

日本語アクセントの範疇知覚検証実験における 端点の正規化について —/aka/を例に—

梁 辰

1. はじめに

範疇知覚の検証実験は人間の言語音に対する知覚様式を明らかにする有効な実験手法である。アメリカ合衆国 Haskins 研究所の研究チームが英語の単音の知覚実験に応用してから、単音レベルにとどまらず、韻律レベルの研究にもしばしば用いられている(稲田, 2013 等)。範疇知覚検証実験を実施する際、基となるのは刺激端点(endpoint)を両端とする刺激連続(stimulus continuum)である。Liberman 他(1957)では、母音の第1フォルマントを一定とし、第2フォルマントの開始部分を急激な上昇から急激な下降まで 120Hz 刻みで変化させることによって、/ba/と/ga/を端点とする刺激連続が合成された。刺激連続の端点をすべて合成によって作成する場合、このように一定の基準を設けることで理想的な刺激を作ることができるが、音声刺激の自然度を保つために人が発話した録音を端点とし、刺激連続を再合成(resynthesize)する際に必ず生じるのは、端点となる録音がどれほどの一般性を持つかという問題である。本稿ではこれを略して「端点の正規化問題」と称することにする。

端点の正規化問題は、日本語アクセントを対象とした範疇知覚の検証実験を実施する際にも生じる。例えば、飴「アメ(0型アクセント)」と雨「アメ(1型アクセント)」を端点として、ピッチが「飴」から「雨」に徐々に変わる刺激連続を再合成することができるが(北原・田嶋, 2011)、発話者によって端点となる「飴」と「雨」のピッチ曲線の形が変わり、また、同一の発話者であっても発話する度に録音語のピッチ曲線が異なることから、1回の録音で抽出したピッチ曲線は必ずしも一般性があるものとは限らない。

一般性を欠くピッチ曲線を端点としては、結果の信頼性を保証できない。この危険を避けるために、録音協力者に複数回発話させる方法が考えられるが、その場合、複数の録音のうち、どの2つを端点として選定すればよいか問題となり、端点に一般性があると言えるだけの客観的な基準が求められることになる。この問題について、本研究では、まず、同一の発話者が複数回発話した同一の語のピッチ曲線が発話によってどの程度変わるかを検証し、端点の正規化問題の重要性を明らかにする。次に、一般性を持つ端点の選定方法の策定を試み、選定した端点の正規化修正の方法について検討する。最後に、選定した録音語をそのまま端点として作成した刺激連続と、正規化修正を経た録音語を端点として作成した刺激連続を用いて、範疇知覚の同定実験を実施し、正規化修正の必要性について述べる。

2. 資料語と録音

2.1 資料語

資料語には、/aka/を1型アクセントで読み上げたもの(「赤」)と、2型アクセントに読み上げたもの(「垢」)を用いる(秋永, 2010)。これらを資料語として用いる理由は次のとおりである。1)「赤」と「垢」はアクセント型のみで弁別されるミニマルペアである。2)前後の母音が同じであり、母音がピッチに及ぼす影響は同一であると考えられる。3)VCVという単純な構造を持ち、中央の子音以外の音によって影響を受ける可能性はない。4)/a/は無声化しにくいいため、中央の子音が無声破裂音であっても、ピッチ曲線が観察できる。5)母音と子音の境界が明瞭である。6)2つの語共にアクセントの妥当性と親密度が高い(天野・近藤, 1999)。

2.2 録音協力者

発話によるピッチ曲線の変動を検証するため、20年以上放送局のアナウンサーを務め、音声学の知識かつ日本語教師としての経験がある成人女性に発話を依頼した。

2.3 録音手順

「赤」と「垢」の他に0型または2型と1型の2モーラの語を多数用意し、キャリアセンテンスは用いずに、まず0型または2型のものを、次に1型のものを読み上げさせ、1通りの録音を

終了した。録音は合計3通り行った。録音協力者による1通りの録音のうち、1資料語を3回以上読み上げたこともあるが、雑音が入ったものを除き、録音されたすべての語を分析対象とした。なお、後続成分がない限り、2モーラにおける2型アクセントは0型アクセントと同様にアクセント核が現れず、また後述の選定基準と正規化修正の方法が0型にも通用できるという意味から、「0型」と表記することにする。

録音は防音室において、コンデンサーマイク (AKG, C4000B) とリニア PCM レコーダー (OLYMPUS, LS-100) によって wav 形式 (サンプリング周波数 44.1 kHz, 量子化ビット数 16 bit) で行い、ノートパソコン (NEC, PC-LZ550LS) 上で音声処理ソフトウェア Praat (Boersma & Weenink, 2014) を用いて処理と計測を行った。

3. ピッチ曲線の変動

録音語のピッチ曲線は Praat によって自動抽出されたものを用いる。ピッチレンジは 100Hz~400Hz で、タイムステップは 7.5ms に設定した。抽出した際に生じた誤り (例えば、無声音/k/のどこかで突然ピッチが検出されたり、語末母音のきしみ声に近い音質により、異常に高いピッチが検出されたりするなど) を除き、一切修正を加えていない。

図1に「垢」と「赤」の合計 19 個の録音語(「垢」10 個, 「赤」9 個)のピッチ曲線を一覧にして示す。図中、左が0型の「垢」で、右が1型の「赤」である。縦軸はピッチ (正確には基本周波数, 単位は Hz) で横軸は時間を示す。観察しやすくするため、/aka/の第1母音 (V1) の終わりと第2母音 (V2) の始まりがそろるように、すべての録音語のピッチ曲線を整列している。よって、それぞれのピッチ曲線の左半分は V1 の、右半分は V2 のピッチ曲線であり、中央の空白は無声破裂音/k/(C)に相当する。なお、この整列方法によって、Cの長短の違いが示されなくなるが、無声破裂音は基本的にピッチを担わないことから、本研究への影響は無視できると考える。

図1によってV1とV2のピッチ曲線を観察すると、以下の2点が見える。第1に、録音語によってピッチ曲線の傾向(上昇か平坦かそれとも下降か)が類似している。しかし、録音語ごとにピッチ曲線の長さが異なるため、ピッチを検出したポイント(以下、ピッチポイントという)の数も異なる。ピッチポイントの数が均一でないことから、10 個(「垢」)ないし 9 個(「赤」)の録音語の V1 と V2 の各ピッチポイントにおけるピッチの値を平均化することによってピッチ曲線を正規化した場合、作成されたピッチ曲線が録音語のピッチ曲線を必ずしも反映でき

なくなる。

図2に、図1に示すピッチ曲線の各ピッチポイントにおけるピッチの平均値で作成したピッチ曲線を示す。

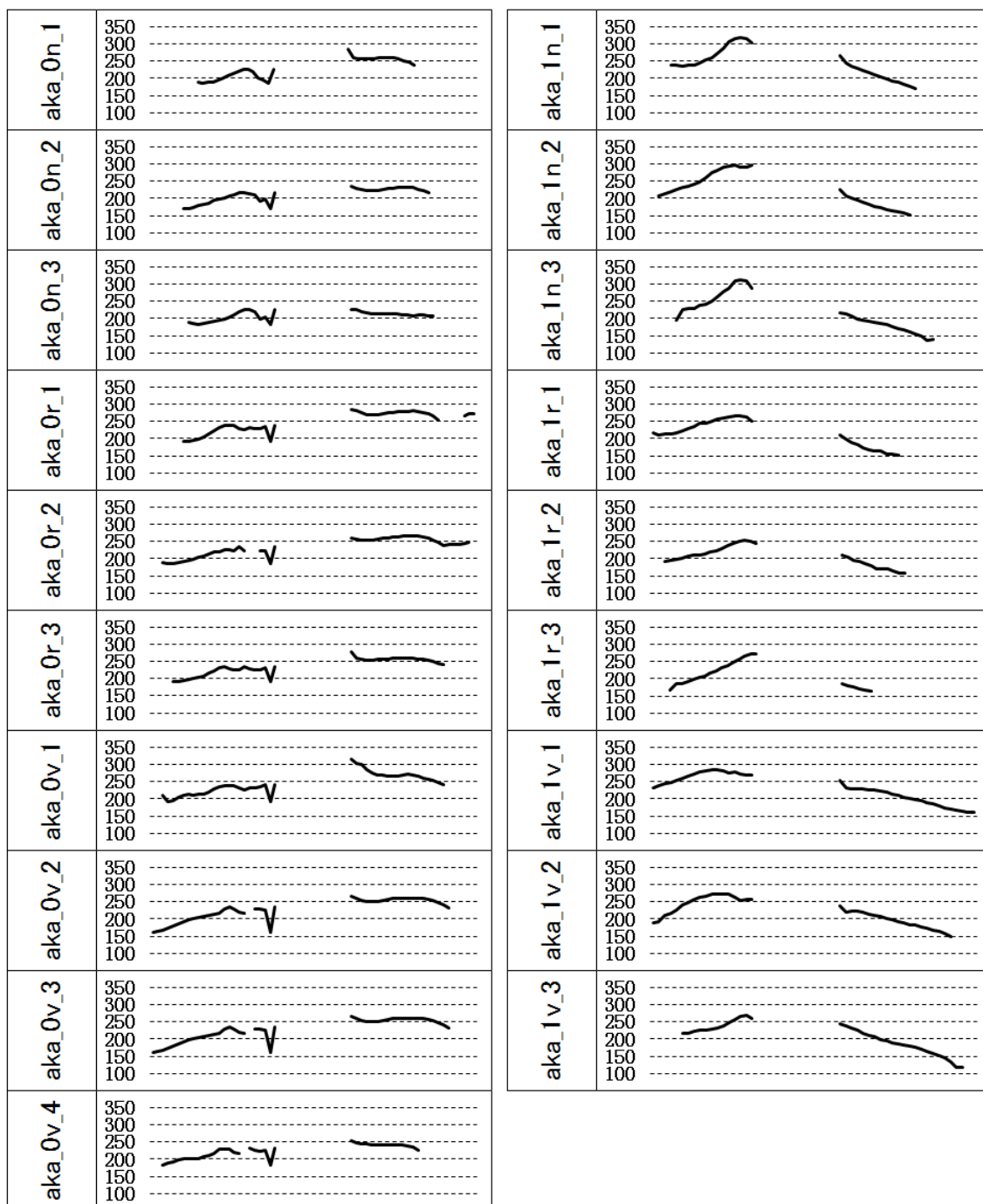


図1. 「アカ」のピッチ曲線の一覧.

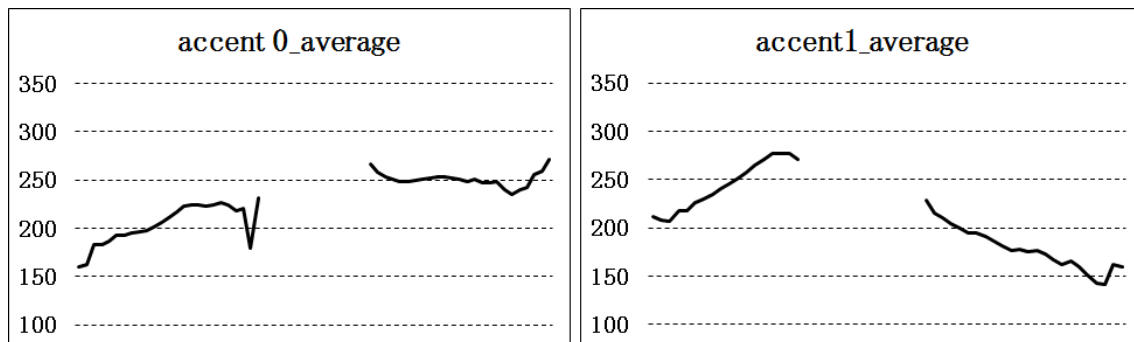


図2. 平均で作成したピッチ曲線.

左側は0型で、右側は1型である。これらのピッチ曲線を録音語から得られたピッチ曲線(図1)と比較すると、明らかに2点で異なる。第1に、平均値によって作成したピッチ曲線は0型、1型ともにV1、V2が長い。これは、録音語から得られたすべてのピッチポイントを用いているからである。このようなピッチ曲線を録音語に付与し、刺激語を合成しようとする時、ピッチ曲線が録音語の持続時間を超えてしまい、ピッチポイントの付与ができなくなる場合がある。第2に、平均値で作成したピッチ曲線の形状を観察すると、全体的には録音語から得られたピッチ曲線に類似しているが、細かなところに違いが見られる。例えば、0型、1型のV2のピッチ曲線の末尾にどちらも顕著な上昇があるが、録音語のピッチ曲線のいずれにもこのような上昇はない。また、0型のV1のピッチ曲線の開始部分に急な上昇が見られるが、これも録音語のピッチ曲線に見られない。

図1に示すV1とV2のピッチ曲線を観察してわかることの第2は、0型のV1とV2、1型のV1とV2にあるピッチ曲線の形状の特徴である。0型では、全体的にV1が上昇し、V2が平坦を保つ。また、V1の最高点がV2より高くなることはない。一方、1型では、全体的にV1が上昇し、V2が下降する。0型と1型を比較すると、V1は0型が1型を上回ることはなく、V2は0型が平坦で1型が下降である。

日本語のアクセントはモーラ間の相対的な高さによって決まる。また、アクセント核を有するモーラに後続するモーラが下降ピッチであることにより、それに先行するモーラにアクセント核があると知覚される(杉藤, 1997)ことから、0型におけるV1とV2の相対的な高さ、1型におけるV1とV2の相対的な高さ、及びV2の下降傾向がアクセント知覚において重要であると考えられる。よって、上記のような、ピッチポイントごとに平均値を取る方法と異なり、0型におけるV1とV2の平均ピッチの差(指標①)、1型におけるV1とV2の平均ピッチの差(指標②)、1型におけるV2の下降の傾斜度(指標③)を基に録音語の選定と正規化を行う可能性を追求したい。

4. ピッチ曲線の正規化の試み

4.1 計測

/aka/は前後に母音, 中央にピッチ曲線が空白となる無声破裂音を配置する音声構造であることから, V1とV2のピッチ曲線それぞれについて平均ピッチを測定するのは容易であるが, 1型におけるV2の傾斜度の計測方法については説明が必要である. 図3では, 録音語 aka_1n_1 の V2 を例として傾斜度の算出方法を示す.

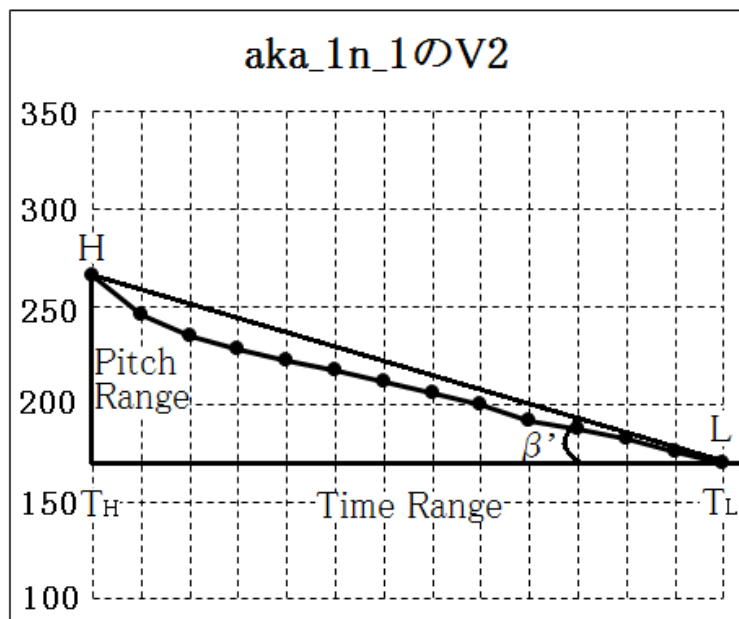


図3. 1型 V2 の傾斜角度.

図中, Hはピッチの最高値, Lはピッチの最低値, T_H はピッチの最高値の出現時点, T_L はピッチの最低値の出現時点を示す. HとLの縦軸上の差を Pitch Range (単位:Hz), T_H と T_L の横軸上の差を Time Range (単位:ms) とする. Pitch Rangeの対角は $\angle \beta'$ である. ピッチ曲線の傾斜角度は $\angle \beta'$ の補角で, $\angle \beta$ と称する(図には示していない).

図3の Pitch Range と Time Range は実測値で, $\angle \beta'$ の補角である $\angle \beta$ は式(1)によって算出する. また, 角の \cos 値は $90^\circ \sim 180^\circ$ の区間ではマイナスの値となり, その変化は $0 \sim -1$ の間にあってわかりやすいことから, 1型のV2の傾斜度を $\angle \beta$ そのものではなく, $\angle \beta$ の \cos 値を使い, 「r」と記す(後述の表1を参照). r値が 0 ($\angle \beta = 90^\circ$)に近ければ近いほど, ピッチ曲線の下降が急であり, -1 ($\angle \beta = 180^\circ$)に近ければ近いほど, ピッチ曲線の下降が緩やかである.

$$\angle \beta = 180^\circ - \arctan \frac{\text{Pitch Range}}{\text{Time Range}} \quad (1)$$

4. 2 計測の結果

図1に示した 19 個の録音語(0型 10 個, 1型 9 個)を対象に, アクセント型別に V1 と V2 の平均ピッチを求め, 両アクセント型の V1 と V2 の平均ピッチの差(DIFF)と, 1型の V2 の傾斜度($\cos \beta$, r と記す)を算出する. 結果は表1のとおりであり, V1, V2, DIFF0, DIFF1の値は小数部分を四捨五入した値を示す.

表1. 録音語の V1 と V2 の平均ピッチ, 平均ピッチの差, 1型の V2 の傾斜度.

0型	MEAN PITCH (Hz)			1型	MEAN PITCH (Hz)			SLOPE
	V1	V2	DIFF0		V1	V2	DIFF1	r
aka_0n_1	204	257	-53	aka_1n_1	270	210	61	-0.71
aka_0n_2	195	227	-32	aka_1n_2	259	181	78	-0.77
aka_0n_3	202	213	-11	aka_1n_3	261	179	82	-0.81
aka_0r_1	220	273	-53	aka_1r_1	240	173	67	-0.80
aka_0r_3	216	256	-40	aka_1r_2	222	180	42	-0.82
aka_0r_2	209	255	-45	aka_1r_3	222	174	49	-0.89
aka_0r_4	211	241	-30					
aka_0v_1	221	271	-50	aka_1v_1	266	202	64	-0.88
aka_0v_2	203	254	-51	aka_1v_2	246	195	51	-0.85
aka_ov_3	203	254	-51	aka_1v_3	239	184	55	-0.77
AVERAGE	208	250	-42	AVERAGE	247	186	61	-0.81

表中, DIFF0 は 0 型の V1-V2, DIFF1 は 1 型の V1-V2 の値である.

0型において V1 が V2 に比べ, 平均ピッチが低く, その差(指標①)はおよそ 42Hz である(表の最下行, AVERAGE の欄に示す). 一方, 1型アクセントにおいて V1 は V2 に比べ, 平均ピッチが高く, その差(指標②)はおよそ 61Hz である. V2 の傾斜度(指標③)の平均値は-0.81 である.

4. 3 端点に使う録音語の選定

録音語の V1 と V2 のピッチ計測を基に、指標①, ②, ③の平均値が得られた。指標①, ②, ③の平均値に近い録音語を端点とすることで録音語全体を代表し、一般性を有する端点が確保されるが、選定に当り、各指標の優先順位が異なる。指標③は1型におけるV2の傾斜度を示す指標であり、1型アクセントの知覚において最も重要な役割を果たす(杉藤, 1997)。また、傾斜度はピッチの継時的な変化であるため調整が最も困難である。従って、指標③を、指標①と指標②に優先して、端点に用いる録音語の選定の第1の基準とする。表1を見ると、指標③の平均的な傾斜度は-0.81 であり、それに最も近い録音語は aka_1n_3 である。よって、録音語 aka_1n_3 を1型の端点として選定する。

指標③によって1型の端点を選定した後、次に指標①により、0型の端点を選定するのが合理的ではあるが、ここで現実的な問題を解決しなければならない。刺激連続を作成する際に、端点となる刺激に一般性がなければならないのは確かであるが、端点となる1対の語のピッチ曲線におけるピッチポイントの数が近いことが望ましい。両端点のピッチ曲線におけるピッチポイント(7.5ms 間隔)の数が大きく異なれば、ピッチポイントの対応をとることができなくなり、中間段階の刺激語を再合成することが困難になる。従って、指標③によって選定された1型の端点となる録音語のピッチ曲線におけるピッチポイントの数に最も近いピッチポイントを持つことが第2の基準となる。紙幅の関係で0型の各録音におけるピッチポイント数の詳細は割愛するが、測定の結果、録音語 aka_0n_3 の V1 と V2 におけるピッチポイントの数が1型端点の aka_1n_3 のそれに最も近いので、これを0型の端点として選定する。表2に、録音語 aka_0n_3(0型端点), aka_1n_3(1型端点)の V1 と V2 の平均値とその差(DIFF), 全資料語の V1 と V2 の平均値とその差を示す(Modification については下記)。なお、V1, V2, DIFF の値は小数部分を四捨五入して整数部分のみを示す。

表2. 端点と修正値.

0型	MEAN PITCH (Hz)			1型	MEAN PITCH (Hz)			SLOPE
	V1	V2	DIFF		V1	V2	DIFF	r
aka_0n_3	202	213	-11	aka_1n_3	261	179	82	-0.81
AVERAGE	208	250	-42	AVERAGE	247	186	61	-0.81
Modification	+7	+37			-13	+8		

表2から aka_0n_3(0型端点), aka_1n_3(1型端点)の V1 と V2 のピッチの平均値は、全録

音語の平均値と差があることがわかる。従って、両録音語をそのまま端点としては、端点に一般性が確保されたとは言えない。そこで、指標①と指標②に基づき、選定された録音語に対する調整を行うことが必要となる。本研究では、aka_0n_3(0型端点)のV1を+7Hz, V2を+37Hz, aka_1n_3(1型端点)のV1を-13Hz, V2を+8Hz, 全体的に移動する(+は上昇, -は下降を示す)。この調整を「正規化修正」と呼ぶことにする。

4.4 正規化修正前の刺激連続と正規化修正後の刺激連続

端点に正規化修正を行っていない刺激連続のピッチ曲線と、正規化修正を行った刺激連続のピッチ曲線を図4に示す。正規化修正前の刺激連続における各ステップをB_ST1～B_ST10と、正規化修正後の刺激連続における各ステップをA_ST1～A_ST10と呼ぶことにする。なお、正規化修正に際して、1型端点と0型端点のV1の長さをそろえるために、0型端点のV1の末尾では2つのピッチポイントを削除したが、V2においてはピッチポイントの数が一致しているため、ピッチポイントの削除はなかった。

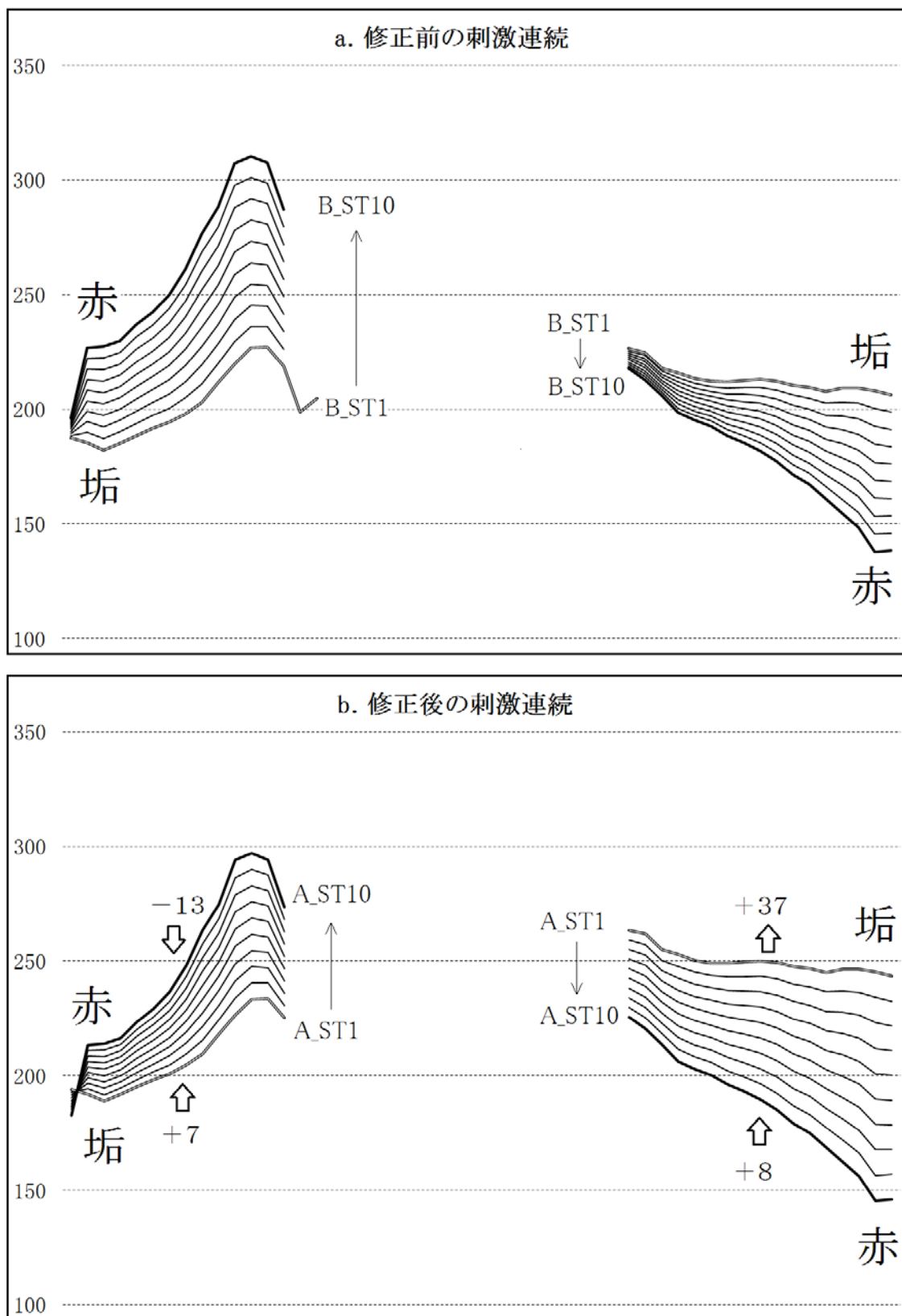


図4. 正規化修正前後の刺激連続のピッチ曲線.

5. 正規化修正の必要性

図4に正規化修正前後の刺激連続における各ステップのピッチ曲線を示しているが、正規化修正が果たして必要であるかはまだ検証する必要がある。端点の正規化修正の必要性を検証するため、日本語母語話者1名(成人男性、音声学知識あり)を対象に、正規化修正前後の刺激連続を用いて同定実験を行った。

5.1 刺激語の再合成

Praat を用いて PSOLA (Pitch Synchronous Overlap and Add) 法によって刺激語を再合成した。再合成する際に、ピッチレンジを 100Hz～400Hz で、タイムステップを 7.5ms に設定した。ピッチ曲線をそれぞれ1型端点と0型端点に付与し、1型をベースとする刺激語と0型をベースとする刺激語を作成した。筆者の聴覚判断で0型ベースの再合成語は自然度がより高いため、0型ベースの合計 20 個の再合成語(修正前 10 個, 修正後 10 個)を同定実験の刺激語として用いることとした。

5.2 範疇知覚の同定実験

同定実験は防音室においてノートパソコン(NEC, PC-LZ550LS)上で行った。図5はパソコン上の実験画面のイメージ図である。実験時、実験協力者はヘッドホン(audio-technica, ATH-A500)を着用した。実験画面の背景は薄い灰色で左上に現在の問題番号と問題の総数を示す数字が示される。画面の上部中央に「Choose the word that you heard」(聞いた語を選んでください)の指示が掲げられ、その下に「赤」(左)と「垢」(右)のボタン(黄色の背景に赤い文字)がある。実験協力者には二者択一の強制選択をさせ、次に自分の判断にどれほどの確信度を持っているかを「1 (unsure)」から「5 (sure)」までの尺度で回答させた。音声は1刺激語につき1回のみ再生されるが、ボタンの押し間違いがあった場合、画面の右下にある「Oops」ボタンを押すことで、再度音声を聴き、判断をやり直すことができるようにした。

実験前に実験データ使用同意書への署名を求め、実験協力者自身に関するアンケート調査を行い、実験内容について説明した。次に、1刺激語を1試行とし、全部で 20 個の刺激語をランダムにまとめた1ブロックを用いた練習を行った。最後に、5ブロック、100 試行の同定実験を行った。実験の所要時間は合計約 40 分であった。

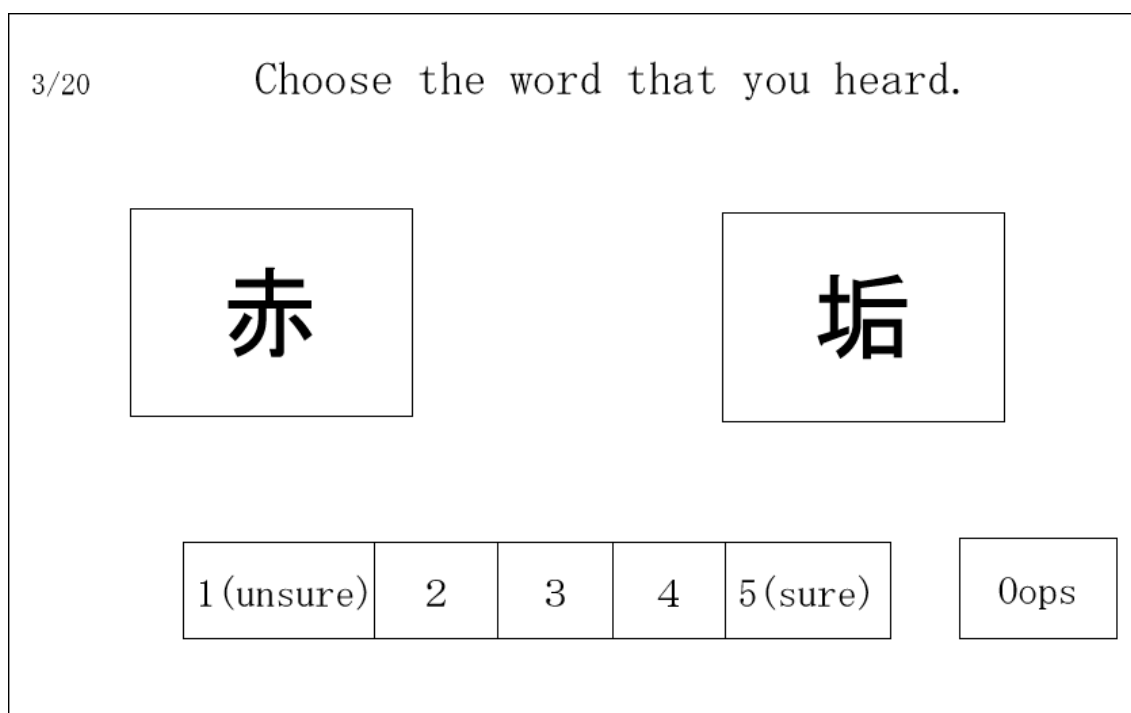


図5. 実験時のパソコン画面のイメージ.

5.3 同定実験の結果と考察

同定実験の結果を図6に示す. 横軸は刺激語の番号で, 縦軸はパーセンテージである. 図6に示す各刺激語が「垢」ないし「赤」に同定される割合と確信度を観察すると, 正規化修正前の刺激連続を用いた同定実験の結果では, B_ST2と B_ST5にかけて「垢」から「赤」への同定の切り替えが現れる. また, B_ST3とB_ST4における判断の確信度が他の刺激語より低い. これらのことから, 正規化修正前の刺激連続においては「垢」と「赤」の範疇境界が B_ST2とB_ST5の間にあることがわかる. 一方, 正規化修正後の場合, 「垢」から「赤」への同定の切り替えは, A_ST4と A_ST7の間に現れ, A_ST6 における判断の確信度が最も低く, A_ST5と A_ST7 においても確信度がやや低い. ただし, A_ST7 に関しては 100%の割合で「垢」と同定されたことから, やや低い確信度は同定判断を揺るがすほどのものではないといえる. 同定判断の結果から, 正規化修正後の刺激連続においては範疇境界が A_ST4 と A_ST7 の間に現れることが確認できる.

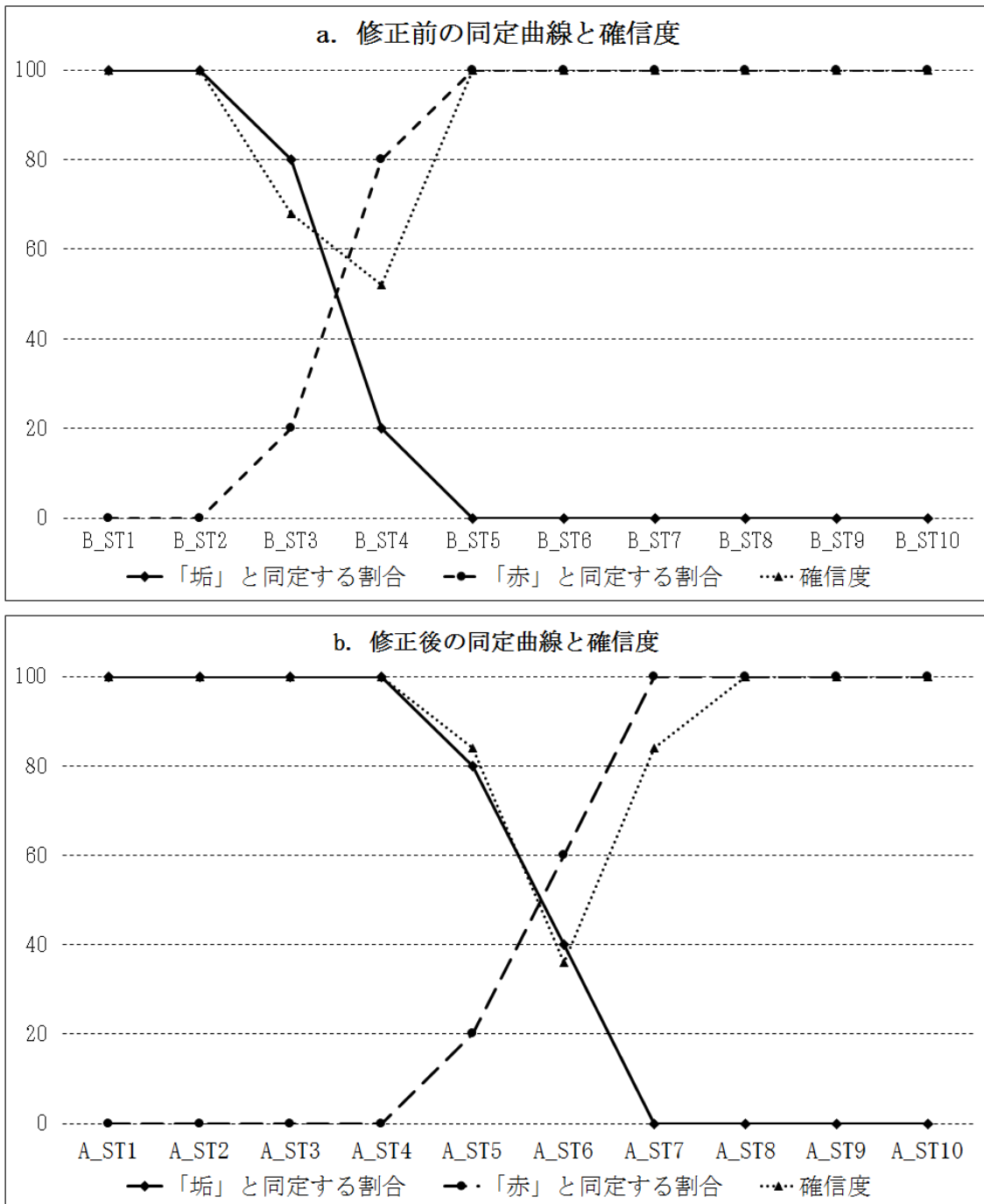


図6. 同定実験の結果.

横軸は刺激語の番号で、縦軸はパーセンテージを示す。実線(ひし形マーカー)は各刺激語が「垢」と同定される割合で、破線(丸マーカー)は各刺激語が「赤」と同定される割合を示す。一方、破線(三角マーカー)は各刺激語における判断の確信度を示し、確信度の尺度1~5は、縦軸の20~100%に対応する。

範疇知覚の同定実験では、結果の分析において同定曲線の3つのパラメータが考察されることが多い。境界位置(boundary position)、境界傾斜度(boundary slope)、及び境界幅

(boundary width)である(加藤他, 2004). 境界幅と傾斜度は, 正規化修正の前後で大きな違いはないが, 境界位置は, 正規化修正前のほうが修正後に比べ, 左に偏っていることがわかる.

なぜベースと端点に用いた音声に違いがないのに, 正規化修正することで同定曲線の境界位置に違いが生じたのか, その原因を明らかにするため, 違いが生じた正規化修正前後の刺激連続におけるステップ1から6までの V1 と V2 の平均ピッチの差(指標①と指標②), 及び V2 の傾斜度(指標③)を表3にまとめる.

表3. B_ST1~B_ST6 と A_ST1~A_ST6 の比較.

	B_ST1	B_ST2	B_ST3	B_ST4	B_ST5	B_ST6
V1	201.36	207.93	214.50	221.07	227.65	234.22
V2	212.99	209.16	205.33	201.50	197.67	193.84
DIFF	-11.63	-1.23	9.17	19.58	29.98	40.38
r	-0.99	-0.98	-0.96	-0.95	-0.93	-0.91
	A_ST1	A_ST2	A_ST3	A_ST4	A_ST5	A_ST6
V1	208.00	212.35	216.70	221.04	225.39	229.74
V2	250.07	242.98	235.88	228.79	221.70	214.60
DIFF	-42.07	-30.63	-19.19	-7.75	3.70	15.14
r	-0.99	-0.98	-0.96	-0.95	-0.93	-0.91

表中, B_ST1~B_ST6 は正規化修正前, A_ST1~A_ST6 は正規化修正後のステップ1~6の刺激語である. V1 は第1母音の平均ピッチ, V2 は第2母音の平均ピッチ, DIFF は V1-V2 によって算出した平均ピッチの差を示す(指標①と②に相当). rは第2母音の傾斜度を示す(指標③に相当).

表3にも示されるとおり, 正規化修正後の刺激連続における各ステップの V2 の傾斜度は, 正規化修正前と変わらない. 端点に対する正規化修正は, 両端点の V1 と V2 に対する全体的な上昇ないし下降であるからである. 従って, 正規化修正による範疇境界の位置変化は, 指標③に由来するものではない. これに対して, V1 と V2 の平均ピッチを比較すると, V2 のほうが高い(つまり DIFF が負の値をとる)のは, 正規化修正前はステップ1, 2のみであるが, 修正後はステップ1から4までとなっている. 前述のとおり, 日本語のアクセントはモーラ間の相対的な高さであり, O型録音語のピッチ曲線を観察すると(図1左), V1 の最高点が V2

より高くないことが特徴的である。以上のことから、修正前の刺激連続では、より若い番号の刺激語から V1 の平均ピッチが V2 のそれを上回るために、範疇境界が左に偏ることとなったと考えられる。

同定実験における範疇境界の位置は、異なる実験協力者群に異なる知覚傾向(例えば、高さを聞きわけると、母語によって、語用論的範疇を用いるか言語学的範疇を用いるか)をもたらす可能性があることから(Chan, Chuan & Wang, 1975), 刺激連続による範疇境界の偏りは望ましくない。端点に対する正規化修正を行わずに実験をすることで、刺激連続による範疇境界の位置が影響を受ける恐れがある。本稿が提案した正規化修正の方法を用いることにより、この問題が改善されることが期待できる。

6. おわりに

本研究は日本語アクセントを対象とした範疇知覚の検証実験を行う際に生じる端点の正規化問題を取り上げた。1名の録音協力者によって発話された複数回の録音語のピッチを計測し、ピッチ曲線の変動を確認した上で、平均値を取る方法では一般性を備えた刺激端点を得られないという問題が解決できないことを示し、端点の正規化問題の重要性を明らかにした。次に、日本語アクセントの知覚特徴を考慮した3つの指標を考案し、複数の録音語から端点とする語を選定する方法、及び選定した端点に対する正規化修正を行う方法を提案した。最後に、同定実験を実施し、正規化修正の必要性和有効性を検証した。今後の課題として、選定基準と正規化修正の汎用性の問題が考えられる。つまり、今回策定した3つの指標が、/aka/のような前後に母音、中央にピッチ曲線が空白となる無声破裂音を配置する音声構造を持たない語へも応用できるかどうかは課題となる。

[謝辞]

本論文の執筆にあたり、中国国家留学基金(China Scholarship Council, CSC)の援助を受けていることを記す。

[参考文献]

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer (Version 5.4.69).

<http://www.praat.org/> (2014/09/11)

- Chan, S. W., Chuan, C.-K. & Wang, W. S.-Y. (1975). Cross-linguistic study of categorical perception for lexical tone. *Journal of the Acoustical Society of America*, 58 (S1), S119.
- Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.
- 天野成昭・近藤公久(1999). 『NTTデータベースシリーズ日本語の語彙特性』三省堂.
- 稲田朋晃(2013). 「ソウル方言話者と慶尚道方言話者による日本語アクセント核のピッチ知覚:知覚の離散性に注目して」『音声研究』17 (1), 6-15.
- 加藤宏明・田嶋圭一・アマンダ ロスウェル・山田玲子・ケビン マンホール(2004). 「母語話者と非母語話者による日本語特殊拍音素の知覚」『電子情報通信学会技術研究報告. SP, 音声』104 (149), 43-48.
- 北原真冬・田嶋圭一(2011). 「音声分析ソフトウェアPraatを用いた聴取実験:F₀再合成による刺激作成と実験の制御」『日本音響学会誌』67 (8), 345-350.
- 秋永一枝(編) (2010). 『新明解日本語アクセント辞典』金田一春彦(監修), 三省堂.
- 杉藤美代子(1997). 『音声波形は語る』和泉書院.