

TR-H-053

0009

英語 /r/-/l/ 非類似度計測における
各種のスペクトル表現の比較

相川 清明
山田 玲子

1994. 2. 15

ATR 人間情報通信研究所

〒619-02 京都府相楽郡精華町光台 2-2 ☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011

Facsimile: +81-7749-5-1008

英語 /r/-/l/ 非類似度計測における各種のスペクトル表現の比較
Comparative Study of Spectral Representations in Measuring the English
/r/-/l/ Acoustic-Perceptual Dissimilarity

相川 清明 山田 玲子

Kiyoaki Aikawa and Reiko A. Yamada

ATR 人間情報通信研究所

京都府相楽郡精華町光台 2-2

ATR Human Information Processing Research Laboratories

2-2 Hikaridai, Seikacho, Sorakugun, Kyoto 619-02 Japan

Email: aik@hip.atr.co.jp and yamada@hip.atr.co.jp

あらまし

日本人被験者による英語 /r/-/l/ 知覚実験における正答率は発話者に依存する。本報告はこの発話者依存性を /r/-/l/ 対の音響的非類似度で説明する試みについて述べる。/r/-/l/ 対の音響的非類似度を動的ケプストラムを用いて計測すると知覚実験による正答率ときわめて相関の高い発話者依存性パターンを示すことがわかった。基本パラメータとしてケプストラム、重み付きケプストラム、メルケプストラムの3種類を用い、これらのパラメータ自身、及び、これより導かれる動的ケプストラム、デルタケプストラム、組合せにして9種類のスペクトル表現の非類似度計測における有効性を比較する。知覚実験としては訓練を伴う場合と伴わない場合の2通りについて行った。部分的音響的非類似性分析により、日本人は /r/, /l/ 直後 250 ms 程度の区間の調音結合により変形した後続音素を利用して /r/-/l/ を聞き分けている可能性があることが示唆された。

Abstract

The talker dependency of the correct response rate on the English /r/-/l/ perception by Japanese listeners has been reported. This paper reports that the acoustical dissimilarity of an /r/-/l/ minimal pair measured by a dynamic cepstrum shows a talker dependency pattern closely correlated with the talker dependency pattern of the correct response rate. Nine spectral representations are compared in measuring the /r/-/l/ acoustical dissimilarity. The nine spectral representations include three base parameters, a cepstrum, a weighted cepstrum, and a mel-cepstrum, and dynamic cepstra and delta cepstra derived from the three base parameters. The experimental results imply that Japanese listeners tend to identify /r/ or /l/ using the succeeding vowels affected by co-articulations.

1 はじめに

日本人による英語 /r/-/l/ 音の同定実験における正答率は発話者に依存することが報告されている [1, 2]。すなわち、発話者により /r/,/l/ の同定しやすさが異なる。しかし、この発話者依存性が何に起因しているのかという説明はなされていない。/r/ と /l/ の正答率の発話者による違いがそれらの物理的な特徴に起因するとすれば、適当な音響特徴量を用いることによりこの現象を説明できると考えられる。ここでは、/r/ と /l/ の音響特徴量の違いが大きい発話者ほど /r/,/l/ を同定しやすいという仮説を設ける音響特徴の表現は人間の聴覚特性にできるだけ近いことが望ましい。すなわち、聴覚系におけるスペクトルの内部表現を模擬している必要がある。

動的ケプストラムは順向マスキングの時間周波数特性を模擬したスペクトル特徴パラメータで、ホルマント遷移などの動的成分を強調したスペクトルを表現する。このパラメータは、対数スペクトルの逆フーリエ変換である従来のケプストラムに比べ優れた音声認識性能を与えるため [3]、ケプストラムよりも聴覚系に近い音響的特徴を抽出している可能性がある。

そこで、英語 /r/-/l/ 知覚実験に見られる正答率の発話者依存性を動的ケプストラムを用いて分析することを試みた [4]。本報告では、各種のスペクトル表現を用いて求めた /r/-/l/ の音響的非類似度の発話者依存性と知覚実験における正答率の発話者依存性の相関を調べ、動的ケプストラムを用いた場合のそれと比較する。

2 基本パラメータ

まず、対数スペクトルの逆フーリエ変換であるケプストラムから導かれる基本スペクトルパラメータを求める。いずれの基本スペクトルパラメータもケプストラム領域のパラメータとして表され、動的ケプストラムなどの非類似度計測に用いられるスペクトル表現は、これらの基本パラメータから導かれる。

2.1 重み付きケプストラム

重み付ケプストラムはケプストラムの係数ごとの分散を正規化する目的で提案された [5]。 k 次のケプストラム係数を c_k とすると、重み付ケプストラム g_k は近似的に

$$g_k = \begin{cases} k c_k & k < k_0 \\ k_0 c_k & k \geq k_0 \end{cases} \quad (1)$$

により求められる。 k_0 は過度の高次成分強調を避けるための飽和ケフレンシである。重み付きケプストラムはピークを強調した対数スペクトルの逆フーリエ変換にあたる。

2.2 メルケプストラム

低域が引き延ばされた聴覚的な周波数軸でのスペクトルはケプストラム係数に対する周波数軸変換により求めることができる。ここでは、双一次変換の位相特性を用いてケプストラムからメルケプストラムを求める方法を用いる [6]。この方法によれば、帯域が 5 kHz 前後であれば周波数歪み率を適当に与えることにより、メルスケールに近い周波数軸における対数スペクトルの逆フーリエ展開係数をメルケプストラムとして求めることができる。

3 スペクトル表現

次に、前述の基本パラメータを基にして各種のスペクトル表現を求める。

3.1 動的ケプストラム

動的ケプストラムは過去の音声による順向マスクングを受けたスペクトルの逆フーリエ変換により定義される。順向マスクングにおいては、マスクングパターンが時間の経過と共に周波数軸上で平滑化される性質があり、動的ケプストラムはこの時間周波数特性を模擬している。マスクされたスペクトルは現時点のスペクトルからマスクングパターンを減算することにより求められ、マスクングパターンは周波数平滑化された過去のスペクトルから求められる。スペクトルの平滑化はスペクトルに対して平滑化フィルタとのコンヴォリューションにより行うことができるが、対数スペクトルを逆フーリエ変換したケプストラム領域ではリフタの乗算により等価的に行うことができる。従って、時点 i における k 次のケプストラム係数を $c_k(i)$ 、マスクングパターンのケプストラム展開係数を $m_k(i)$ とすると、動的ケプストラム $b_k(i)$ は、

$$b_k(i) = c_k(i) - m_k(i) \quad (2)$$

$$m_k(i) = \sum_{n=1}^N c_k(i-n)l_k(n) \quad (3)$$

により求められる。 $l_k(n)$ は n 時点前のスペクトルを平滑化するリフタの k 次の係数である。 N はマスクングの影響の及ぶ最大時間長を表している。スペクトルを平滑化するために以下のようなガウス型リフタを用いると良好な音声認識率が得られる。

$$l_k(n) = \alpha\beta^{n-1} \exp\left(-\frac{k^2}{2(q_0 - \nu(n-1))^2}\right) \quad (4)$$

ここで、 q_0 は 1 時点前の平滑化リフタのガウス分布の標準偏差、 ν は標準偏差が時間をさかのぼるに従って小さくなる速度を表している。 β はマスクング減衰率、 α はその初期減衰率である。

3.2 デルタケプストラム

動的特徴を捉えるパラメータとしてデルタケプストラムが音声認識に広く用いられてきた [7, 8]。デルタケプストラムはケプストラム時系列の 1 次回帰係数として定義される。時点 i における k 次のデルタケプストラム $\Delta c_k(i)$ は

$$\Delta c_k(i) = \frac{\sum_{n=-L/2}^{L/2} w(n)c_k(i+n)n}{\sum_{n=-L/2}^{L/2} w(n)n^2} \quad (5)$$

により求められる。ここで、 $w(n)$ は長さ L の時間窓で二等辺三角形形状の窓が良く用いられる。デルタケプストラムはスペクトルのスペクトル時系列の各時点における時間的な傾斜を表し、スペクトル形状の情報を持たないので、通常スペクトル形状を表すケプストラムと併用する。ケプストラム距離と併用して用いる場合には

$$d = (1 - \lambda) \sum_k (c_k - c_k^*)^2 + \lambda \sum_k (\Delta c_k - \Delta c_k^*)^2 \quad (6)$$

を用いる。ここで、 c_k^* 、 Δc_k^* は比較の対象となるケプストラム、デルタケプストラムの各係数、 λ は結合係数である。

4 音響的非類似度計測法

/r/-/l/ 単語対の音響的距離を図 1 のように動的時間伸縮法 (DTW: Dynamic Time Warping) により求める [9]。DTW による整合パスに沿ってマッチング距離を平均したものを単語対の音響的非類似度とする。DTW アルゴリズムは疑似両端点フリー対称型で、端点領域は単語マッチング時には 10 フレー

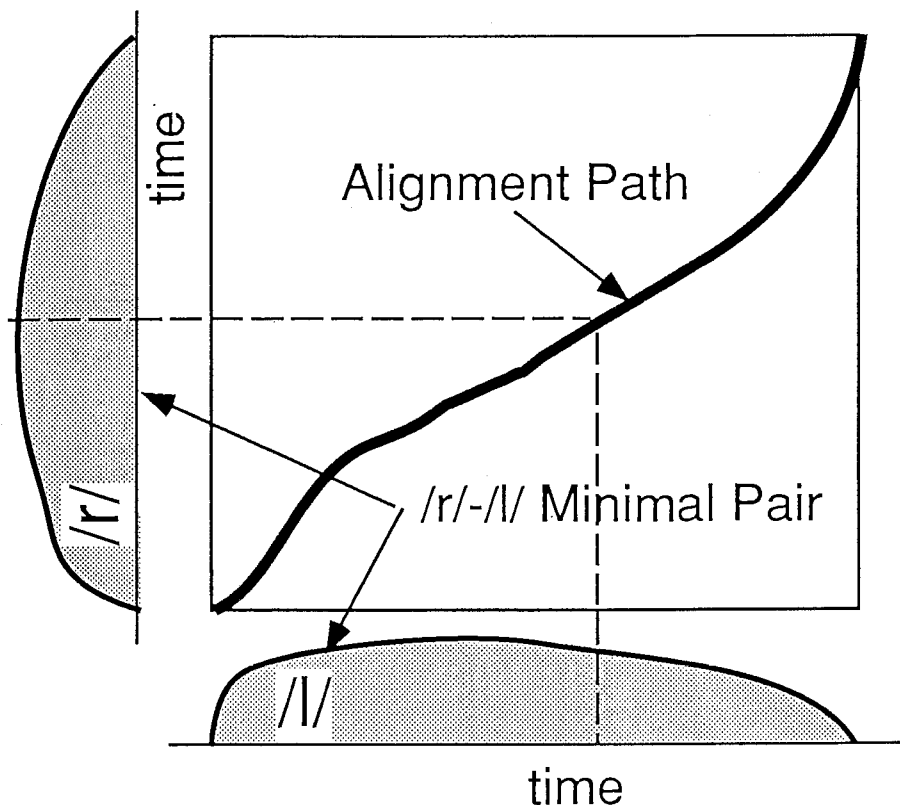


図 1: A Dynamic Time Warping algorithm for measuring the acoustical dissimilarity of a minimal pair.

ム、音素等局部マッチング時には 3 フレームとした。フレーム周期は 10 ms である。局所的パス制限は最大傾き 2 最小傾き 1/2 とした。

両音声の整合時点の非類似度は対応するフレームのスペクトルを表す特徴量ベクトル間のユークリッド距離により求める。特徴ベクトルの種類は、3 種の基本パラメータ、ケプストラム、重み付きケプストラム、メルケプストラムに対して 3 種類のスペクトル表現：無変換、動的ケプストラム、デルタケプストラムを求めているので 9 種類である。この計測法では同一条件で収録した同一発話者の単語音声間の距離を用いるので収録系や発話者固有の平均的スペクトル形状の影響は受けにくいと考えられる。

5 発話者依存性パターン間距離尺度

知覚実験における正答率の発話者依存性パターンは発話者ごとに平均した各被験者の /r/-/l/ 対の正答率で表される。一方、音響的非類似度の発話者依存性パターンは発話者ごとに平均した各 /r/-/l/ 対の音響的非類似度で表される。これらは単位が異なる量であるため、これらのパターンの間の違いを計測するためにパターン間の相関を用いることにする。正答率の発話者依存性パターンを $\vec{x} = (x_1, \dots, x_I)$ 、音響的非類似度の発話者依存性パターンを $\vec{y} = (y_1, \dots, y_I)$ で表す。I は発話者数にあたる。データの組 $(x_i, y_i), i = 1, 2, \dots, I$ の相関は

$$\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (7)$$

で表される。これは 2 つのベクトル \vec{x} 、 \vec{y} に関する

$$\frac{(\vec{x} - \bar{x}, \vec{y} - \bar{y})}{\|\vec{x} - \bar{x}\| \|\vec{y} - \bar{y}\|} \quad (8)$$

表 1: Number of minimal pairs in each phonetic context.

Category	Phonetic Context	Exp.1	Exp.2
IP	r/l ...	21	13
IC	C r/l V	32	24
IN	.. V r/l V ..	6	5
FP	.. V r/l	19	15
FC	.. C r/l	15	11

と等価であり、平均値を 0 に正規化したベクトル $\vec{x} - \bar{x}$ と $\vec{y} - \bar{y}$ の余弦を表す。

6 知覚実験

6.1 音声資料

/r/-/l/ を同じ位置に含み他の部分が同一である英単語対 (ミニマルペア) 93 種類を資料として用いた (表 2)。これは Logan らが訓練に用いた 68 単語対に試験に用いている 25 対を加えたものである [1]。これらを 5 人の米語を母国語とする話者 (男性 3 名、女性 2 名) が発話した音声 (文献 [1, 2, 4] と同じデータベース) を刺激音として用いた。発話者 1, 3, 5 が男性で発話者 2, 4 が女性である。被験者には刺激音がヘッドホンより呈示され、その刺激を含む単語対 (例えば "red" の場合 "red" と "led") から成る反応選択肢のうちひとつを強制選択することが課せられた。

本報告では 2 種類の知覚実験について正答率と音響的非類似度の関係の分析を行った。第 1 の実験は 93 単語対を用いた訓練を行わない状態での知覚実験である。第 2 の実験は Logan が訓練に用いた 68 単語対を用いた訓練効果を調べる実験である。単語対は /r/, /l/ の音素文脈により 5 種類のカテゴリーに分類できる [1]。各カテゴリーに含まれる単語対の数を表 1 に示す。

6.2 実験 1 : 同定実験

第 1 の実験では訓練を行わない同定実験における正答率の発話者依存性を調べた。5 名の発話者の 93 単語対 (930 単語) をランダムな順序で呈示して同定試験を行い、結果を発話者ごとに平均した。呈示順序は被験者ごとに異なる。被験者は海外滞在経験を持たない日本人成人 11 名である。すべてのカテゴリーの単語対の正答率を発話者ごとに平均して求めた発話者依存性パターンを図 2 に示す。単語対のうち、/r/-/l/ 以外の部分の音素の一致が良いカテゴリー IP IC IN のみについて正答率を平均したものを図 3 に示す。この図から発話者 4 の /r/, /l/ が日本人にとって聞き取りやすいことがわかる。

6.3 実験 2 : 同定訓練実験

実験 2 では、各単語同定試験毎に解答の正誤を示す訓練実験を行った。この実験は本来、訓練効果を調べるための実験であり、正答率は一連の実験中に徐々に向上する [2]。各セッションは 1 発話者の発話した 68 単語対 (136 単語) をそれぞれ 2 回ずつ含む 272 試行からなる。単語はランダムな順序で呈示され、被験者は /r/-/l/ 単語対のどちらの単語かを答える。発話者 1 から 5 までをこの順序で 9 回繰り返す (総計 45 セッション)。ここでは発話者 1 から 5 までの一連のセッションをセッションブロックと呼ぶことにする。セッション内での刺激の呈示順序は被験者、セッションごとに異なる。被験者は海外滞在経験を持たない日本人成人 13 名である。実験 1、2 は同じ音声データベース中の音声資料を用いているが、実験 2 の被験者は実験 1 とは異なる。

すべてのセッションにおける平均正答率の変化を図示したものが図4である。また、この学習により正答率が上昇する過程を発話者ごとに示したものが図5である。この図にはブロック1、3、6、9のみを示した。

このように学習が進むにつれて正答率が変化してしまうので、このままでは最初に聞いた発話者1に比べ、セッションブロックの最後に聞く発話者5の正答率は高くなる傾向を持つことになる。/r/-/l/の音響的非類似度との対応を取るためには、学習効果を除去した正答率を求める必要がある。セッション*i*の正答率を u_i とすると誤り率の対数

$$\ln(1 - u_i) \quad (9)$$

は図6に示すように各発話者毎にほぼ直線に乗ることがわかる。従って、同定誤りは指数関数的に減衰すると言える。この平均的な傾きを除去した対数誤り率を v_i とする。全45セッションの中で同一発話者は周期的に現れるので、発話者 j に対する未学習時の推定正答率 x_j は

$$v_i = \ln(1 - u_i) + \kappa i \quad (10)$$

$$x_j = 1 - \exp\left(\frac{1}{9} \sum_{k=1}^9 v_{((k-1) \times 5 + j)}\right) \quad (11)$$

となる。 κ は1セッションごとの正答率の向上を表す係数で、すべての音素文脈を用いた場合0.02623、IP IC INに限った場合は0.02412であった。 k はセッションブロックの番号である。IP IC INに限った場合の誤りの減衰率は各セッションにつき平均0.976で、5名の発話者を1巡すると誤りは平均0.886に減少する。学習効果を除去した推定未学習時正答率を図7に示す。実験2ではこれを発話者依存性パターンとする。この正答率の発話者依存性パターンは実験1での図3と良く一致している。なお、この図は音素文脈をIP IC INに限定した場合であるがすべての音素文脈を用いる場合も図8に示すようにほぼ同じパターンとなる。

7 音響的非類似度分析

刺激音声の/r/-/l/音響的非類似度を求めるため、すべての刺激音声のスペクトル表現を求めた。まず、基本パラメータを求めるために10ms毎に16次の線形予測分析により32次のLPCケプストラム係数を求めた。このケプストラム時系列から2種類の基本パラメータ、重み付きケプストラム、およびメルケプストラムを求めた。重み付きケプストラムの飽和ケブレンシは5とした。メルケプストラムについては、帯域5kHzの場合、双一次変換による周波数変換係数が0.5のときにBarkスケールに近くなるためこの値を用いて周波数軸変換を行った。ここでは慣例に従ってこのパラメータをメルケプストラムと呼ぶことにする。

次に基本パラメータ時系列から2種のスペクトルパラメータ、動的ケプストラム、およびデルタケプストラムを求めた。動的ケプストラムのリフタのパラメータは音声認識において最も良い結果が得られている組合せ、 $N = 4$ 、 $q_0 = 18$ 、 $\nu = 1$ 、 $\alpha = 0.3$ 、 $\beta = 0.7$ 、を用いた。デルタケプストラムは70msの二等辺三角形の時間窓を用いて求めた。ケプストラム距離に対するデルタケプストラム距離の重みは $\lambda = 0.95$ とした。

以上の9種類のパラメータについて/r/-/l/単語対の音響的非類似度をDTWを用いて計測した。この音響的非類似度の発話者依存性パターンを実験1、2で用いた音声のセットそれぞれに対して求めた。

知覚実験に用いたすべての刺激音声にはあらかじめ次の部位、および音がわかるようにラベルが付与された。単語の先頭、末端、/r/または/l/の部分とその前後の母音または子音部。/r/または/l/と前後の母音の境界は遷移部(わたり)の中央部付近とした。以下、/r/,/l/部位は/r/,/l/のラベルが付与された部分を指すこととする。

8 正答率と音響的非類似度の関係

8.1 実験1の分析

まず5種のカテゴリすべてを用い、知覚実験による正答率の発話者依存性パターンと、/r/,/l/ 単語対の単語全区間での音響的非類似度の発話者依存性パターンの類似度を両者の相関係数により測定した。実験1で用いたすべての単語対の音響的非類似度の発話者依存性パターンをケブストラムを用いて求めたものが図9、動的ケブストラムを用いて求めたものが図10である。両者を音素文脈 IP IC IN について求めたものが図11（ケブストラムの場合）と図12（動的ケブストラムの場合）である。これらの結果からスペクトル表現として動的ケブストラムを用いた場合の音響的非類似度（図10）は知覚実験における正答率（図2）と高い相関があることが分かる。これに対して、ケブストラムによる音響的非類似度は正答率との相関は弱いことがわかる。すべての単語対を用い、スペクトル表現として通常のケブストラムを用いる場合の正答率と音響的非類似度の関係を図13に示す。また、動的ケブストラムを用いる場合の正答率と音響的非類似度の関係を図14に示す。

以上が単語全体での非類似度と正答率の関係を分析した結果である。次に、/r/,/l/ 部分以外が良く一致しているカテゴリ IP, IC, IN について単語のどの部分の音響的非類似度が知覚実験による正答率と関係が深いかを調べた。非類似度計測を行った部分は単語全体、/r/,/l/、/r/,/l/ + 後続母音、後続母音、後続母音へのわたり（/r/,/l/ 中央から後続母音中央まで）である。これらの計測部分を図示したものが図15である。/r/,/l/ 部分の指定は音素ラベルによる。各部の整合パスを求める際の位置的な制約は端点のみである。すなわち、ラベル上での音素境界は、端点として与えられたもの以外は単語対の整合パスの上で一致している保証はない。後続母音の種類は単語により異なる。また、発話速度が発話者毎に異なっていることも加わって、継続時間もまちまちである。そこで、/r/,/l/ 直後の一定部分を後続母音と仮定して非類似度計測を行うことにした。後続母音区間の指定には2種類の方法を用いた。第1の方法は /r/,/l/ 直後の一定の時間長を後続母音とする。第2の方法では /r/-/l/ から語尾までの長さの一定割合を後続母音とする。第1の方法では発話速度が一定であれば単語によって語尾までの音素数が異なっても指定した区間に一定数の音素が含まれると考えられる。第2の方法では後続する音素数が一定であれば発話速度に依らず指定した区間に一定数の音素が含まれると考えられる。

各部の音響的非類似度と正答率の相関を調べた結果を基本パラメータがケブストラムの場合を表3に、重み付ケブストラムの場合を表4に、メルケブストラムの場合を表5に示す。このうち基本パラメータがケブストラムで、後続母音区間長を 250 ms に指定した場合の結果を図16に示す。ケブストラムと動的ケブストラムを比較している。この図から後続母音部分を含む部分について動的ケブストラムにより計測された音響的非類似度が正答率と高い相関を持つことがわかる。特に /r/,/l/ + 後続母音 の部分において高い相関を示している。

次に9種類のスペクトル表現の非類似度計測における有効性比較結果を示す。非類似度計測を行った部分は /r/,/l/ + 後続母音の部分で、非類似度と正答率の相関を幾つかの後続母音区間長について調べた。ケブストラムを基本パラメータとする場合を結果を図17に、重み付きケブストラムを基本パラメータとする場合を結果を図18に、メルケブストラムを基本パラメータとする場合を結果を図19に示す。これより、音響的非類似度と正答率の相関において、ケブストラムとデルタケブストラムの組合せはケブストラムより良く、動的ケブストラムはこれよりさらに良い結果を示すことがわかる。基本パラメータを比較してみると、ケブストラムとメルケブストラムでは大きな差は認められなかった。重み付きケブストラムは母音区間長に対する感度が高く、部位を限定すると高い相関が得られる。また、これらの図から /r/,/l/ を含む 250 ms 程度の比較的長い区間の音響的非類似度と正答率の相関が強いことが分かる。

また、後続母音区間を割合で与える場合の /r/,/l/ と 後続母音の区間の音響的非類似度と正答率の関係を調べた結果を図20に示す。スペクトル表現としては動的ケブストラムを用いている。この結果も、比較的長い母音区間の音響的非類似度が正答率と相関が高いことを示している。なお、/r/,/l/ と 後続

母音 250 ms を含む区間の動的ケプストラムを用いて計測した音響的非類似度と正答率の相関における各発話者の分布を図 21 に示す。

これらの結果は、同定しやすい発話者の /r/, /l/ 単語対は長い区間にわたってスペクトルに違いがあることを示している。以上の結果から、日本人は /r/, /l/ の局所的な情報のみではなく調音結合により変化した後続母音を手がかりにして /r/, /l/ を同定している可能性があることが示唆される。

8.2 実験 2 の分析

音素文脈が IP IC IN である音声データについて知覚実験による正答率の発話者依存性パターンと、/r/, /l/ 単語対各部の音響的非類似度の発話者依存性パターンとの相関を求めた。ただし、この分析では正答率の発話者依存性パターンは訓練による学習の影響を補整した推定値である。各部の音響的非類似度と正答率の相関を調べた結果を基本パラメータがケプストラムの場合を表 6 に、重み付ケプストラムの場合を表 7 に、メルケプストラムの場合を表 8 に示す。母音区間長を変えて求めた /r/, /l/ と後続母音の区間の音響的非類似度と推定正答率との相関を図 22 に示す。基本パラメータはケプストラムで、これから導かれるケプストラム、デルタケプストラム、動的ケプストラムを比較している。この場合も実験 1 とほぼ同じ分析結果が得られた。

9 むすび

各種のスペクトル表現を用いて英語 /r/-/l/ 対の音響的非類似度を計測し、知覚実験における正答率との相関を調べた。この結果、動的ケプストラムにより計測した音響的非類似度は日本人被験者による /r/-/l/ 知覚実験における正答率の発話者依存性パターンときわめて高い相関を示した。また、デルタケプストラムを用いた場合に、ケプストラムにくらべて良い音響的非類似度と正答率の対応が得られることから、スペクトルの動的特徴が知覚実験結果の裏付けに重要であると言える。部分的な非類似度計測実験により、日本人は /r/, /l/ のみでなく調音結合により変形した後続母音の情報を用いてこれらの音素を同定している可能性があることが示唆された。これらの実験を通して、動的ケプストラムが音声認識のみでなく、聴知覚実験結果の音響分析にも有効である可能性が示された。

謝辞 音声データを提供して頂いたインディアナ大学 D.B. Pisoni 博士、音素ラベリングツールを提供して頂いた旧 ATR 自動翻訳電話研究所杉山雅英博士（現在会津大学教授）に感謝する。また、本研究を進めるにあたって貴重な助言を頂いた当研究所社長東倉洋一博士に感謝する。

参考文献

- [1] John S. Logan, Scott E. Lively, and David B. Pisoni. Training Japanese listeners to identify English /r/ and /l/: A first report. *J. of Acoust. Soc. of Am.*, 89(2):874-886, 1991.
- [2] 山田玲子, 東倉洋一, Scott E. Lively, David B. Pisoni, 山田恒夫. 英語 /r//l/ 音の聴取訓練. 日本音響学会講演論文集, I:315-316, 1993-03.
- [3] Kiyooki Aikawa, Harald Singer, Hideki Kawahara, and Yoh'ichi Tohkura. A dynamic cepstrum incorporating time-frequency masking and its application to continuous speech recognition. *Proc. ICASSP93*, II:668-671, 1993.
- [4] 相川清明, 山田玲子. 英語 /r/-/l/ 非類似性計測における動的ケプストラムの有効性. 日本音響学会講演論文集, I:609-610, 1993.
- [5] Yoh'ichi Tohkura. A weighted cepstral distance measure for speech recognition. *IEEE Trans.*, ASSP-35(10):1414-1422, 1987-10.
- [6] Alan V. Oppenheim and Donald H. Johnson. Discrete representation of signals. *Proc. of the IEEE*, 60(6):681-691, 1972.

- [7] Shigeki Sagayama and Fumitada Itakura. On individuality in a dynamic measure of speech. *Acoust. Soc. Jpn. meeting*, pages 589–590, 1979-06.
- [8] Sadaoki Furui. Speaker-independent isolated word recognition using dynamic features of speech spectrum. *IEEE Trans.*, ASSP-34(1):52–59, 1986-02.
- [9] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans.*, ASSP-26(1):43–49, 1978.

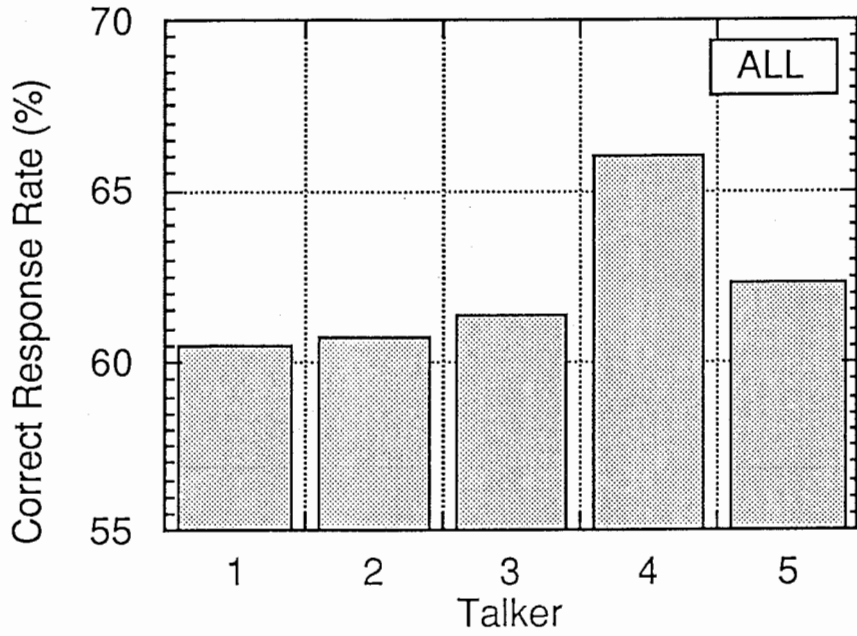


Figure 2: The talker dependency pattern on the English /r-/l/ identification (Exp.1).

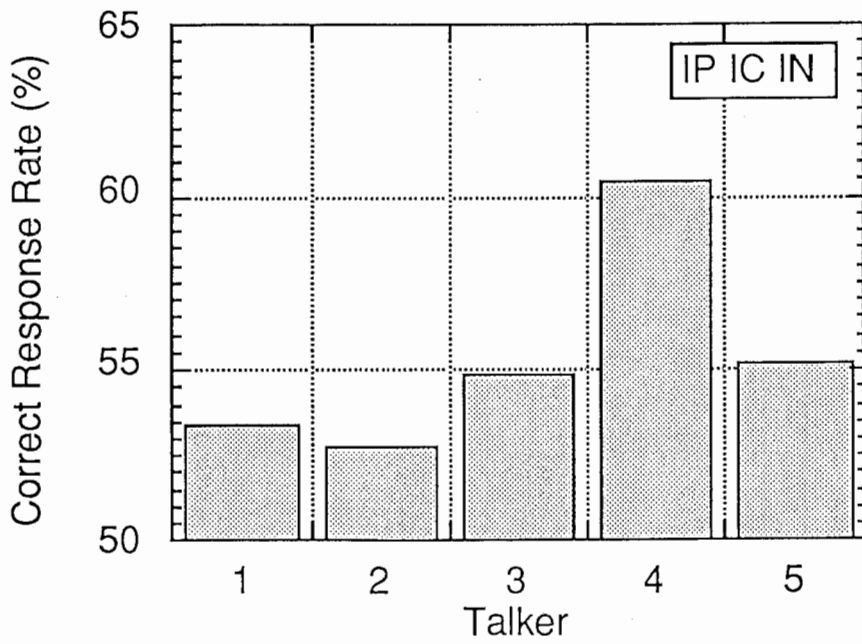


Figure 3: The talker dependency pattern on the English /r-/l/ identification (Exp.1).

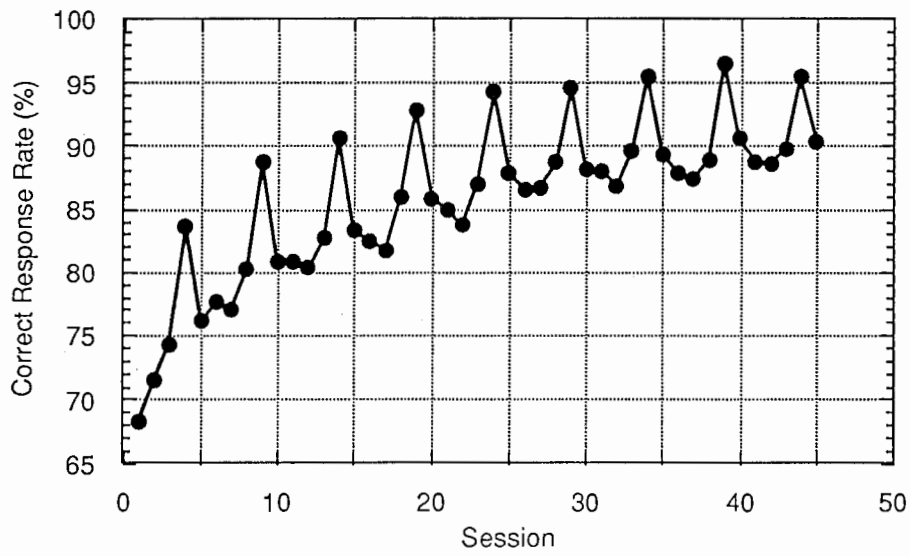


图 4: The change of correct response rate over a series of training sessions (Exp.2).

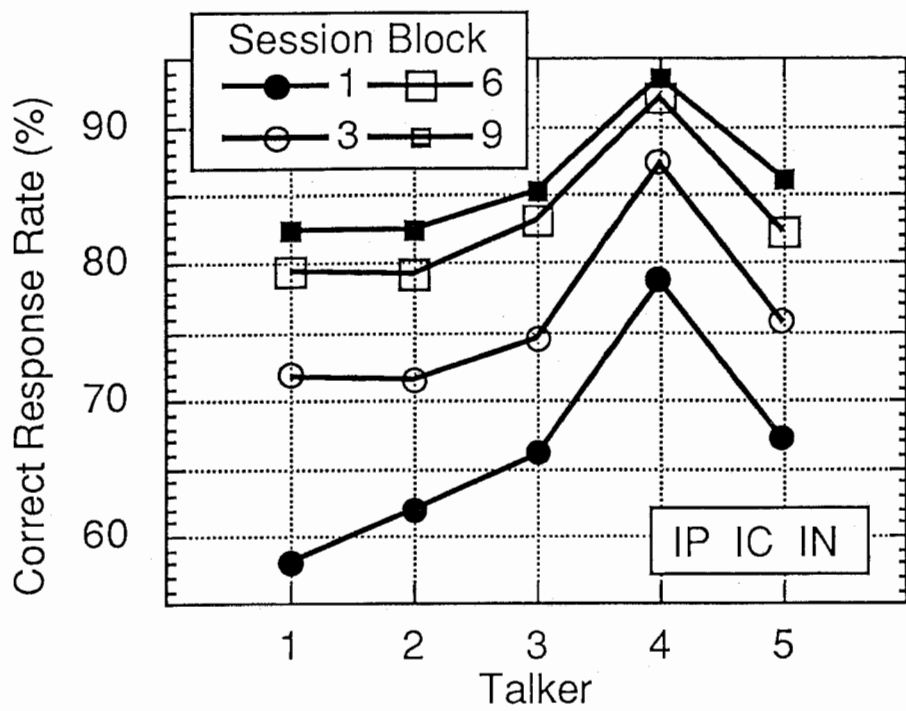


Figure 5: The change of correct response rate for each session block (Exp.2).

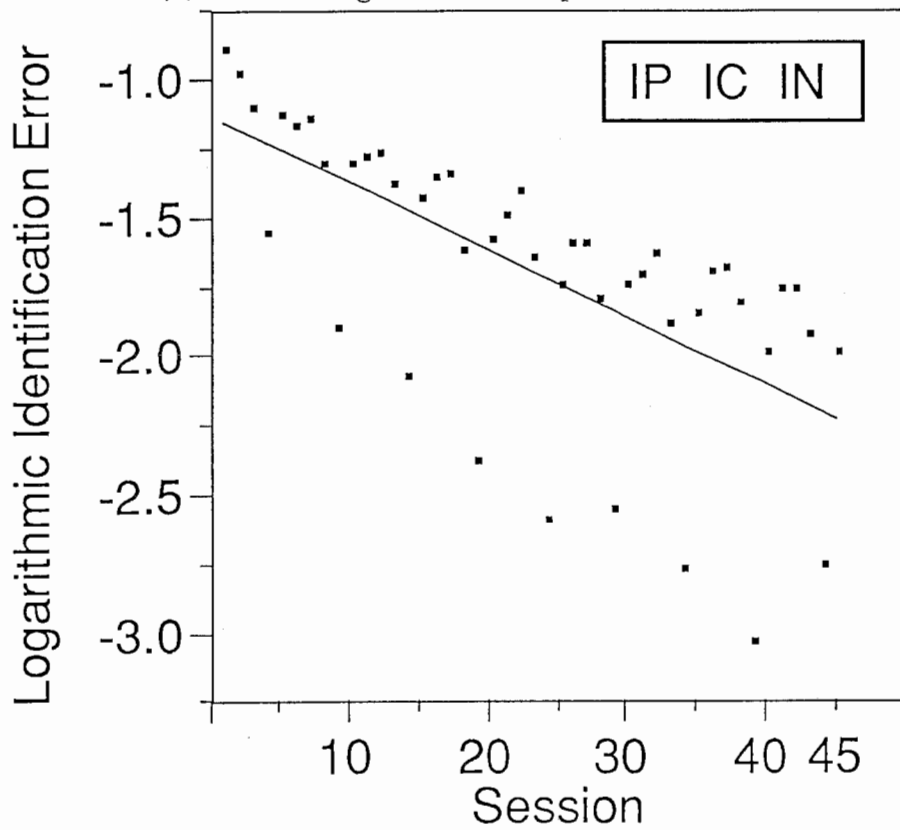


Figure 6: The decrease of the logarithmic identification error over the whole training sessions (Exp.2).

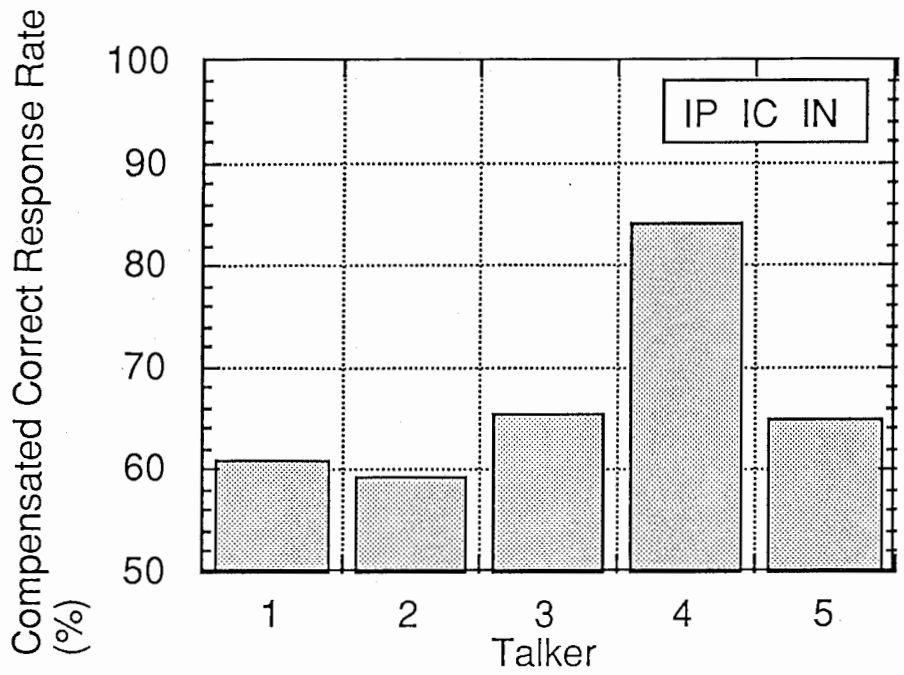


Figure 7: The estimated correct response rates by compensating the training effect for categories IP IC IN (Exp.2).

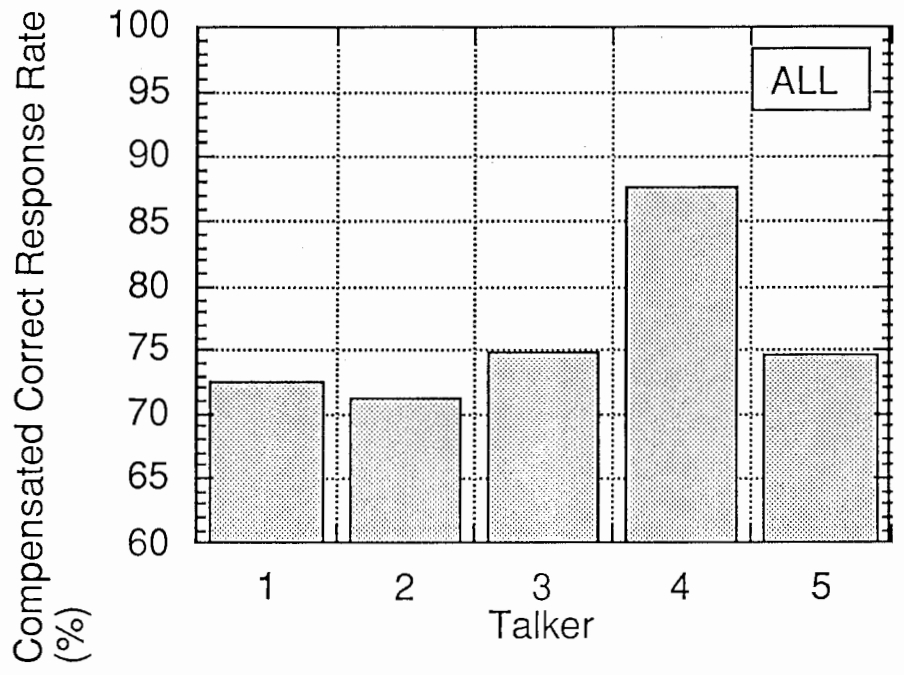
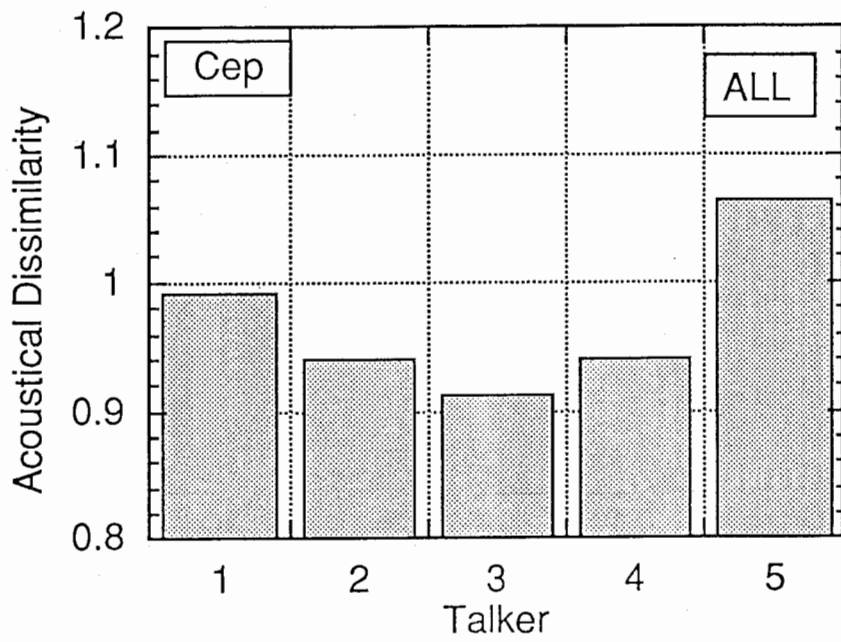
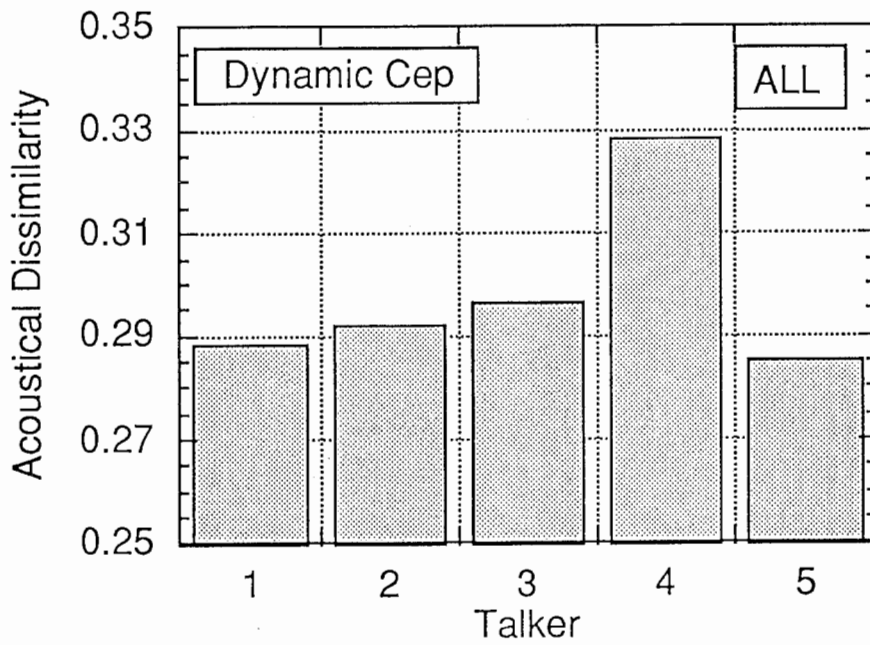


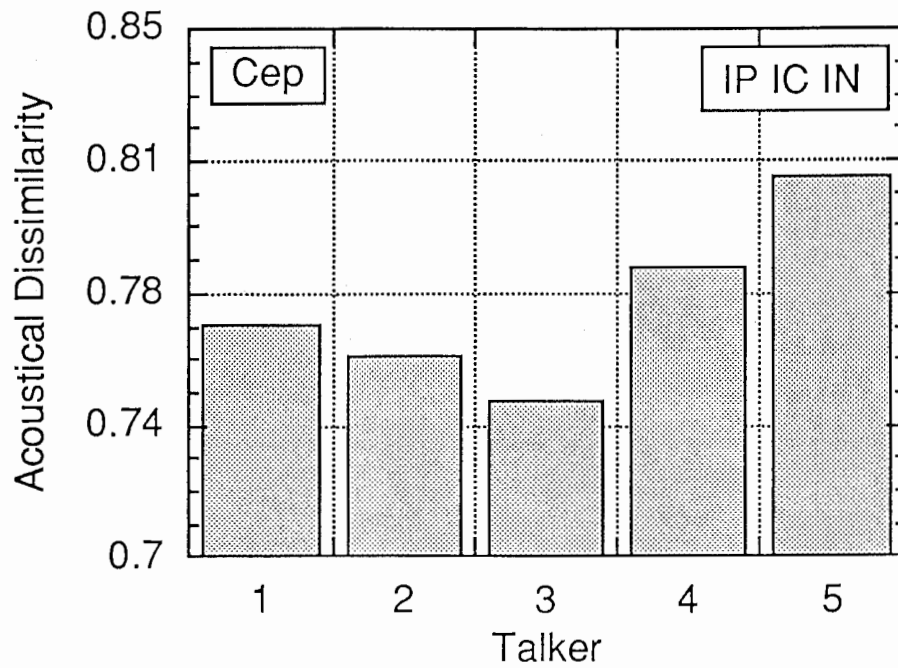
Figure 8: The estimated correct response rates by compensating the training effect for all categories (Exp.2).



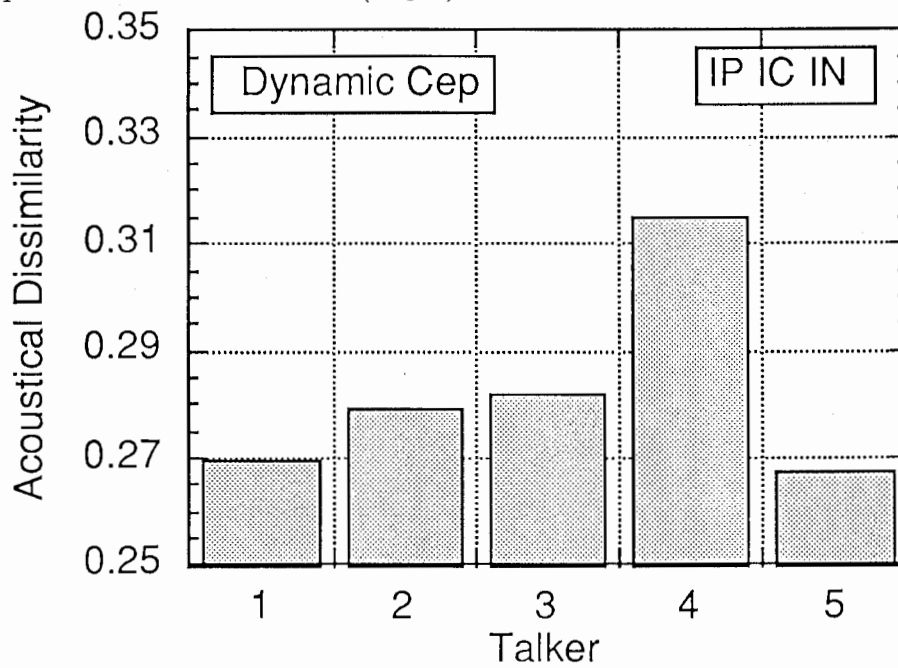
⊠ 9: The talker dependency pattern of the acoustic dissimilarity measured by a cepstrum (Exp.1).



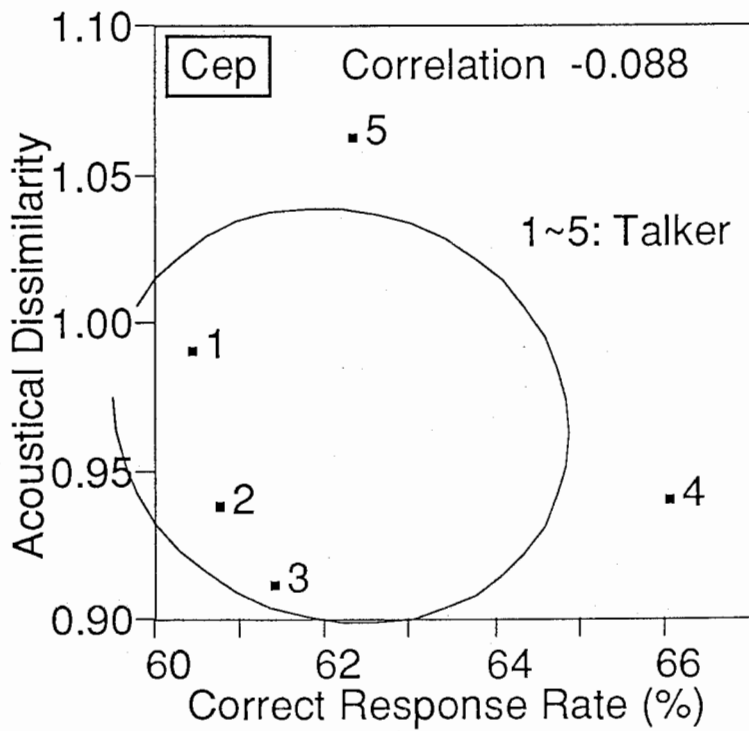
⊠ 10: The talker dependency pattern of the acoustic dissimilarity measured by a dynamic cepstrum (Exp.1).



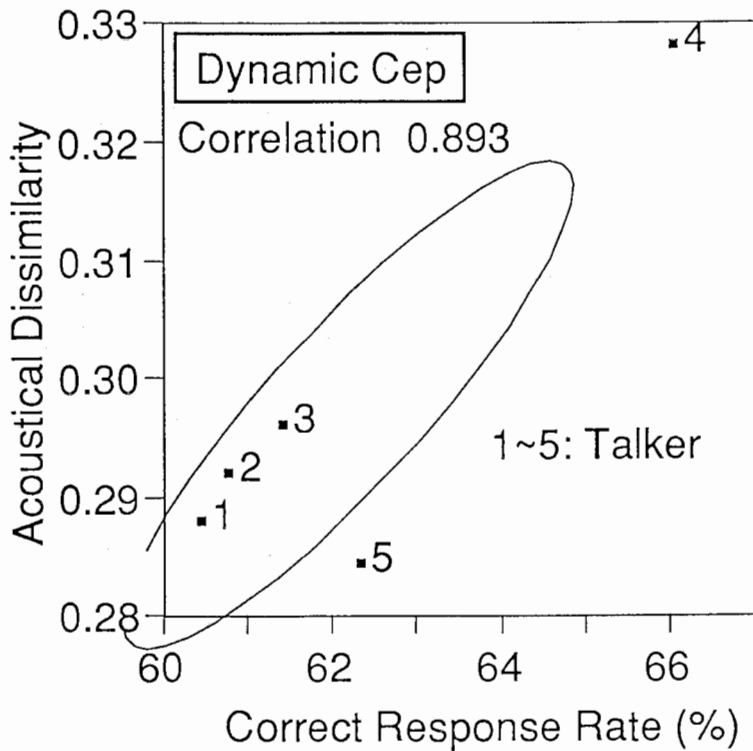
⊠ 11: The talker dependency pattern of the acoustic dissimilarity measured by a cepstrum for phonetic contexts IP IC IN (Exp.1).



⊠ 12: The talker dependency pattern of the acoustic dissimilarity measured by a dynamic cepstrum for phonetic contexts IP IC IN (Exp.1).



☒ 13: Correlation between the correct response rate and the acoustic dissimilarity measured by a cepstrum.



☒ 14: Correlation between the correct response rate and the acoustic dissimilarity measured by a dynamic cepstrum.

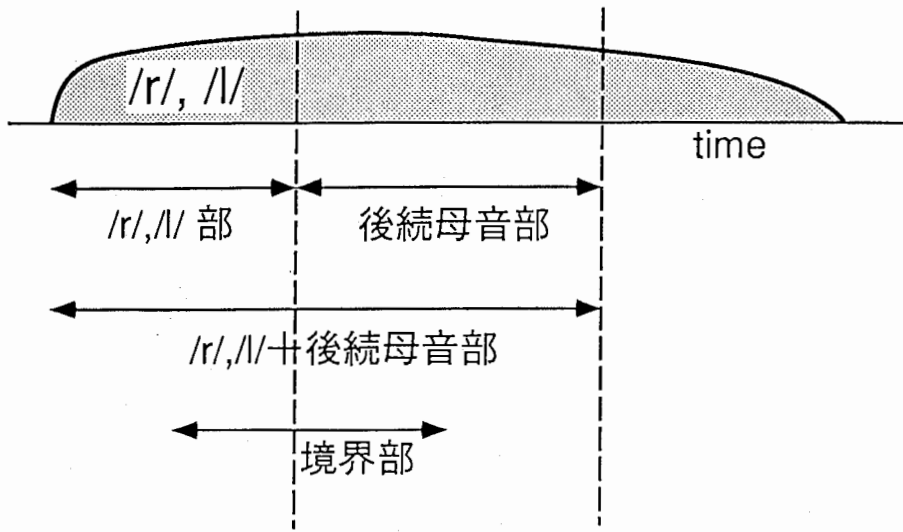


図 15: Portions for measuring the acoustic dissimilarity of a minimal pair.

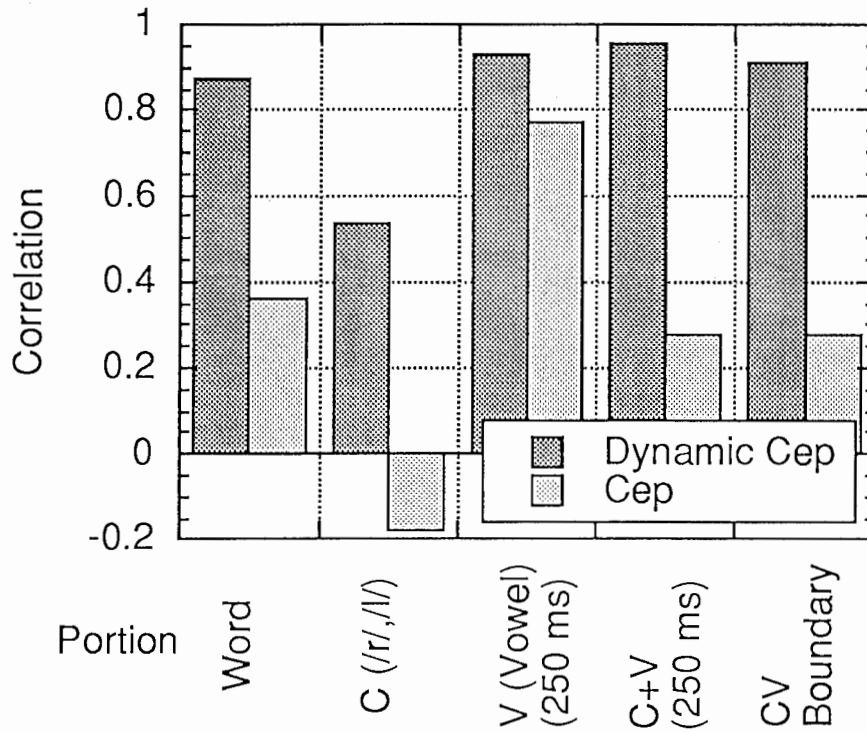


図 16: Correlation between the partial acoustic dissimilarity and the correct response rate (Exp.1).

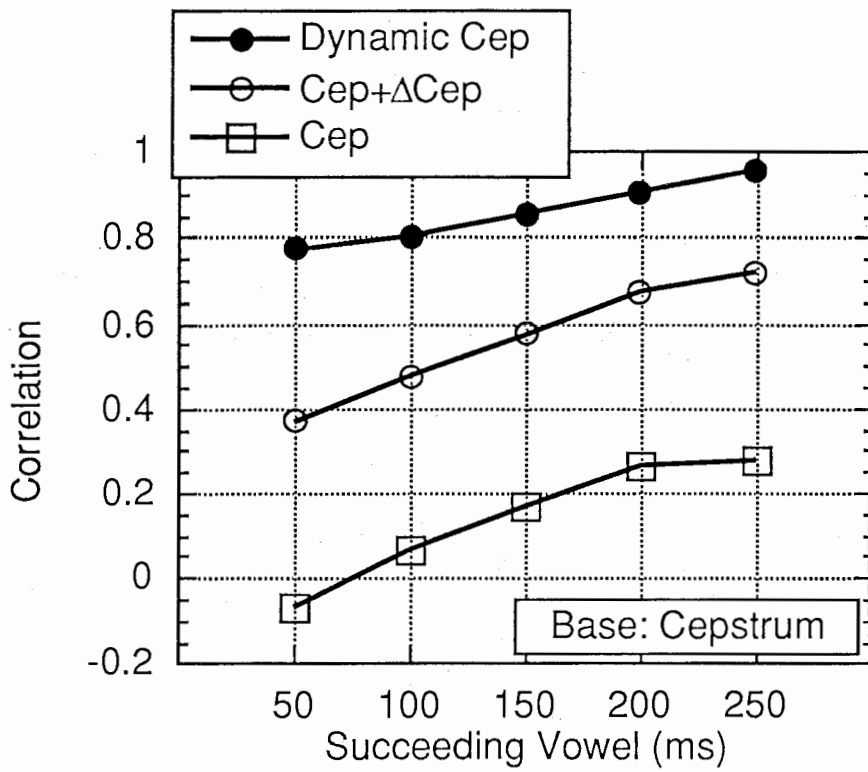


Figure 17: Correlation between the correct response rate and the dissimilarity of the /r/,/l/ + succeeding vowel using a cepstrum as the base parameter (Exp.1).

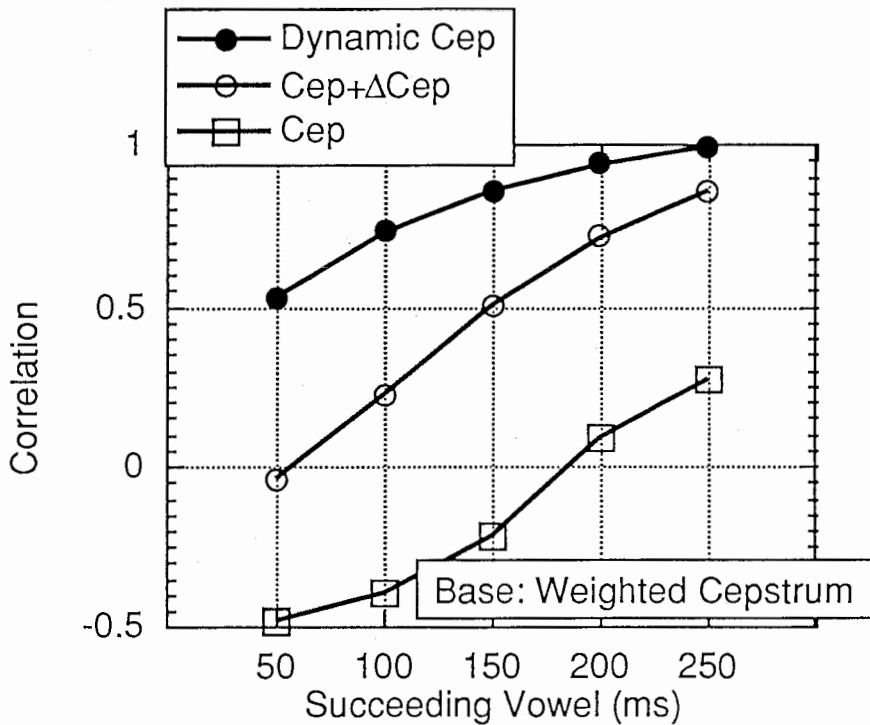


Figure 18: Correlation between the correct response rate and the dissimilarity of the /r/,/l/ + succeeding vowel using a weighted cepstrum as the base parameter (Exp.1).

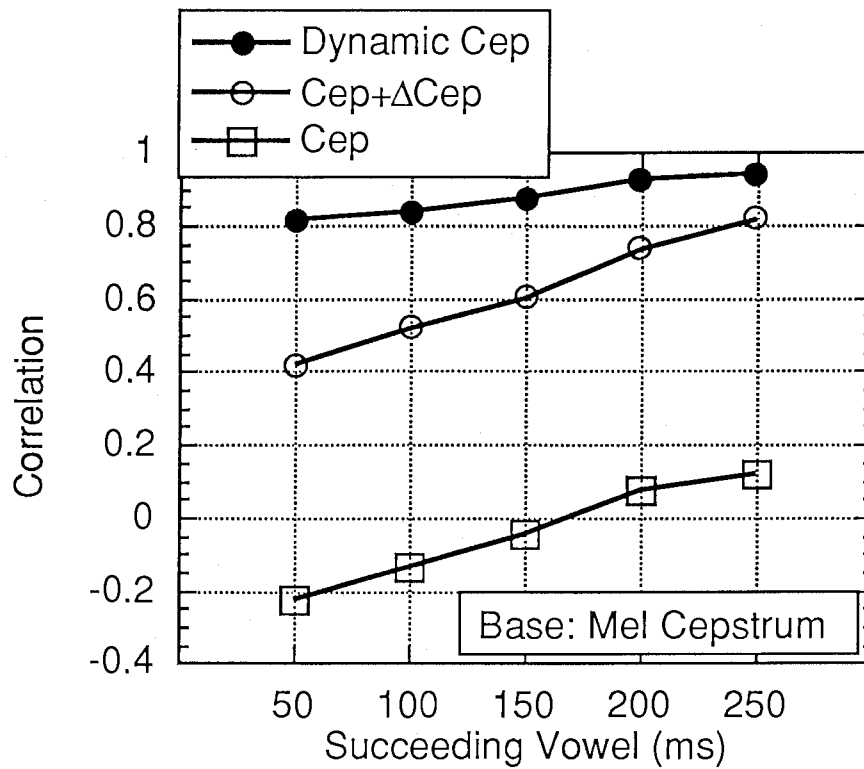


Figure 19: Correlation between the correct response rate and the dissimilarity of the /r/,/l/ + succeeding vowel using a mel-cepstrum as the base parameter (Exp.1).

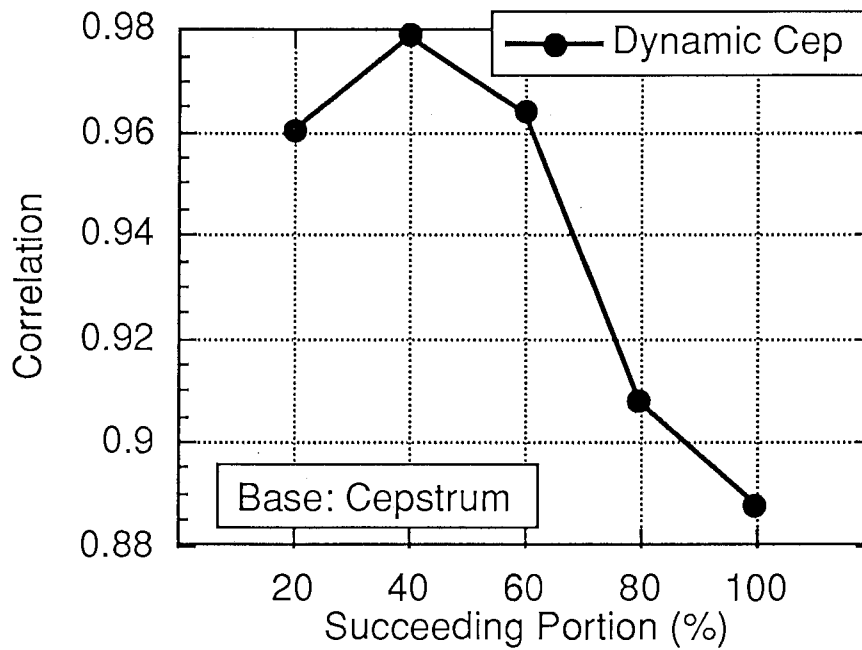


Figure 20: Correlation between the correct response rate and the dissimilarity of the /r/,/l/ + succeeding speech portion (Exp.1).

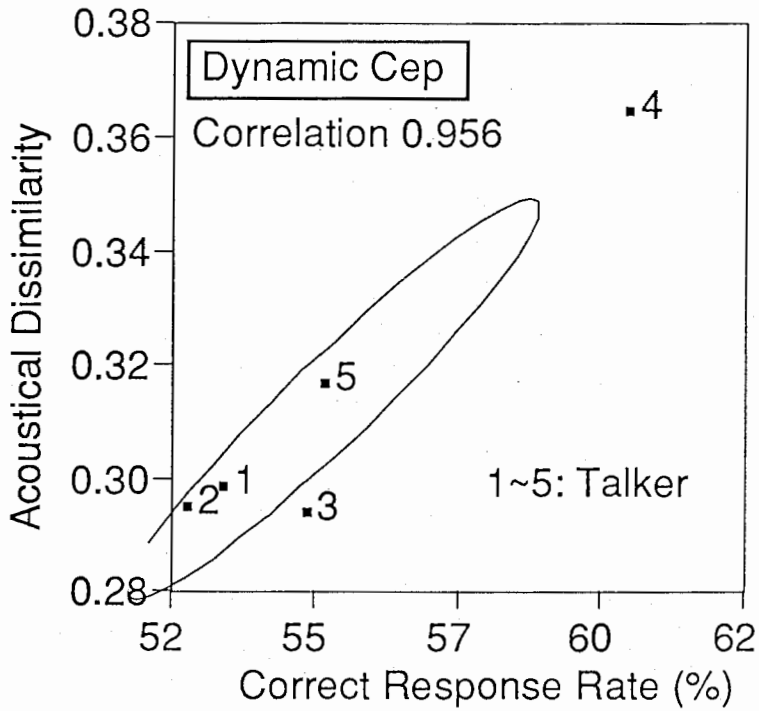


Figure 21: Distribution of talkers on the plane of the correct response rate and the dissimilarity of the /r/,/l/ + succeeding vowel (250 ms) (Exp.1).

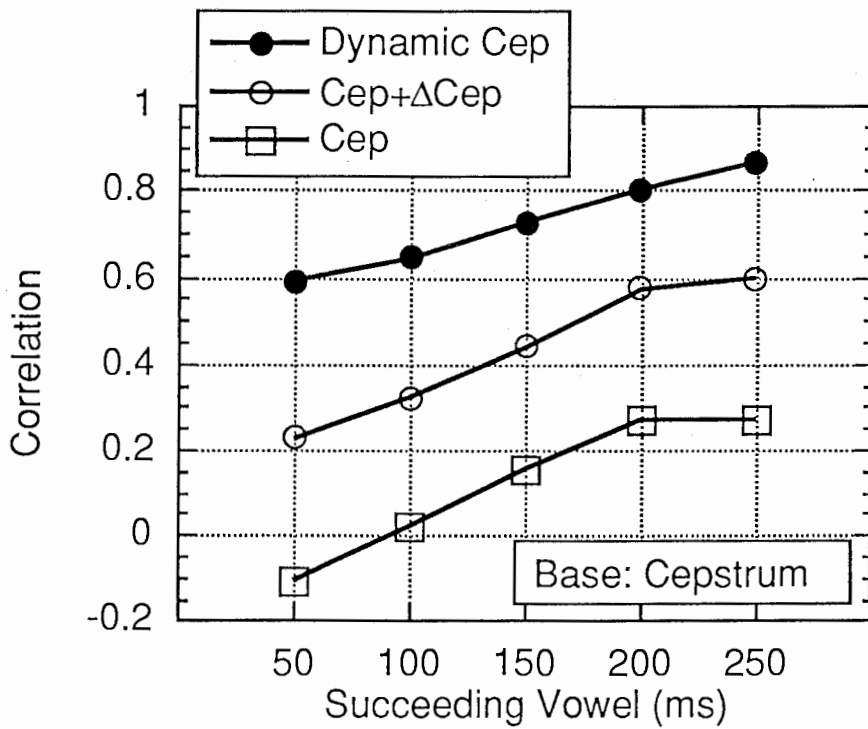


Figure 22: Correlation between the compensated correct response rate and the dissimilarities of the /r/,/l/ + succeeding vowel (Exp.2).

(b) IP IC IN

表 2: Minimal pair list.

(a) FC and FP

ID	L	R
FC002	BOLD	BOARD
FC003	BUILD	BIRD
FC004	CALLED	CARD
FC005	COLD	CORD
FC008	GOLD	GOURD
FC009	HAULED	HARD
FC010	HALT	HEART
FC013	KNEELED	NEARED
FC014	MALT	MART
FC016	MILD	MIRE
FC017	PELT	PERT
FC018	PULSE	PURSE
FC019	SHIELD	SHEARED
FC020	TILED	TIRED
FC021	WIELD	WIRED
FP001	POLL	PORE
FP002	MISFILE	MISFIRE
FP003	BAIL	BARE
FP004	BOWL	BOAR
FP006	DALE	DARE
FP008	DIAL	DIRE
FP009	FAIL	FAIR
FP010	FEEL	FEAR
FP012	HAIL	HARE
FP017	MAIL	MARE
FP018	MEAL	MERE
FP019	MILE	MIRE
FP020	PAIL	PARE
FP021	PEAL	PEER
FP024	REEL	REAR
FP025	ROLL	ROAR
FP026	SOUL	SORE
FP028	TAIL	TEAR
FP034	WHILE	WIRE

ID	L	R
IC001	BLUE	BREW
IC007	FLOW	FRO
IC010	GLUE	GREW
IC011	PLAY	PRAY
IC013	PLY	PRY
IC014	BLADE	BRAID
IC015	BLAIN	BRAIN
IC017	BLAZE	BRAZE
IC018	BLEACH	BREACH
IC019	BLED	BREAD
IC022	BLUES	BRUISE
IC023	BLUSH	BRUSH
IC025	CLAM	CRAM
IC026	CLASH	CRASH
IC027	CLASS	CRASS
IC028	CLICK	CRICK
IC029	CLIMB	CRIME
IC030	CLIQUE	CREEK
IC031	CLOAK	CROAK
IC032	CLOCK	CROCK
IC033	CLOUD	CROWD
IC034	CLOUT	KRAUT
IC035	CLOWN	CROWN
IC036	CLUTCH	CRUTCH
IC037	CLONE	CRONE
IC044	FLOCK	FROCK
IC045	FLOG	FROG
IC051	GLOAT	GROAT
IC055	PLOP	PROP
IC056	PLOD	PROD
IC057	BLACKEN	BRACKEN
IC058	BLACKISH	BRACKISH
IN003	PALLET	PARROT
IN004	PILOT	PIRATE
IN005	RAILING	RARING
IN006	ALLAY	ARRAY
IN008	ELECT	ERECT
IN009	OLEO	OREO
IP002	LACK	RACK
IP004	LAI	RAID
IP009	LAP	RAP
IP010	LASH	RASH
IP011	LATE	RATE
IP012	LAVE	RAVE
IP021	LED	RED
IP022	LEER	REAR
IP024	LEWD	RUDE
IP025	LICE	RICE
IP031	LIMB	RIM
IP032	LIME	RHYME
IP034	LIP	RIP
IP035	LIT	WRIT
IP040	LOAN	ROAN
IP043	LOCK	ROCK
IP044	LONG	WRONG
IP045	LOOK	ROOK
IP049	LOT	ROT
IP056	LUGE	ROUGE
IP058	LATH	WRATH

表 3: Correlation between the correct response rate and acoustical dissimilarity using a cepstrum as the basic parameter (Exp.1).

Vowel Portion = 5 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Word	0.874141	0.901418	0.359033
/r/,/l/	0.535530	0.152227	-0.178197
Vowel	-0.087733	-0.087725	-0.087735
/r/,/l/ + Vowel	0.753081	0.281186	-0.161522
Boundary	0.781507	0.367870	-0.091028

Vowel Portion = 10 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.728472	0.693672	0.363963
/r/,/l/ + Vowel	0.851166	0.393776	-0.121231
Boundary	0.873883	0.419403	-0.043269

Vowel Portion = 20 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.844971	0.945781	0.917679
/r/,/l/ + Vowel	0.960366	0.632055	0.041182
Boundary	0.936003	0.525895	0.020608

Vowel Portion = 40 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.887548	0.941033	0.875991
/r/,/l/ + Vowel	0.978610	0.890058	0.327433
Boundary	0.994012	0.773948	0.293945

Vowel Portion = 60 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.872843	0.886868	0.866240
/r/,/l/ + Vowel	0.963800	0.967924	0.540126
Boundary	0.997809	0.903001	0.536417

Vowel Portion = 80 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.807149	0.845967	0.815902
/r/,/l/ + Vowel	0.907732	0.987771	0.597899
Boundary	0.990271	0.939319	0.741194

Vowel Portion = 100 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.791223	0.863248	0.845155
/r/,/l/ + Vowel	0.887600	0.971084	0.714050
Boundary	0.991795	0.967086	0.863938

Vowel Portion = 150 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.791223	0.863248	0.845155
/r/,/l/ + Vowel	0.887600	0.971084	0.714050
Boundary	0.925688	0.945975	0.940278

Vowel Portion = 200 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.791223	0.863248	0.845155
/r/,/l/ + Vowel	0.887600	0.971084	0.714050
Boundary	0.893422	0.953546	0.967342

Vowel Period = 50 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.694146	0.527179	0.328265
/r/,/l/ + Vowel	0.772965	0.370030	-0.068016
Boundary	0.848665	0.406068	-0.043043

Vowel Period = 100 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.791113	0.706120	0.595996
/r/,/l/ + Vowel	0.803017	0.473423	0.065454
Boundary	0.870583	0.458810	0.058746

Vowel Period = 150 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.864870	0.811897	0.696789
/r/,/l/ + Vowel	0.851980	0.573520	0.165195
Boundary	0.880823	0.517854	0.129759

Vowel Period = 200 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.933214	0.882363	0.759329
/r/,/l/ + Vowel	0.904549	0.671825	0.262266
Boundary	0.899365	0.582201	0.234690

Vowel Period = 250 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.929590	0.906433	0.773507
/r/,/l/ + Vowel	0.956461	0.717002	0.279447
Boundary	0.910882	0.624471	0.278475

表 4: Correlation between the correct response rate and acoustical dissimilarity using a weighted cepstrum as the basic parameter (Exp.1).

Vowel Portion = 20 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Word	0.790739	0.731841	0.054882
/r/,/l/	0.153603	-0.180118	-0.439091
Vowel	0.619121	0.563517	-0.052051
/r/,/l/ + Vowel	0.884472	0.212817	-0.478263
Boundary	0.768307	0.074793	-0.460601

Vowel Portion = 40 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.700791	0.635163	0.345431
/r/,/l/ + Vowel	0.949505	0.792213	-0.242897
Boundary	0.955510	0.489488	-0.356717

Vowel Portion = 60 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.713012	0.650965	0.522826
/r/,/l/ + Vowel	0.897028	0.759866	0.092611
Boundary	0.947901	0.690686	-0.173491

Vowel Portion = 80 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.679732	0.646126	0.534517
/r/,/l/ + Vowel	0.805191	0.707414	0.131478
Boundary	0.925812	0.738718	0.093812

Vowel Portion = 100 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.668736	0.671148	0.549457
/r/,/l/ + Vowel	0.788495	0.741513	0.204914
Boundary	0.928826	0.782200	0.393644

Vowel Period = 50 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.365114	0.245951	-0.242866
/r/,/l/ + Vowel	0.527200	-0.041711	-0.478148
Boundary	0.538875	0.059421	-0.434005

Vowel Period = 100 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.579203	0.508584	0.144521
/r/,/l/ + Vowel	0.733294	0.225197	-0.391340
Boundary	0.701560	0.106486	-0.402477

Vowel Period = 150 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.691452	0.651129	0.431338
/r/,/l/ + Vowel	0.853723	0.504856	-0.219569
Boundary	0.795429	0.254580	-0.351424

Vowel Period = 200 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.806217	0.756270	0.611418
/r/,/l/ + Vowel	0.941693	0.715357	0.084630
Boundary	0.843986	0.381201	-0.223175

Vowel Period = 250 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.826607	0.790365	0.667767
/r/,/l/ + Vowel	0.993280	0.854546	0.269969
Boundary	0.880420	0.478450	-0.132750

表 5: Correlation between the correct response rate and acoustical dissimilarity using a mel cepstrum as the basic parameter (Exp.1).

Vowel Portion = 20 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Word	0.697313	0.657438	-0.483590
/r/,/l/	0.545385	0.153071	-0.303147
Vowel	0.900864	0.918418	0.831635
/r/,/l/ + Vowel	0.934088	0.739735	-0.283918
Boundary	0.938945	0.697738	-0.136209

Vowel Portion = 40 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.837892	0.831056	0.667151
/r/,/l/ + Vowel	0.877520	0.906340	-0.246933
Boundary	0.975573	0.904219	0.071384

Vowel Portion = 60 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.787438	0.718160	0.662736
/r/,/l/ + Vowel	0.829251	0.804512	0.072906
Boundary	0.961376	0.957719	0.419705

Vowel Portion = 80 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.582003	0.577937	0.576987
/r/,/l/ + Vowel	0.632863	0.613287	0.026677
Boundary	0.910696	0.911062	0.723033

Vowel Portion = 100

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.610674	0.660156	0.642203
/r/,/l/ + Vowel	0.650536	0.703173	0.518540
Boundary	0.937635	0.910583	0.931008

Vowel Period = 50 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.836962	0.706848	0.491189
/r/,/l/ + Vowel	0.815013	0.418254	-0.228225
Boundary	0.845284	0.486365	-0.165185

Vowel Period = 100 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.875679	0.805398	0.631334
/r/,/l/ + Vowel	0.834119	0.517828	-0.138601
Boundary	0.913881	0.577526	-0.040933

Vowel Period = 150 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.916850	0.861470	0.709440
/r/,/l/ + Vowel	0.873070	0.600697	-0.043567
Boundary	0.920523	0.634747	0.012282

Vowel Period = 200 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.962165	0.912089	0.781916
/r/,/l/ + Vowel	0.924350	0.734952	0.076065
Boundary	0.934932	0.679854	0.100969

Vowel Period = 250 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.908407	0.905771	0.802670
/r/,/l/ + Vowel	0.941260	0.817013	0.119301
Boundary	0.945097	0.706681	0.149517

表 6: Correlation between the correct response rate and the acoustical dissimilarity using a cepstrum as the basic parameter (Exp.2).

Vowel Portion = 5 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Word	0.877934	0.701985	0.301873
/r/,/l/	0.253350	-0.121148	-0.267269
Vowel	-0.081294	-0.081287	-0.081302
/r/,/l/ + Vowel	0.294876	-0.046276	-0.272842
Boundary	0.529624	0.115832	-0.138424

Vowel Portion = 10 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.324810	0.231954	0.425246
/r/,/l/ + Vowel	0.470370	0.003067	-0.196952
Boundary	0.589947	0.131503	-0.118025

Vowel Portion = 20 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.642252	0.735099	0.924217
/r/,/l/ + Vowel	0.663439	0.177736	0.024055
Boundary	0.721555	0.193174	0.012740

Vowel Portion = 40 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.782094	0.894285	0.900349
/r/,/l/ + Vowel	0.841512	0.588041	0.416012
Boundary	0.865885	0.443316	0.369547

Vowel Portion = 60 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.768577	0.844225	0.881938
/r/,/l/ + Vowel	0.881348	0.764774	0.583063
Boundary	0.935008	0.642914	0.622484

Vowel Portion = 80 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.685167	0.810959	0.823902
/r/,/l/ + Vowel	0.847040	0.819632	0.572467
Boundary	0.948064	0.774536	0.789287

Vowel Portion = 100 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.702241	0.827815	0.866384
/r/,/l/ + Vowel	0.854811	0.814612	0.696295
Boundary	0.956108	0.854587	0.859985

Vowel Portion = 150 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.702241	0.827815	0.866384
/r/,/l/ + Vowel	0.854811	0.814612	0.696295
Boundary	0.892525	0.841735	0.879056

Vowel Portion = 200 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.702241	0.827815	0.866384
/r/,/l/ + Vowel	0.854811	0.814612	0.696295
Boundary	0.880095	0.878202	0.915481

Vowel Period = 50 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.293400	0.213670	0.383478
/r/,/l/ + Vowel	0.355912	0.013394	-0.127174
Boundary	0.534632	0.116180	-0.109465

Vowel Period = 100 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.491719	0.483447	0.641651
/r/,/l/ + Vowel	0.393702	0.125507	0.021481
Boundary	0.576428	0.172208	0.052811

Vowel Period = 150 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.708758	0.678781	0.739911
/r/,/l/ + Vowel	0.514398	0.259551	0.156718
Boundary	0.589071	0.213674	0.148245

Vowel Period = 200 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.870902	0.796567	0.789013
/r/,/l/ + Vowel	0.650829	0.432150	0.267164
Boundary	0.634625	0.311651	0.246776

Vowel Period = 250 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.921091	0.822767	0.788392
/r/,/l/ + Vowel	0.750644	0.454877	0.261784
Boundary	0.679257	0.353911	0.297613

表 7: Correlation between the correct response rate and the acoustical dissimilarity using a weighted cepstrum as the basic parameter (Exp.2).

Vowel Portion = 5 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Word	0.842004	0.768201	0.343502
/r/,/l/	-0.037483	-0.273060	-0.506079
Vowel	-0.081296	-0.081298	-0.081306
/r/,/l/ + Vowel	0.015358	-0.195973	-0.588728
Boundary	0.284509	-0.045464	-0.429992

Vowel Portion = 10 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.191970	0.148066	-0.164271
/r/,/l/ + Vowel	0.248229	-0.233823	-0.565057
Boundary	0.369349	-0.012474	-0.473852

Vowel Portion = 20 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.601505	0.615553	0.152478
/r/,/l/ + Vowel	0.657460	0.082350	-0.457146
Boundary	0.576151	-0.002636	-0.424884

Vowel Portion = 40 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.723591	0.712377	0.484642
/r/,/l/ + Vowel	0.985084	0.890853	-0.059522
Boundary	0.911417	0.503253	-0.203534

Vowel Portion = 60 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.714339	0.687528	0.582080
/r/,/l/ + Vowel	0.939137	0.789260	0.217581
Boundary	0.975659	0.773252	0.047845

Vowel Portion = 80 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.689988	0.695857	0.601479
/r/,/l/ + Vowel	0.851174	0.760860	0.278573
Boundary	0.960443	0.828293	0.330032

Vowel Portion = 100 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.681626	0.700416	0.615200
/r/,/l/ + Vowel	0.826524	0.761292	0.348695
Boundary	0.957964	0.855349	0.548887

Vowel Portion = 150 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.681626	0.700416	0.615200
/r/,/l/ + Vowel	0.826524	0.761292	0.348695
Boundary	0.856240	0.760645	0.574736

Vowel Portion = 200 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.681626	0.700416	0.615200
/r/,/l/ + Vowel	0.826524	0.761292	0.348695
Boundary	0.820823	0.769185	0.614713

Vowel Period = 50 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.205167	0.190990	-0.032727
/r/,/l/ + Vowel	0.251601	-0.139298	-0.515083
Boundary	0.332205	-0.014380	-0.451865

Vowel Period = 100 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.459307	0.481948	0.295134
/r/,/l/ + Vowel	0.446913	0.095576	-0.421855
Boundary	0.494523	0.047136	-0.361589

Vowel Period = 150 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.630373	0.645721	0.509684
/r/,/l/ + Vowel	0.651295	0.390210	-0.184896
Boundary	0.595778	0.180536	-0.287223

Vowel Period = 200 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.778986	0.744582	0.648850
/r/,/l/ + Vowel	0.810772	0.612703	0.205207
Boundary	0.694497	0.333138	-0.122102

Vowel Period = 250 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.828639	0.792251	0.679323
/r/,/l/ + Vowel	0.936568	0.763717	0.350034
Boundary	0.762698	0.439848	0.003714

表 8: Correlation between the correct response rate and the acoustical dissimilarity using a mel cepstrum as the basic parameter (Exp.2).

Vowel Portion = 5 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Word	0.540541	0.311955	-0.464837
/r/,/l/	0.293344	-0.167371	-0.450228
Vowel	-0.081287	-0.081264	-0.081300
/r/,/l/ + Vowel	0.402690	-0.083779	-0.514380
Boundary	0.630967	0.171564	-0.292070

Vowel Portion = 10 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.629876	0.467081	0.452098
/r/,/l/ + Vowel	0.602510	-0.073846	-0.506932
Boundary	0.701113	0.183103	-0.302162

Vowel Portion = 20 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.888262	0.837250	0.827157
/r/,/l/ + Vowel	0.821652	0.030504	-0.525017
Boundary	0.834905	0.249945	-0.217876

Vowel Portion = 40 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.675866	0.745081	0.688948
/r/,/l/ + Vowel	0.753912	0.511118	-0.415875
Boundary	0.939953	0.488073	-0.033794

Vowel Portion = 60 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.634374	0.598827	0.658324
/r/,/l/ + Vowel	0.711242	0.506864	-0.131645
Boundary	0.937964	0.727844	0.406732

Vowel Portion = 80 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.370344	0.445938	0.537573
/r/,/l/ + Vowel	0.425555	0.300088	-0.210748
Boundary	0.851839	0.776718	0.728814

Vowel Portion = 100 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.423767	0.516998	0.616931
/r/,/l/ + Vowel	0.472242	0.413747	0.122984
Boundary	0.920299	0.842255	0.880892

Vowel Portion = 150 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.423767	0.516998	0.616931
/r/,/l/ + Vowel	0.472242	0.413747	0.122984
Boundary	0.591889	0.430625	0.520290

Vowel Portion = 200 %

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.423767	0.516998	0.616931
/r/,/l/ + Vowel	0.472242	0.413747	0.122984
Boundary	0.552325	0.545705	0.656149

Vowel Period = 50 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.557914	0.337788	0.356470
/r/,/l/ + Vowel	0.469543	-0.052120	-0.398812
Boundary	0.656049	0.139025	-0.283252

Vowel Period = 100 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.710696	0.560024	0.548985
/r/,/l/ + Vowel	0.520935	0.041032	-0.314707
Boundary	0.721176	0.187976	-0.137542

Vowel Period = 150 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.857858	0.725420	0.694736
/r/,/l/ + Vowel	0.630843	0.164162	-0.158440
Boundary	0.726071	0.214535	-0.077535

Vowel Period = 200 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.905284	0.830657	0.781252
/r/,/l/ + Vowel	0.738144	0.420721	0.019079
Boundary	0.776273	0.313226	-0.002497

Vowel Period = 250 ms

Portion	Dynamic Cep.	Cep. + Δ Cep.	Cep.
Vowel	0.887459	0.836861	0.784565
/r/,/l/ + Vowel	0.835659	0.496638	0.026288
Boundary	0.802528	0.337268	0.065695