

TR-I-0136

音韻環境に応じた音声合成素片の接続方法の検討

On the Concatenation of Speech Synthesis Units  
According to Unit Extraction context

安部勝雄、武田一哉、匂坂芳典

Katsuo ABE, Kazuya TAKEDA and Yoshinori SAGISAKA

1990. 2

#### 内容梗概

我々が用いている複合音声単位を用いる音声合成方式では、音韻環境を考慮して合成素片が選択される。本稿では、選択された合成素片の素片境界近傍の音韻環境に応じて素片の接続を行う方法を提案し、その有効性について述べた。

合成素片は音韻ラベル付き音声データファイルから切り出されるが、視察によるラベルに基づく素片境界での接続歪を抑圧するために、素片境界を固定せず隣接する素片にあわせて視察境界近傍で歪の小さい接続点を探索することにした。接続点の探索は、切り出す素片の音韻環境に応じて探索区間、探索基準を設定している。探索基準として、素片間のケプストラム距離最小、パワー最小、低周波数スペクトルパワー最小、スペクトル変化率最大の4つの基準を導入した。音韻ラベルに基づく素片境界での接続の場合に算出された接続歪との比較、及び聴取実験により、本手法の有効性が確認された。

ATR 自動翻訳電話研究所  
ATR Interpreting Telephony Research Laboratories

## 1. はじめに

規則音声合成において合成素片の接続方法は素片の選択方法と同様に合成音の品質を決定する上で重要な要因である。素片選択法により適切な音声スペクトルを選択し、その音声スペクトルの性質を損なわないよう接続することが要求される。

従来の合成方式では画一的に合成素片を選択し、素片毎に固定した境界を用いて補間による接続が多く用いられる。これらの方法では、隣接する素片間の境界でのスペクトル距離が大きい場合、補間によって各々の境界を滑らかにつなぐ接続スペクトルを生成することは難しく、少なからず接続歪を生じていた。このような接続歪を抑えるには、補間にたよる以前に、まず素片間の境界でのスペクトル距離を小さくするような素片の切り出しが重要と思われる。

複合音声単位を用いる合成方式<sup>[1]</sup>では、素片の選択及び接続という二つの処理を相互に関連性を持たせて取り扱い、接続歪の低減を図ろうとしている。即ち素片の選択には音声スペクトルに影響の大きい音韻環境<sup>[2]</sup>を考慮するとともに、接続の容易さなども選択基準に組み込み<sup>[3]</sup>、接続側ではそれに応じて接続歪が小さい素片の切り出し、接続を行う。

本稿では、音韻環境を考慮した合成素片の接続方法について検討し、一対比較による受聴試験の結果、提案する素片接続法の効果を確認したので報告する。

## 2. 接続方法

合成素片の接続によって生じる接続歪は全ての接続点(切り出し点)で一様に生じるのではなく、接続点近傍の音韻環境によって大きな歪の発生しやすい所とそうでない所がある。例えば母音から有声子音への遷移区間で素片をつなごうとすれば大きな歪が発生しやすく、無声破裂子音の閉鎖区間などで接続すれば接続歪は殆ど目立たない。

複合音声単位を用いる合成方式では、音韻ラベル付き音声データベース<sup>[4]</sup>から素片境界近傍の音韻環境が隣接素片の音韻環境に類似するよう素片が選ばれるが、ラベルによる素片境界点は音韻の遷移区間にあるので、ここでの接続には歪を伴い易い。従来の音声合成方式では均一な音韻環境から単位を切り出し固定した単位境界を用いるためこのような接続歪みに対処することは難しかった。我々の合成方式では予め単位の境界を固定しないため、選出された単位どうしの音韻環境に応じて接続点をラベルによる素片境界の前後にずらすことにより接続歪を抑えることが期待できる。以下、合成素片の音韻環境に応じた接続点の探索区間、及び探索方法を詳細に述べる。

### 2.1 素片境界近傍の音韻環境

接続点を探索する際に考慮すべき素片境界近傍の音韻環境を図1に示す。ここで、

A、B: 隣接する合成素片

$a_1, a_2$ : 合成素片Aの末尾、及び後続音韻

$b_1, b_2$ : 合成素片Bの先行、及び先頭音韻

$la_1, la_2$ : 音韻 $a_1, a_2$ 内での接続点探索区間

$lb_1, lb_2$ : 音韻 $b_1, b_2$ 内での接続点探索区間

また、音韻環境及び接続を考慮した素片選択法<sup>[3]</sup>では母音(撥音)から子音への遷移区間に素片境界がくるように合成素片を選択しているため、 $a_1$ 及び $b_1$ は母音(撥音)である。接続に際しては、

- ① 無声子音部での接続は容易である。
- ② 有声の同一音韻での接続は音韻間のスペクトル距離最小となる点で素片をつなげば接続歪を抑えることができる。

ということから、I.無声子音部での接続、II.同一有声音韻での接続、III.その他の接続に分けて音韻環境を考える。音韻環境として次の6つの場合を考慮した。例を図2に示す。

[場合 I ] 無声子音部での接続

$a_2, b_2$ が無声子音,  $a_2 = b_2$ .

[場合 II-1] 有声音全体での接続

$a_2, b_2$ が有声音,  $a_1 = b_1, a_2 = b_2$ .

[場合 II-2] 母音部での接続

$a_1 = b_1$ .

[場合 II-3] 有声音部での接続

$a_2, b_2$ が有声音,  $a_2 = b_2$ .

[場合 II-4] 同一母音連続の接続

$a_1, b_2$ が母音,  $a_1 = b_2$ .

[場合 III] その他の接続

$a_1 \neq b_1, a_2 \neq b_2, a_1 \neq b_2$

## 2.2 接続点探索のための基準

合成素片を抽出する母体となる音声データ(サンプリング周波数12kHz、語長16bit)は、フレーム長21.3ms、フレーム周期5msで分析され、30次の改良ケプストラム<sup>[5]</sup>の形でファイル化されている。合成素片の接続点を探索するために、ケプストラムから計算される次の4つの基準を用いた。

[基準1] ケプストラム距離最小

素片AとBとの間でケプストラムのユークリッド距離が最小となるフレーム。

パワーに相当するケプストラム零次項も距離計算に含める。有声の同一音韻で接続する場合の歪最小点の探索に用いる。

[基準2] パワー最小

素片内でケプストラムの零次項が最小となるフレーム。無声破裂音の閉鎖区間内の探索に用いる。

[基準3] 低周波数スペクトルパワー最小

素片内で50Hzから1.5kHzまでの低周波数スペクトルパワーが最小となるフレーム。無声摩擦音の摩擦区間での探索に用いる。

[基準4] スペクトル変化率最大

素片内で時間方向のケプストラム変化率(隣接フレーム間のケプストラム距離)が最大となるフレーム。連続する有声音の遷移点の探索に用いる。

これらは次項で述べるように、音韻のカテゴリーに応じて用いられる。

## 2.3 接続点の探索

2.1項で述べた音韻環境に応じて、探索区間及び探索基準を設定した(表1参照)。

[場合Ⅰ] 無声子音部での接続

無声子音部での接続は厳密に接続点を求める必要はなく、無声破裂音のときには閉鎖区間、無声摩擦音のときには安定した摩擦区間であればどこで接続しても接続歪は十分小さい。探索区間を音韻ラベルに基づく素片境界点から無声子音の中央部までとし、先行音韻(母音)の特徴の無くなるフレームを接続点とした。探索基準は素片間のケプストラム距離ではなく、無声破裂音では各素片内でのパワー最小、無声摩擦音では低周波数スペクトルパワーの最小を用いた。

[場合Ⅱ] 有声音での接続

有声音での接続では、同一の有声音韻区間でケプストラム距離最小の探索基準を用いる。探索区間は一致している音韻全域にわたるのではなく、各音韻の半分までとした。これは、音韻の立ち上がり区間どうし、立ち下がり区間どうしでの接続が行われるように考慮したものである。[場合Ⅱ]の接続を行ったとき、特に[場合Ⅱ-4]同一母音連続の接続で効果は顕著に現れ、単純な接続により生ずるクリック音をかなり抑えることができるようになった。

[場合Ⅲ] 前後の音韻が異なる接続

素片境界近傍で一致する音韻がない場合は、素片間にわたる接続点の探索は意味がなく、各素片毎に素片境界を探索する。素片境界を挟む前の音韻が無声子音の場合には[場合Ⅰ]無声子音部での接続と同様に探索し、有声音の場合には時間方向のスペクトル変化が最も大きい箇所を接続点とした。現在素片の抽出母体として5240単語を用いているが、抽出母体を整備しこの接続環境の適用を極力少な

くすることが望まれる。

### 3. 素片境界近傍の音韻環境と接続歪の関係

本稿で提案した接続方法の有効性を確認し、さらに適切な接続点を探索できるようにするために、素片境界近傍の音韻環境と接続歪の関係を調べた。接続歪として零次項を含む改良ケプストラム30次の接続フレーム間でのユークリッド距離を用いた。サンプル文章は合計193文章、総接続箇所は1,347個である。

#### 3.1 後続素片の先頭音韻と接続歪の関係

まず後続素片Bの先頭音韻 $b_2$ の音韻カテゴリ別に次の二種類の接続歪を求めた。

X: 音韻ラベルによる素片境界での接続歪

Y: 本手法により決定した接続点での接続歪

$b_2$ の音韻カテゴリ別のX、Yの接続歪の平均と分散を求め、あわせて図3に示す。

どの音韻カテゴリにおいてもYはXより改善されている。Yの中では音韻 $b_2$ が半母音、鼻音、母音のとき比較的大きな値をとるので、これらの場合について音韻環境と接続歪の関係を調べた。音韻環境は表1の場合をさらに細分して次の7通りとした。データの少ない[場合II-4]は省いた。

(II-1) [場合II-1]

(II-2-a) [場合II-2]で $a_2$ 、 $b_2$ が同一音韻カテゴリ

(II-2-b) [場合II-2]で $a_2$ が無音( $a_1$ は語尾)

(II-2-c) [場合II-2]で(II-2-a)、(II-2-b)以外

(II-3-a) [場合II-3]で $b_1$ が無音( $b_2$ は語頭)

(II-3-b) [場合II-3]で(II-3-a)以外

(III) [場合III]

結果を図4に示す。この図に示されるように以下のことが判る。

(1) (II-1)、(II-2-a)、(II-2-c)では、自然音声の隣接フレーム間ケプストラム距離程度(1.0から1.5)に接続歪は抑えられる。

(2) (II-2-b)が示すように、先行素片Aの末尾の音韻 $a_1$ と後続素片Bに先行する音韻 $b_1$ が一致するが、 $a_1$ が語尾であるとき、他の場合に比べ接続歪は大きい。この結果は、「音韻表記が同じ母音でも、語中のスペクトルは語頭、あるいは語尾のものとは異なる」という文献[2]の報告内容に一致する。

(3) (II-3-a)が示すように、先行素片Aに後続する音韻 $a_2$ と後続素片Bの先頭音韻 $b_2$ が一致するが、 $b_2$ が語頭るとき、他の場合に比べ接続歪は大きい。

#### 3.2 ラベル境界での接続歪と、本手法による接続歪の相関

合成素片の選択の際、接続の容易さを音韻環境だけではなく、ラベルによる素片境界での接続歪から定量的に知ることができれば、より有効な素片選択が可能になると思われる。音韻 $b_2$ が半母音、鼻音の場合のラベルによる素片境界での接続歪と、本稿で提案する方法での接続歪の関係を図5、図6に示した。この図に示されるように、データにばらつきはあるが、音韻ラベルによる素片境界での接続歪と本手法による接続歪にはある程度の相関が見られる。この結果から、音韻ラベルによる素片境界での接続歪を選択尺度に導入できる可能性が示された。

### 3.3 受聴試験

本稿で提案した接続方法の有効性を確認するために、音韻ラベルに基づく素片境界を接続点とする接続法との一対比較による簡単な受聴試験を行った。合成音の作成に用いた文章は3文節から成る6文章で、合成音の提示は、各文章について、(本稿で提案する接続)-(ラベル境界での接続)、またその順番を入れ替えたものをペアとして計24文章をランダムに配し、プリファレンスをとった。被験者は男性3名、女性4名の計7名である。実験の結果、本稿で提案する接続による合成音が約76%のプリファレンス・スコアを得た。

### 4. おわりに

音韻環境を考慮して選択された音声合成素片を滑らかに接続するために、素片境界の音韻環境に応じて接続点を探索する方法を提案し、その有効性を確認した。今後は素片選択法との処理の融合性を高め、接続歪を抑えきれない場合には合成素片の再選択を行う等選択方法ともあわせ合成音の品質向上を図る。

### 謝辞

日頃御指導頂く樽松社長、鹿野室長、並びに討論頂いたATR 自動翻訳電話研究所の皆様、データ作成に協力頂いた早稲田大学の橋本賢治氏に深謝致します。

### 文献

- [1] 匂坂「種々の複合音声単位からの音声合成」、音響学会講演論文集、1987.10
- [2] 安部、匂坂「抽出環境が音節スペクトルに与える影響の分析」、音響学会講演論文集、1987.10
- [3] 武田、匂坂、桑原「合成単位の選択における音韻的知識と韻律的属性の処理の統合」、音響学会講演論文集、1989.10
- [4] 武田、匂坂、片桐、桑原「音韻ラベルを持つ日本語音声データベースの構築」、信学技報SP87-19、1987.7
- [5] 今井、阿部「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(A)、J62-A、4、p.217、1979

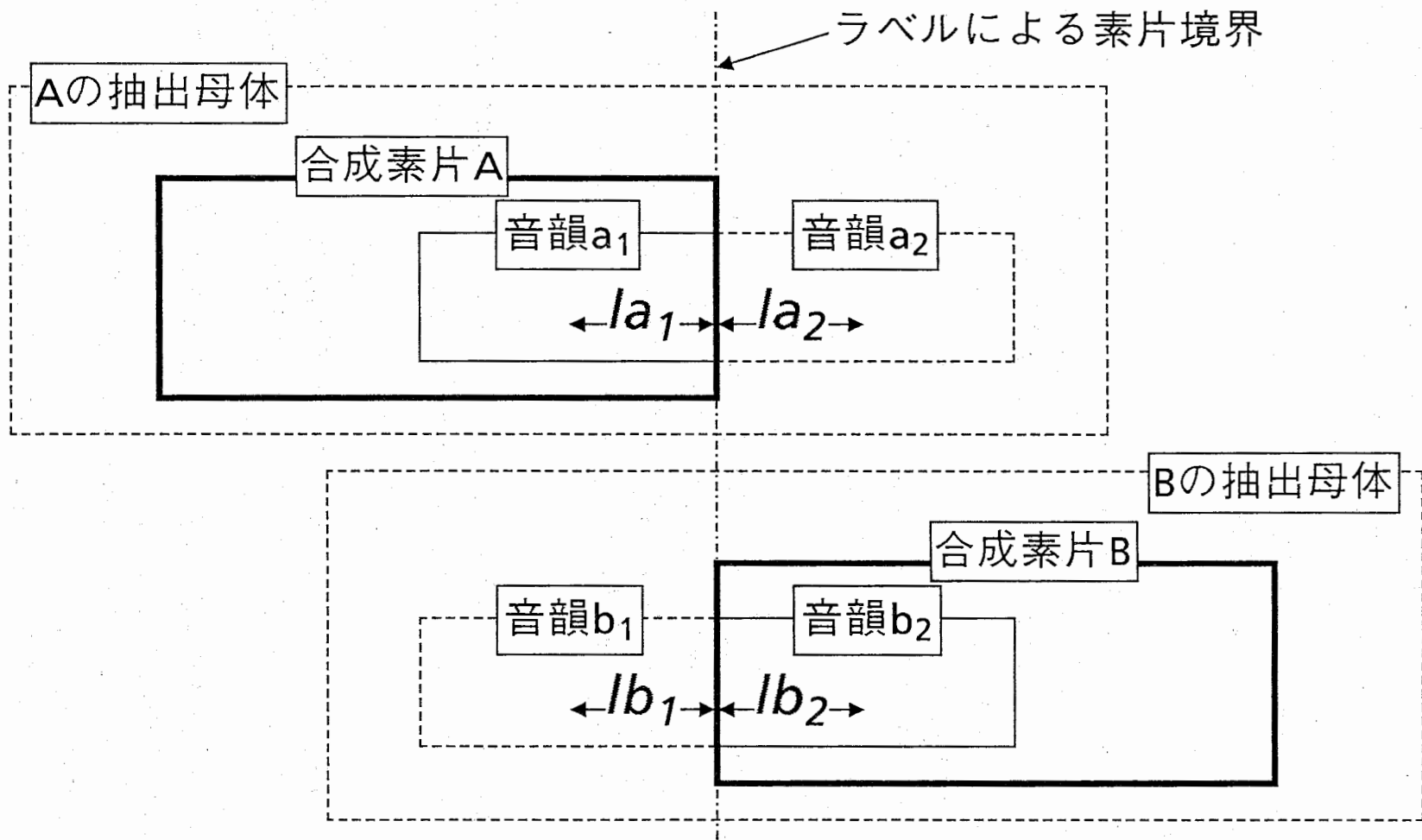
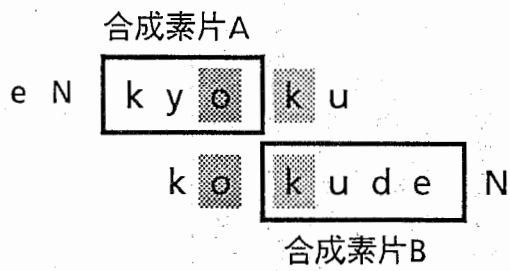
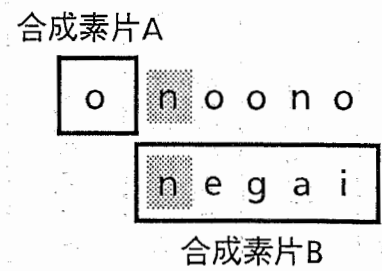


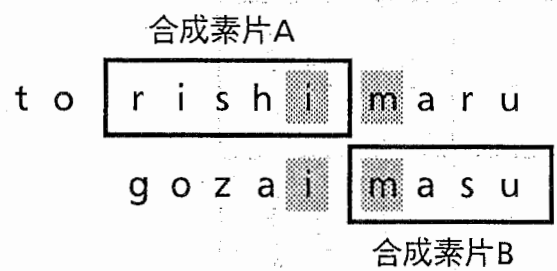
図1. 素片境界近傍の音韻環境



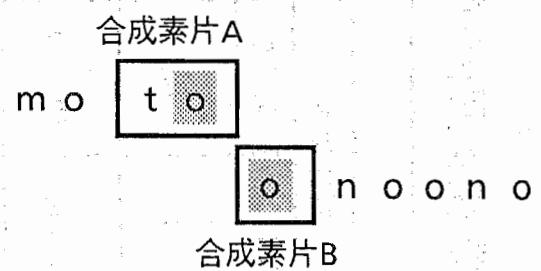
場合Ⅰ. 無声子音部での接続



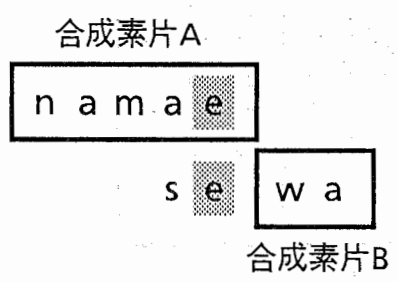
場合Ⅱ-3. 有声子音部での接続



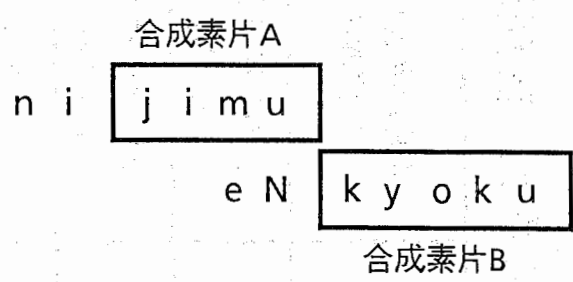
場合Ⅱ-1. 有声音全体での接続



場合Ⅱ-4. 同一母音連続の接続



場合Ⅱ-2. 母音部での接続



場合Ⅲ. その他

図2. 素片境界近傍の音韻環境による接続の場合分け

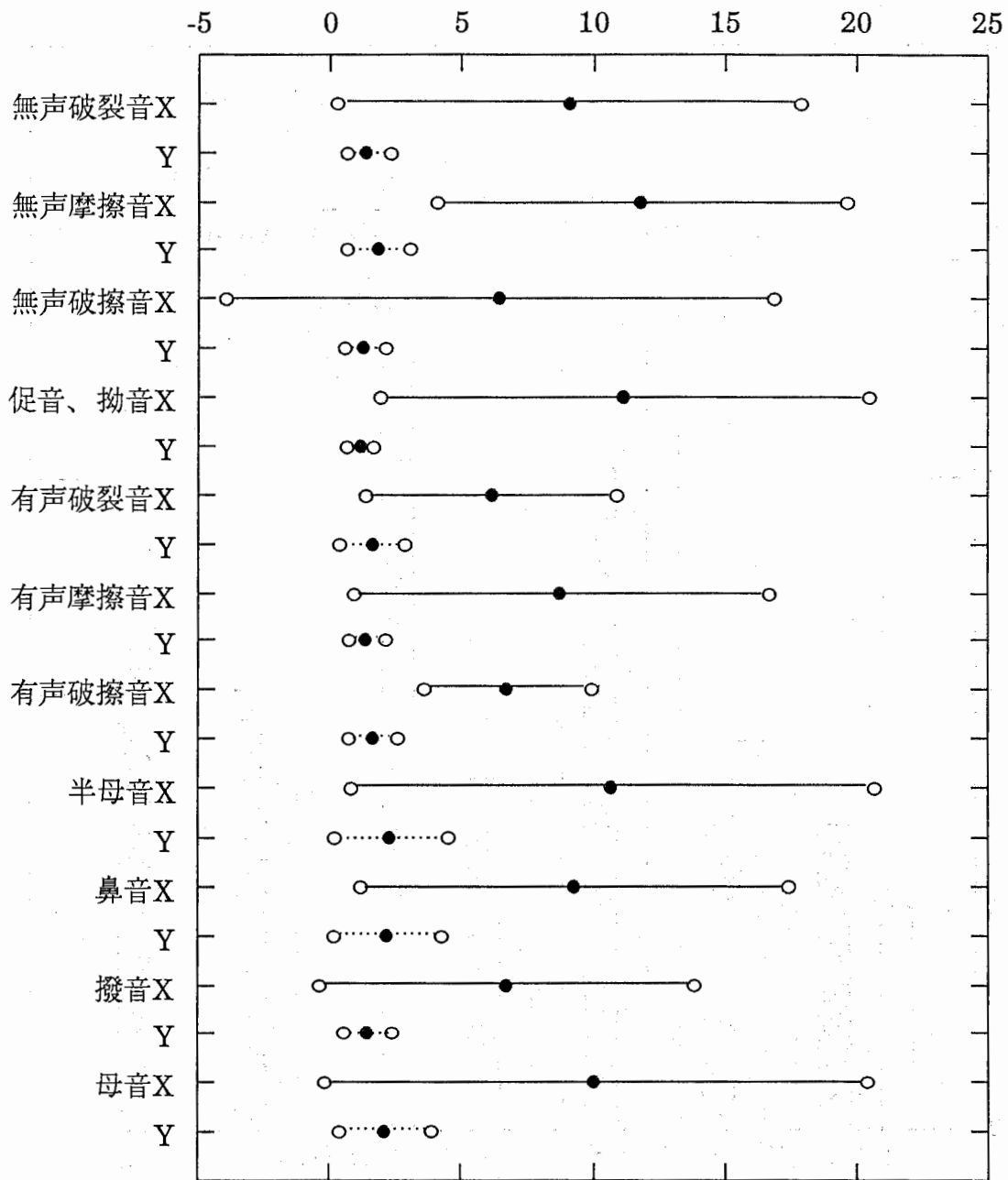


表1 接続点の探索 (音韻環境、探索区間の項は図1を参照)

音韻環境による 接続の場合分け	素片境界近傍の 音韻環境	接続点の探索		
		探索区間		探索基準
		素片A	素片B	
[場合Ⅰ] 無声子音部での接続	$a_2, b_2$ が無声子音 $a_2 = b_2$	$la_2$	$lb_2$	パワー最小(無声破裂音) 低周波数パワー最小(無声摩擦音)
[場合Ⅱ-1] 有声音全体での接続	$a_2, b_2$ が有声音 $a_1 = b_1$ かつ $a_2 = b_2$	$la_1, la_2$	$lb_1, lb_2$	ケプストラム距離最小
[場合Ⅱ-2] 母音部での接続	$a_1 = b_1$	$la_1$	$lb_1$	ケプストラム距離最小
[場合Ⅱ-3] 有声子音部での接続	$a_2, b_2$ は有声音 $a_2 = b_2$	$la_2$	$lb_2$	ケプストラム距離最小
[場合Ⅱ-4] 同一母音連続の接続	$a_1 = b_2$	$la_1$	$lb_2$	ケプストラム距離最小
[場合Ⅲ] その他	$a_1 \neq b_1$ かつ $a_2 \neq b_2$ かつ $a_1 \neq b_2$	$la_2$ $la_1, la_2$	$lb_2$ $lb_1, lb_2$	パワー最小(無声破裂音) 低周波数パワー最小(無声摩擦音) スペクトル変化率最大(その他)

b<sub>2</sub>の音韻カテゴリ

接続歪



X: 音韻ラベルによる素片境界での接続

Y: 本手法による接続

図3 b<sub>2</sub>の音韻カテゴリ別の接続歪

接続歪

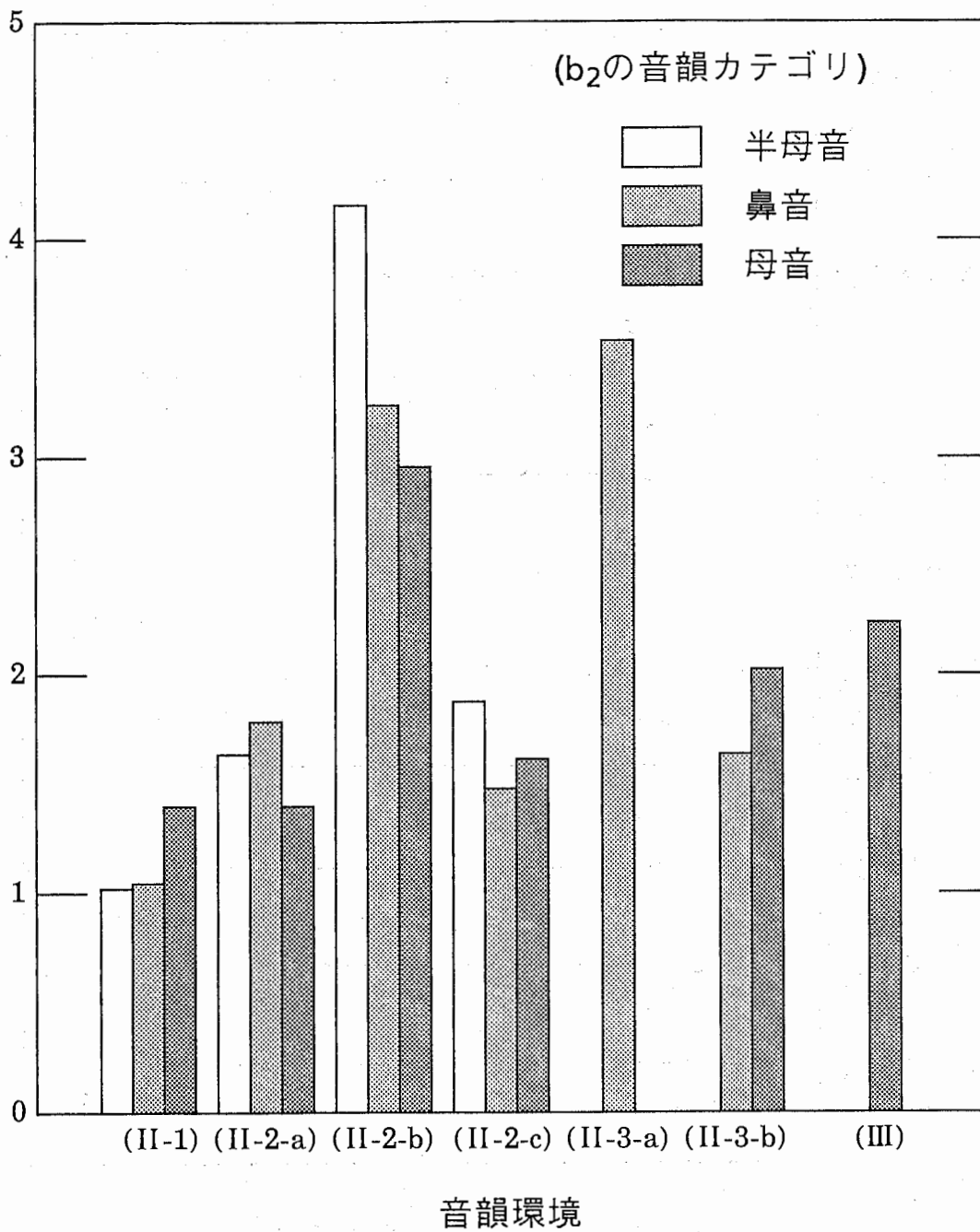
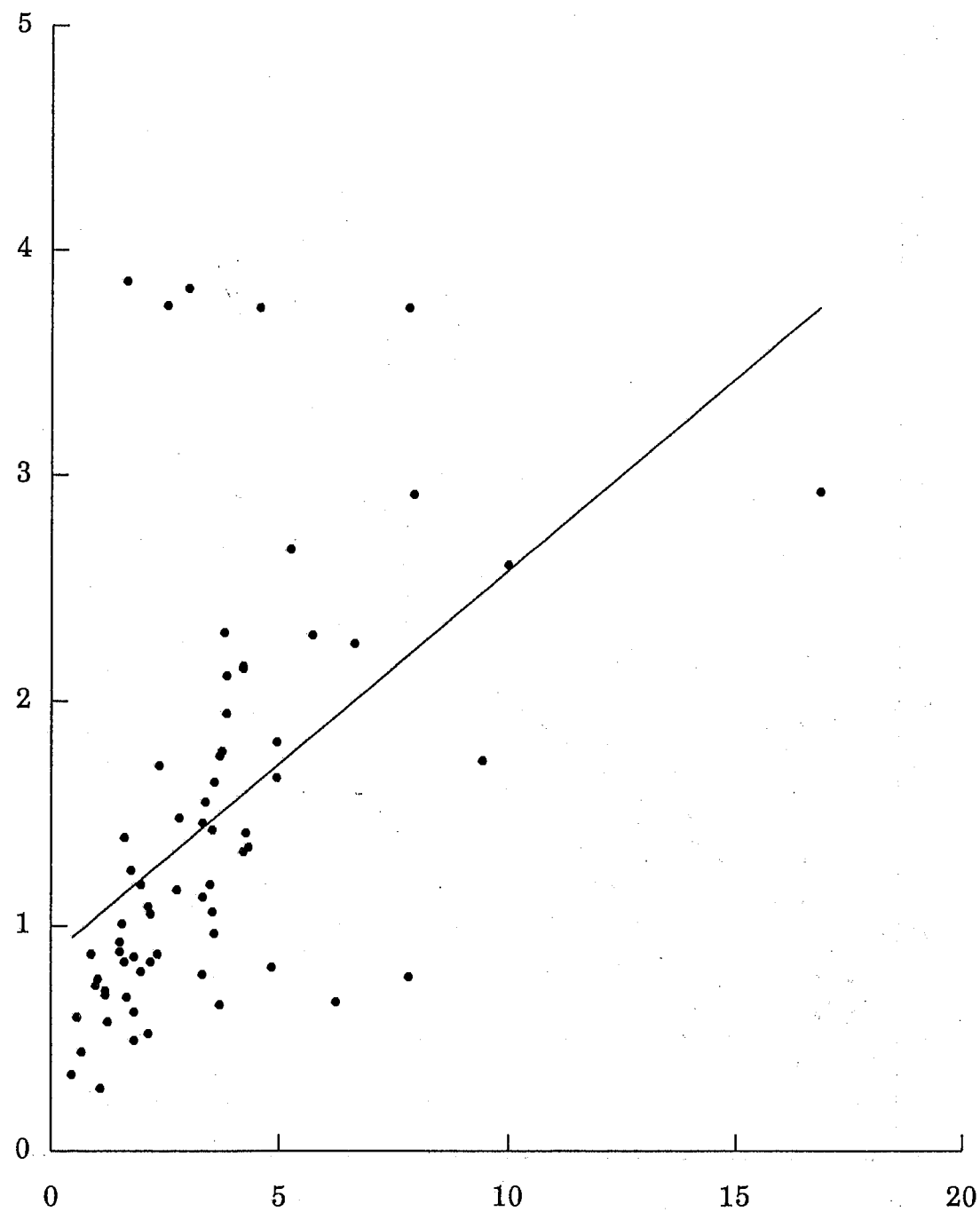


図4 素片境界近傍の音韻環境と接続歪

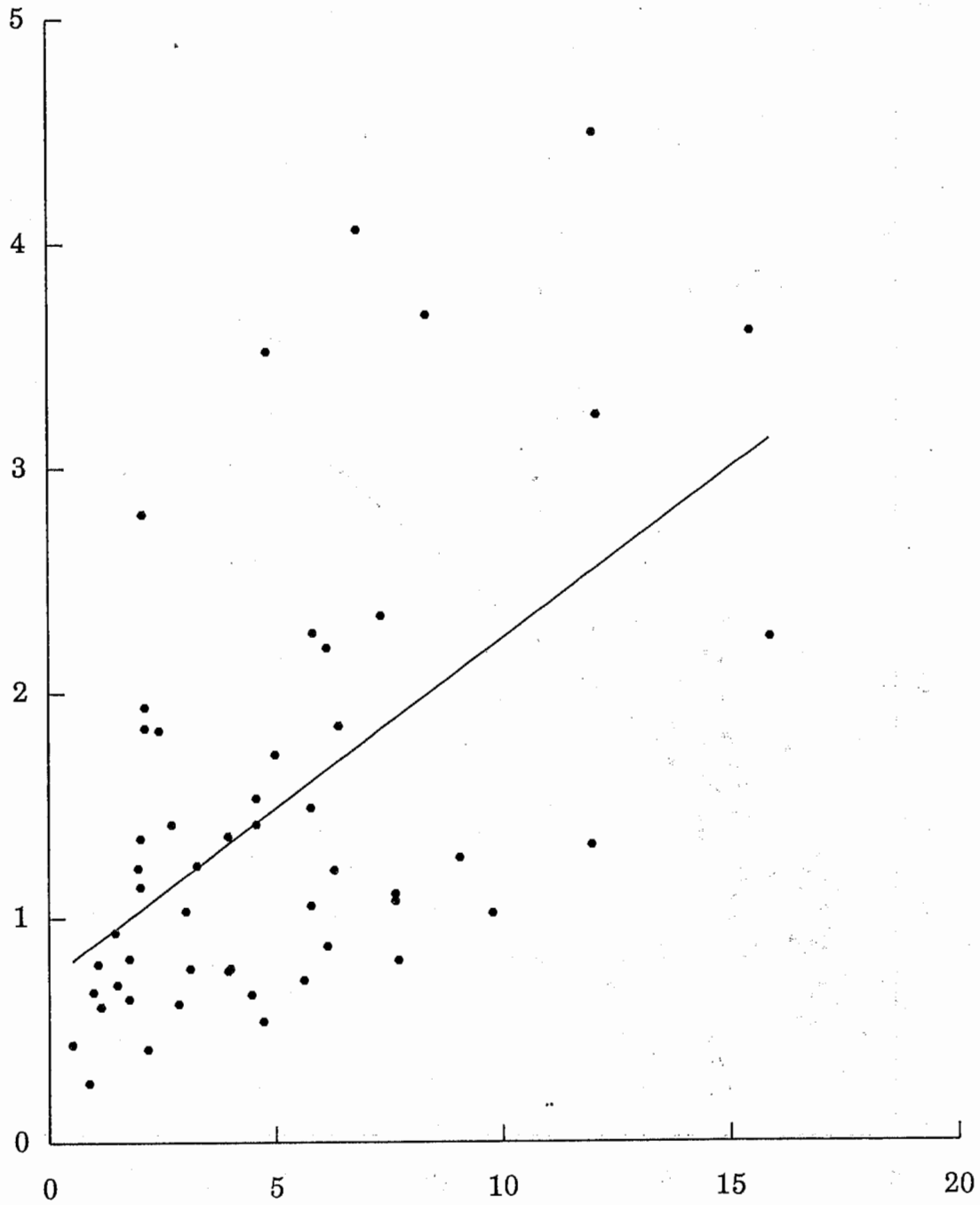
本手法による接続歪



音韻ラベルによる素片境界での接続歪

図5  $b_2$ が半母音の場合の接続歪の相関

本手法による接続歪



音韻ラベルによる素片境界での接続歪

図6  $b_2$ が鼻音の場合の接続歪の相関