音素遷移区間

各音素遷移に対して得られた音響的分離度を**第 2.2 節**の (1) 〜 (4) の遷移に分類した上で、これらの中から任意の二つの遷移を選んで分散分析を行い、二種類の遷移の間に有意な差が見られるのかを調べる。すべての遷移の組み合わせに対して分散分析を行い、有意差の有無を調べることで、モーラ境界とそれ以外の境界との音響的分離度の違いを分析した。

2.4 分析結果

表 4 に各遷移における分離度の平均値を示す。また分散分析の結果を**表 5**, **表 6** に示す。以上の結果より、ピーク分離度を見た場合、以下のような大小関係があることがわかる。

(2)**VC** ＞ (3)**CV** ＞ (4)**CC** ≳ (1)**VV**

同様にして、トータル分離度については、

(1)**VV** ≳ (2)**VC** ＞ (3)**CV** ＞ (4)**CC**

という大小関係になった。ピーク分離度を見た場合には、**VC**遷移、また、トータル分離度について見た場合には、**VV**遷移と **VC**遷移において、音響的分離度がより大きいことが確認できる。即ち、これらの遷移において音響的により明確な境界があると考察される。

以上の結果より、モーラ境界に相当する遷移である **VV**遷移と **VC**遷移では、それ以外の遷移に比べて、より音響的な分離度が大きいことが確認できた。**VV** 遷移については、ピーク分離度の場合では最も小さくなってるが、ほかの遷移に比べて時間的に遷移区間が長くなり比較的緩やかにスペクトルが変化するためであると考えられる。スペクトルの急峻な変化はみられないものの、トータルのスペクトルの遷移量

表 4. 各遷移における分離度の平均値

	ピーク分離度	トータル分離度
(1) **VV**	8.44×10^1	$\mathbf{1.27 \times 10^3}$
(2) **VC**	$\mathbf{1.03 \times 10^2}$	1.16×10^3
(3) **CV**	9.57×10^1	9.85×10^2
(4) **CC**	8.94×10^1	8.37×10^2

表 5. 分散分析による有意水準 (ピーク分離度)

	(1)**VV**	(2)**VC**	(3)**CV**	(4)**CC**
(1)**VV**	-	<1%	<1%	**>10%**
(2)**VC**	-	-	<1%	<1%
(3)**CV**	-	-	-	<1%
(4)**CC**	-	-	-	-

表 6. 分散分析による有意水準 (トータル分離度)

	(1)**VV**	(2)**VC**	(3)**CV**	(4)**CC**
(1)**VV**	-	**>5%**	<1%	<1%
(2)**VC**	-	-	<1%	<1%
(3)**CV**	-	-	-	<1%
(4)**CC**	-	-	-	-

が大きいために音響的にセグメンテーションしやすいと考えられる。

この結果から、モーラ言語である日本語は、シラブル言語である英語に比較すると音響的にセグメンテーションしやすい言語であるともいえる。一般によく言われている、「日本語を母国語とする日本人は英語の聞き取りが苦手である」ということに対する音響的観点からの一つの定量的説明であると考えている。

3 まとめ

音声ストリームをセグメンテーションする音響的イベントとしてケプストラムの動的特徴量に着目し、英語音声中のシラブル境界において、モーラ境界に相当する遷移である **VV**遷移及び **VC**遷移と、それ以外の遷移について、その音響的分離度の様子に分析した。モーラ境界に相当する遷移はそれ以外の遷移に比べて音響的分離度がより大きいことが確認できた。

今回は一人の話者でのみの分析であったので、今後は複数の話者に対して同様の分析を行うことが必要であると思われる。また、子音を弁別素性などにより分類してより詳細な分析を行うことや、非母国語話者の英語発話に対して同様の分析を行い、母国語話者との比較を行うことなども挙げられる。

参 考 文 献

[1] S.Furui: “On the role of spectral transition in speech perception”, The Journal of the Acoustical Society of America, Vol.80, No.4, pp.1016-1025, 1986

[2] 嵯峨山他: “音声の動的尺度に含まれる個人性情報”, 日本音響学会講演論文集, 3-2-7, pp.589-590, 1979