

有声音の寸法知覚に対する聴覚計算モデル

瀧本 恵理^{1,a)} 入野 俊夫^{1,b)} 松井 淑恵^{1,c)} Roy D. Patterson^{2,d)}

概要: 人間の聴覚系では寸法情報と形状情報を分離抽出しているという理論が提案されている。これに対し、理論を支持する寸法弁別の心理物理実験が得られている。また、スペクトル傾斜の影響についても無声音を用いて知覚実験が行われ、計算モデル化で説明できている。最近、有声音についても同様な知覚実験が行われ、無声音と同様な結果が得られることがわかった。しかし、無声音に対するモデルでは、この有声音の結果を説明できないことがわかった。ここでは、安定化ウェーブレットメリン変換による寸法形状分離の理論に基づいた重み (SSI weight) を導入した。これにより、有声音に含まれる調波構造の影響を低減させた。また、スペクトル傾斜補正の度合いを調整することにより、寸法弁別の手がかりの個人差を説明できることを示した。

キーワード: 寸法弁別, スペクトル傾斜, 計算モデル, dcGC, 安定化ウェーブレットメリン変換

A computational model of speaker size perception for voiced speech sounds

TAKIMOTO ERI^{1,a)} IRINO TOSHIO^{1,b)} MATSUI TOSHIE^{1,c)} ROY D. PATTERSON^{2,d)}

Abstract: An auditory model was developed to explain the results of behavioral experiments on perception of speaker size with voiced speech sounds. It is based on the dynamic, compressive gammachirp (dcGC) filterbank and a weighting function (SSI weight) derived from a theory of size-shape segregation in the auditory system. Voiced words with and without high-frequency emphasis (+6 dB/octave) were produced using a speech vocoder (STRAIGHT). The SSI weighting function reduces the effect of glottal pulse excitation in voiced speech, which, in turn, makes it possible for the model to explain the individual subject variability in the data.

Keywords: speaker size perception, spectral tilt, computational model, dcGC, stabilized wavelet-Mellin transform

1. はじめに

人間は音声を聞いたとき、男性や女性、子供では声道長 (vocal tract length, VTL) が大きく異なるのに関わらず、同じ母音であるとわかる。また同時に、それが同じ話者だ

とわかる。Iriano and Patterson [1] は、聴覚系において音声から声道形状に関する音響的特徴をどのように分離しているかを説明するための計算理論とアルゴリズムを提案した。このアルゴリズムでは話者の寸法を正規化することで、寸法の相違に対して不変な形状を音声の内部表現として生成している。これはウェーブレット変換、イメージ安定化、メリン変換からなる安定化ウェーブレットメリン変換 (SWMT) である。

この理論の裏づけとなる人間を対象とした聴取実験では、STRAIGHT[2], [3] を自然発声に対する声道長の操作に使用している。声道長の伸縮は、その逆数となる平均

¹ 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama University

² CNBH, Department of Physiology, Development, and Neuroscience, University of Cambridge

a) s185063@center.wakayama-u.ac.jp

b) irino@sys.wakayama-u.ac.jp

c) tmatsui@sys.wakayama-u.ac.jp

d) rdp1@cam.ac.uk

フォルマント周波数 (mean formant frequency, MFF) を操作することにより実現した。STRAIGHT を用いた先行研究により、人間は有声音でも無声音でも寸法弁別ができることが示された [4], [5]。実験では、話者の寸法の弁別閾 (just noticeable difference, JND) は、母音に対しては声道長 (または MFF) で約 7%、音節や単語に対しては MFF で約 5% であることがわかっている。例えばラウドネスの JND が約 11% であることを鑑みると、かなり小さい JND で感度が高い知覚現象であると言える。これらの研究では、たとえ話者の寸法が通常発声の範囲外であったとしても、音声認識の成績には影響がなく頑健であることも示されている。

最近、Yamamoto ら [6] は 2 種類の雑音駆動の音声を用いて、寸法知覚に対するスペクトル傾斜の影響を明らかにする実験を行った。ひとつはオリジナルの有声音と同じスペクトル包絡をもつ無声音 “Unvoiced” で、もうひとつはオリジナルの有声音に比べて +6 dB/Oct の高域強調を施した無声音 “Whispered” である。刺激名の由来は、後者は前者よりささやき声に似ているためである。Unvoiced と Whispered による話者の寸法判断実験を行い、心理物理曲線を適合した (心理物理曲線の例として図 3 を参照)。その結果、スペクトル傾斜の影響は被験者個人に依存することが明らかになった。心理物理曲線がスペクトルの高域強調によってシフトした被験者もいれば、シフトしない被験者もいた。

この結果を説明するために、Yamamoto ら [6] は、SWMT[1] の初段のウェーブレット変換に基づいた寸法弁別の計算モデルを構築した。このモデルでは、動的圧縮型ガンマチャープ (dynamic compressive GammaChirp, dcGC) フィルタバンク [7], [8], [9], [10](dcGC-FB) を用いて音声の聴覚スペクトログラムをまず生成する。dcGC-FB のブロックダイアグラムを図 1 の上半分に示す。これは、受動的ガンマチャープフィルタバンク (passive GC-FB), 高域通過非対称フィルタ群 (HP-AF), 各 HP-AF の位置を制御するためのレベル推定ユニットの 3 つの要素で構成されている。図 1 の中央のパネルに、6 つのチャンネルの 7 音圧レベルに対する合成フィルタの周波数特性を示す。フィルタバンクは 100-6000 Hz の範囲で、準対数軸である健聴者の矩形等価帯域 (ERB_N) 軸に沿って均等に配置された 100 チャンネルからなる。dcGC-FB の各チャンネルの出力の時間波形に、25 ms の窓長、5 ms のシフト長のハニング窓をかけて RMS 値を計算し、図 1 の下パネルにあるような聴覚スペクトログラムを出力した。母音のスペクトログラムは、高域強調処理をした音声では高い周波数帯域に相対的に大きなエネルギーを持ち、このことが一部の被験者の寸法弁別過程に影響したと Yamamoto ら [6] は議論している。さらに、被験者ごとの違いも含めて説明できる聴覚計算モデルも提案した。

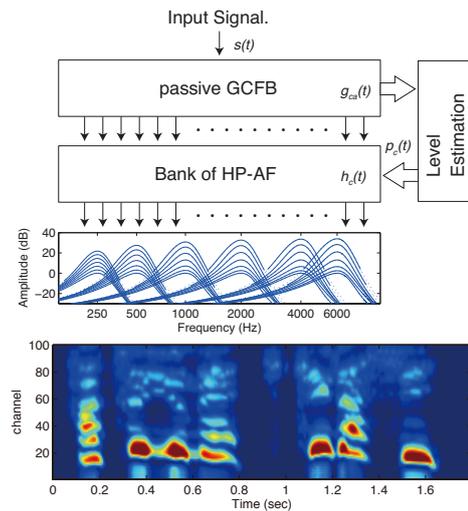


図 1 dcGC-FB から算出された聴覚スペクトログラム (下図)。上半分は、dcGC-FB の 6 つのチャンネル分の、7 つの刺激音圧レベルの時のフィルタ形状と利得の変化を示す。刺激レベルが 90 dB から 30 dB に 10 dB ステップで下がるにつれ、フィルタのゲインは上昇する。

これらの実験とモデルの結果を受けて、松井ら [11] は通常の有声音とそれに対して +6 dB のスペクトル傾斜のある有声音を用いて Yamamoto らと同様の寸法弁別実験を行った。その結果、無声音を用いた実験と同様にスペクトル傾斜の寸法弁別への効果は個人に依存することがわかった。ところが、Yamamoto らが開発した Unvoiced と Whispered 音声の寸法弁別の計算モデルを有声音刺激に適用したところ、実験結果を説明することはできなかった。有声音における声帯振動が、聴覚スペクトログラム上に調波成分のピークを生じさせる。さらにこのピークがフォルマント成分に対応するピークよりも大きく、声帯振動周波数に依存して大きく変化することがモデルの予測に悪影響を与えていると考えられる。

本論文では、SWMT モデル [1] の第 2 段階であるイメージ安定化を取り込むことで、Yamamoto らの寸法知覚モデルを拡張し、有声音にも対応する寸法知覚モデルを構築することを目指した。有声音の寸法弁別実験の結果をモデルで説明することを行ったので報告する。自動音声認識研究における声道長正規化に関して、本論文と同様の有声音に関する問題を持っている。本論文における寸法知覚における解決法は、今まで用いられてきた声道長正規化アルゴリズムにも適用可能である。

2. 寸法弁別実験

有声音を対象として、スペクトル傾斜が寸法弁別判断に及ぼす影響を明らかにするための実験を行った。詳細は同時発表の研究会資料 [11] を参照されたい。

2.1 使用した音声データ

自然に発話された 4 モーラ単語音声を収録した親密度別音声データベース FW03[12] から、平均基本周波数が約 150 Hz の男性発話者 1 名 (ラベル名 mya) の、高親密度セット 4 と 3 を原音声として用いた。

2.2 合成音声の声道長伸縮における条件

スケール変形音声の合成には TANDEM-STRAIGHT [13] を用いた。音声合成は 3 つの手順で行った。(1)TANDEM-STRAIGHT を用いて原音声から平滑化した STRAIGHT スペクトルを求めた。(2) 求めたスペクトルの周波数軸を伸縮し、平均フォルマント周波数 (mean formant frequency, MFF) を操作した。MFF と声道長 (Vocal Tract Length, VTL) はほぼ逆数の関係にあるので、等価的に声道長の伸縮を行ったことになる。(3) 抽出した基本周波数 (glottal pulse rate, GPR) を変形させたパルス波で STRAIGHT スペクトルを駆動させることで再合成した。基本周波数は原音声に対して 0.5, 1, 2 倍したものを再合成に用いた。また、時間微分により +6 dB/Oct のスペクトル強調処理を行った。スペクトルの強調処理を行っていないものを “Original” (以下 “Or”), 強調処理を施したものを “Emphasized” (以下 “Em”) とした。基準となる音は常に Or とし、比較音は Or と Em を均等に割り当て、提示順をランダムにした。Or-Or と Or-Em の 2 種類の刺激対による 2 区間 2 肢強制選択法を用いた。実験は図 2 に示す GPR-MFF の 5 つの領域について行った。

2.3 実験結果

図 3 は, Or-Or と Or-Em のそれぞれ 8 人の被験者の結果の平均に対して適合した心理物理曲線を 5 つの領域に分けて示している。横軸は原音声に対する MFF 比、縦軸は比較音を小さい話者と判断した割合を示す。累積ガウス分布をブートストラップ法で適合させて、表示している心理物理曲線を求めた。Or-Or の心理物理曲線 (赤) は、傾斜が急で偏りが無い (基準音の MFF に対してグラフの中央に曲線が位置している)。Or-Em の心理物理曲線 (青) は、Or-Or に比べてわずかに傾斜が緩やかである。また、図からわかるように心理物理曲線の中心が Or-Or に比べて MFF の値が小さい方へ偏っている。Or-Em の心理物理曲線が MFF 比の小さい方にシフトしているということは、高域を強調した音声の方が小さい話者と判断されているということであり、先行研究 [6] の結果と同様である。

弁別閾 (JND) は「比較音が小さい」と回答した割合が 50% から 76% に増加したときの MFF ratio の増分である。Or-Or の結果 (赤) では、JND は 5.2-5.8% であり、平均すると 5.5% である。これは有音声を用いた先行研究 [14] の JND の値とほぼ一致する。これに対して、Or-Em の結果 (青) では、JND は 4.5-7.9% で、平均が 5.9% と Or-Or の場

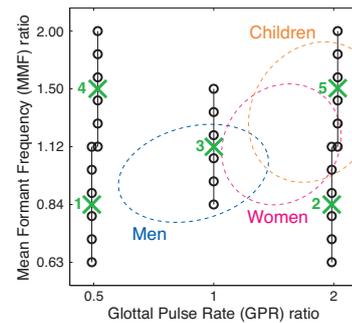


図 2 寸法弁別実験で提示された有音声の GPR-MFF の組み合わせ。5 つの基準音を ×印で示し、6 つの比較音を白抜きで示す。破線の楕円は通常の音声の GPR および MFF の範囲を示す。

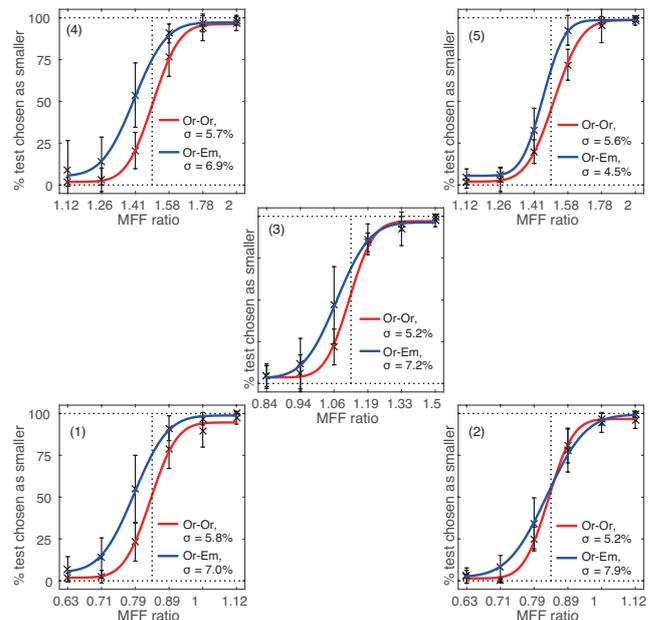


図 3 寸法弁別実験の 8 名の結果平均に適合した心理物理曲線を 5 つの基準音ごとに示した。基準音ごとに Original-Original (Or-Or, 赤) と Original-Emphasized (Or-Em, 青) のそれぞれの心理物理曲線を示す。エラーバーは被験者間の ± 1 標準偏差。弁別閾 (JND, σ) は比較音が小さいと選択された比率が 50% から 76% に増加したときの MFF の増分を百分率で示したものである。

合に比べて少し大きい。Or-Em の結果が Or-Or に比べて被験者間でばらつきがあるのは、何人かの被験者はスペクトル強調に影響を受けなかったのに対して、他の被験者は影響されたためである。

3. 寸法知覚の聴覚計算モデル

Yamamoto らの寸法知覚計算モデル [6] を、前節で述べた実験で用いた有音声に適用して寸法予測をしたところ、GPR が 0.5 の場合の心理物理曲線は再現できたが、GPR が 1 および 2 の場合の結果は説明することができなかった。そこで、寸法知覚の SWMT モデルの 2 番目の段階であるイメージ安定化と、イメージ安定化により生成する音

声の表現，すなわち寸法形状イメージ (size-shape-image, SSI) を用いて，有声音の寸法弁別実験結果も説明できるようなモデル拡張を試みる．

3.1 音声スペクトログラムの操作

イメージ安定化は，繰り返しのある時間間隔パターンを強調するピッチ同期式の時間積分である．有声音はウェーブレット変換の出力の段階で周期性を持っているため，これに同期して聴覚イメージ (Auditory Image, AI) を生成する (イメージ安定化の詳細については，先行研究 [1] を参照のこと)．

この聴覚イメージを，“スケールシフト共変”の寸法-形状イメージ (Size-Shape Image, SSI) に変換する．SSI の例として母音 /o/ の場合を図 4(a) に示す．これは，AI における“時間間隔軸”の横軸から，フィルタバンク・チャンネルのピーク周波数と時間間隔の積 h で表される横軸に変換することによって SSI は計算される．この計算は，声帯振動の一周分ごとに計算され，次の周期分は切り捨てられる．結果として，イメージの右下の境界より下の領域には活性度は存在しない．この境界は基本周波数で変化する．寸法情報の処理を容易にするイメージの領域は，図 4(a) の境界 (カーブした実線) の上部に収まる．Yamamoto らの寸法知覚モデル [6] の基礎となるスペクトログラムは境界下の活動の除外を考慮していなかった．これが比較的高い GPR でモデルが機能しない原因となっていたと考えられる．このため，GPR の低次の高調波に関連するスペクトログラムの領域での活動を減少させる重み付け関数を導入した．この関数を図 4(b) に示す．基本周波数が F_0 の場合，周期は $1/F_0$ であり，重み付け関数 (以下，“SSI weight”) w_{SSI} は次式で求められる．

$$w_{SSI}(e_f) = \frac{\min(f_p(e_f)/F_0, h_{max})}{h_{max}} \quad (1)$$

$f_p(e_f)$ は ERB_N 軸上の値 e_f の dcGC 聴覚フィルタのピーク周波数であり， h_{max} はピーク周波数と時間間隔の積 h の最大値で，SSI の水平方向の範囲を指定する．計算を簡単にするため，音声において時間と共に変動する基本周波数 F_0 は定数値 $F_0^{(lim)}$ に固定した．この値は， w_{SSI} によって重み付けされる領域の限度 (図 4(a) の境界線) を規定する．

図 5 に，聴覚スペクトログラムの算出に w_{SSI} を適用した場合の，単語のスペクトログラムを一对示す．上部スペクトログラムは高域強調した場合，下部スペクトログラムは高域強調していない単語を示している．スペクトログラムを比較すると，上部スペクトログラムは高い周波数領域が低い周波数領域よりも強調されていることがわかる．

3.2 寸法判断の手順

寸法知覚モデルによる寸法弁別の心理物理曲線の計算では，被験者実験と全く同様に 2 区間 2 肢強制選択法を用い

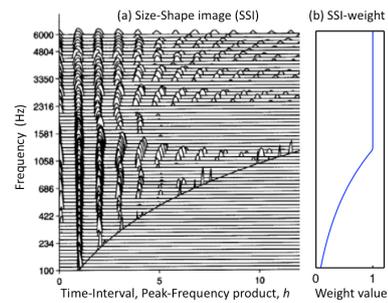


図 4 SSI weight モデル．(a) 母音 /o/ の寸法形状イメージ ([1] より改変)．(b) SSI が有効な領域に基づく SSI weight ．

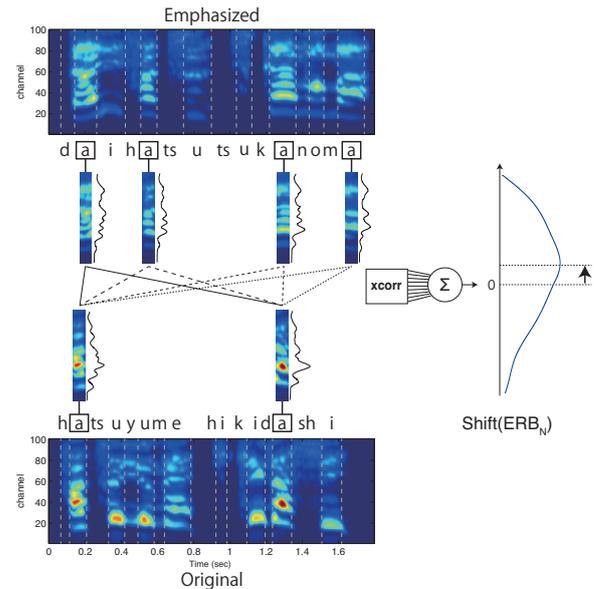


図 5 2 区間で示された音声に対する聴覚スペクトログラムに基づく相互相関モデル．最初の区間 (上パネル) は Emphasized 音声の“daihatsu”と“tsukanoma”である．2 番目の区間 (下パネル) は，Original 音声の“hatsuyume”と“hikidashi”である．

る．各試行において，寸法弁別の被験者実験で被験者に提示された単語と全く同じ単語を処理した後に，より小さな話者を含む区間をモデルが判断する．この手順で得られた結果によって，図 3 に示すような心理物理曲線をモデルの弁別結果に対しても得ることが可能になる．

3.3 MFF シフト量の算出

図 5 は，寸法判断の手順を説明している．被験者実験でもモデルでの実験でも，一試行の 2 区間ではそれぞれ単語が 2 つずつ呈示される．その 2 単語の音声から，重みを付与した聴覚スペクトログラムを計算した．音声データから自動音声認識プログラムである Julius[15] を用いて母音区間を検出し，母音区間中心から ± 25 ms を抽出した．この ± 25 ms スペクトログラムから RMS 値を計算し，聴覚的スペクトル - エキサイテーションパターン E_p - を算出した．これの計算を全ての母音 $v = \{/a/, /e/, /i/, /o/, /u/\}$ で行った．次に，図 5 の母音 /a/ の例のように，母音ごと

に、区間 1 における全ての母音に対する E_p と、区間 2 に出現する同じ母音全てと各々相互相関をとった。一つの対の両区間の相互相関関数は次式で表される。

$$R_{Ep12}(e_s) = \sum_{i=-N+1}^{N-1} Ep1(e_{f_i})Ep2(e_{f_i} + e_s) \quad (2)$$

e_{f_i} は ERB_N の値、 e_s は ERB_N 軸上の中心からのシフト量、 N はチャンネル数を示す。比較音が基準音よりも小さい話者の場合、この相互相関関数のピークは平均して 0 を上回る値にシフトする。そのため、“シフト量のピーク”は話者の寸法を決定するための妥当な変数である。母音ごとにスペクトル特徴は違うため、まず同一母音ごとの組み合わせ (index k) での相互相関関数をすべて計算し、それをすべての母音 (index v) で統合した。

$$R_{Epvk}(e_s) = \frac{R_{Ep12,v,k}(e_s)}{\sqrt{R_{Ep11,v,k}(0) \cdot R_{Ep22,v,k}(0)}} \quad (3)$$

$$\Phi(e_s) = \sum_{v=a/k}^{u/k} R_{Epvk}(e_s) \quad (4)$$

図 5 の右側に、統合相互相関関数 $\Phi(e_s)$ の分布の例を示す。 ERB_N の単位でシフト量のピークは相関軸の中心から上にシフトしている。ピークのシフト量によって、実験の各試行における刺激に対する被験者応答を予測する。シフト量のピーク S_{Ep} は次のように求められる。

$$S_{Ep} = \arg \max_{e_s} \Phi(e_s). \quad (5)$$

この S_{Ep} が正の話者区間を小さい話者として選択した。 $S_{Ep} = 0$ の場合は小さい話者の区間をランダムに選択した。これにより、比較音を小さい話者と判断する選択率を計算した。後は、被験者実験と全く同じように心理物理曲線を生成し、被験者実験結果と比較した。

3.3.1 異なるスペクトル傾斜での個人差の表現

被験者実験における Or-Em 条件の心理物理曲線のシフト量には大きな個人差があった。何人かの被験者はスペクトルの高域強調の効果を無視するか抑圧しているようである。このことに関連した個人差を特徴づけるため、 E_p の傾きに重み付けをすることを考えた。線形回帰から求めた傾きが α_r であるとき、個人パターンである Ep^c は、

$$Ep^c(e_f) = Ep(e_f) - w_r \alpha_r e_f \quad (6)$$

と表現される。ここで重み w_r はその被験者の判断が高域強調に影響される度合いを示す。このように改変した Ep^c から、式 5 の S_{Ep} を計算した。ちなみに、ここで α_r は線形回帰から求めたが、2 次および 3 次の回帰としても、結果にほとんど影響がなかったことを記しておく。

3.4 結果

図 6 は、図 2 で示す GPR-MFF 領域 (5) の心理物理曲線

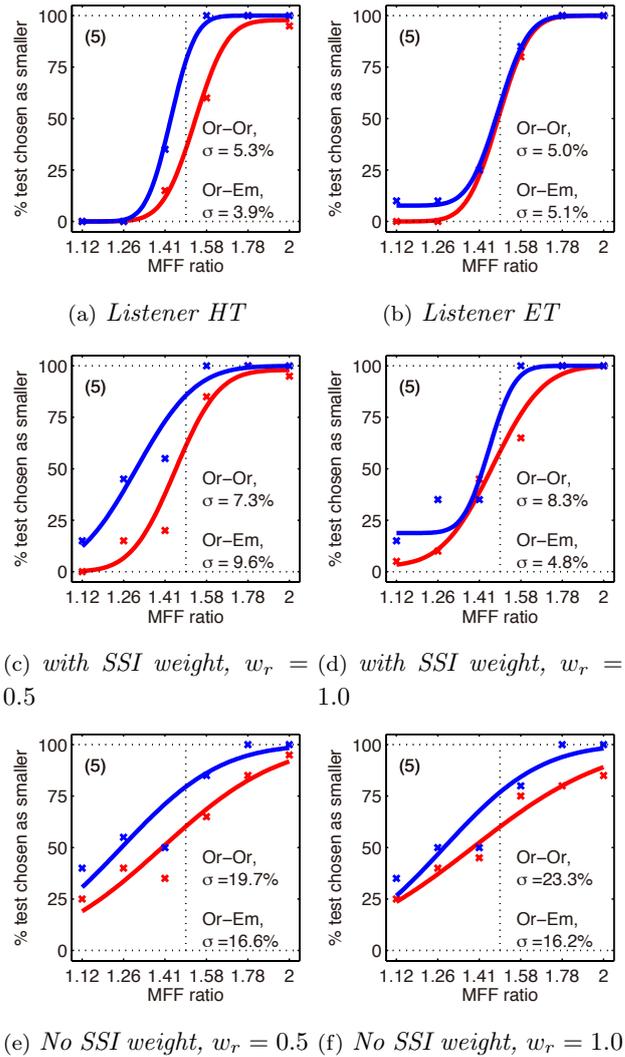


図 6 (a)(b)GPR-MFF 領域 (5) に対する 2 名分の心理物理曲線の例。(c)(d) w_{SSI} を含む場合の SWMT モデルによる心理物理曲線のシミュレーション。(e)(f) w_{SSI} を含まない場合のシミュレーション。赤い曲線は Or-Or 条件、青い曲線は Or-Em 条件を示す。

を示している。(a)(b) は被験者実験の結果から 2 名の被験者 (HT と ET) の心理物理曲線を示す。(c)(d) は w_{SSI} を含む場合の SWMT モデルによるシミュレーションの心理物理曲線で、(e)(f) は w_{SSI} を含まない場合のシミュレーションの心理物理曲線である。被験者実験の結果 (図 6 における一番上のパネル 2 つ) を比べると、(a) の被験者 HT の結果では Or-Em (青) が Or-Or (赤) に比べて左にシフトしているが、(b) の被験者 ET の結果はシフトしていないことがわかる。JND は約 5% である。 w_{SSI} ありでシミュレーションされた心理物理曲線 (図 6 における中段のパネル 2 つ) では、(c) は (a) の被験者 HT と同様の左シフトを示すが、(d) は (b) の被験者 ET と傾向が異なる。JND は被験者 HT では、Or-Or と Or-Em とともに 2-4% 大きい。 w_{SSI} を含まない場合の心理物理曲線 (図 6 における下段のパネル 2 つ) は、無声音を対象に開発された先行研究のモデル [6] と同様のものであり、被験者のデータから直接

生成された心理物理曲線（上段）よりも明らかに傾きがゆるやかである。JND は 16%以上であり，被験者のデータから直接推定された JND である 5%よりはるかに大きい。これは有声音の声帯音源によるエキサイテーションパターン E_p のピークが，フォルマントのピークと区別がつかなくなったためであると考えられる。SWMT の SSI に由来する重み付け関数 w_{SSI} は， E_p における声帯音源の効果を減少させる。図 6 の中段に示す重み付きシミュレーションでは，式 1 における $F_0^{(lim)}$ を 150 Hz に固定した。ただし， $F_0^{(lim)}$ を 200 Hz あるいは 300 Hz に増加させても，心理物理曲線にはほとんど影響がなかった。GPR-MFF 領域 (3) および (5) のように GPR 比が 2 で平均 F_0 が 300 Hz であるときでも，心理物理曲線はほとんど変わらなかった。このことは， w_{SSI} の計算に正確な F_0 推定を必要とせず，有声音の寸法推定が単純な関数でできることを示唆している。

4. おわりに

dcGC を用いて有声音の寸法弁別結果を説明できる寸法知覚モデルを検討した。dcGC-FB は，SWMT モデルの音声処理の最初の段階にあたる [1]。寸法弁別実験のデータより，有声音に含まれる低域の分解された調波成分がフォルマント周波数のピークと混在することで寸法推定が妨げられることがわかった。この問題は，SWMT モデルの第 2 段階の処理を簡素化した SSI weight を導入することで解決できた。これは声帯振動に対応する聴覚スペクトログラムのエキサイテーションのピーク値を低減させる効果がある。様々な信号処理で有声音を対象とした声道長や話者の寸法の推定を行うことが性能向上に結びつく場合がある。SSI weight は簡単なスペクトル重み付け関数であるため，ガンマトーンやメル周波数，1/3 オクターブフィルタバンクなど線形スペクトル分析手法に対しても簡便に適用できるであろう。

謝辞 本研究は JSPS 科研費 JP25280063，JP15H02726 および JP16H01734 の助成を受けた。

参考文献

- [1] Irino, T. and Patterson, Roy D.: Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform, *Speech Communication*, Vol. 36, No. 3, pp. 181–203 (2002).
- [2] Kawahara, H., Masuda-Katsuse, I. and De Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f_0 extraction: Possible role of a repetitive structure in sounds, *Speech Communication*, Vol. 27, No. 3, pp. 187–207 (1999).
- [3] Kawahara, H. and Irino, T.: *Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation*, Pierre Divenyi, Ed. Springer (2005).
- [4] Smith, D. R. R., Patterson, Roy D., Turner, R., Kawa-

- hara, H. and Irino, T.: The processing and perception of size information in speech sounds, *The Journal of the Acoustical Society of America*, Vol. 117, No. 1, pp. 305–318 (2005).
- [5] Ives, D. T., Smith, D. R. R. and Patterson, Roy D.: Discrimination of speaker size from syllable phrases, *The Journal of the Acoustical Society of America*, Vol. 118, No. 6, pp. 3816–3822 (2005).
- [6] Yamamoto, K., Irino, T., Nisimura, R., Kawahara, H. and Patterson, Roy D.: How the slope of the speech spectrum affects the perception of speaker size, *Proc. INTERSPEECH*, pp. 1556–1560 (2015).
- [7] Irino, T. and Patterson, Roy D.: A time-domain, level-dependent auditory filter: The gammachirp, *The Journal of the Acoustical Society of America*, Vol. 101, No. 1, pp. 412–419 (1997).
- [8] Irino, T. and Patterson, Roy D.: A compressive gammachirp auditory filter for both physiological and psychophysical data, *The Journal of the Acoustical Society of America*, Vol. 109, No. 5, pp. 2008–2022 (2001).
- [9] Patterson, Roy D. and Unoki, M. and Irino, T.: Extending the domain of center frequencies for the compressive gammachirp auditory filter, *The Journal of the Acoustical Society of America*, Vol. 114, No. 3, pp. 1529–1542 (2003).
- [10] Irino, T. and Patterson, Roy D.: A dynamic compressive gammachirp auditory filterbank, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 6, pp. 2222–2232 (2006).
- [11] 松井淑恵, 入野俊夫, 山本航大, 河原英紀, Roy D. Patterson: 有声音の寸法知覚における高域強調処理の影響, 第 115 回音楽情報科学研究会資料 (2017).
- [12] 坂本修一, 鈴木陽一, 天野成昭, 小澤賢司, 近藤公久, 曾根敏夫: 親密度と音韻バランスを考慮した単語了解度試験用リストの構築. *日本音響学会誌*, Vol. 54, No. 12, pp. 842–849 (1998).
- [13] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F_0 , and aperiodicity estimation, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3933–3936 (2008).
- [14] Irino, T., Aoki, Y., Kawahara, H. and Patterson, Roy D.: Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination, *Speech Communication*, Vol. 54, No. 9, pp. 998–1013 (2012).
- [15] Lee, A., Kawahara, T. and Shikano, K.: Julius—an open source real-time large vocabulary recognition engine, *The 7th European Conference on Speech Communication and Technology (EUROSPEECH2001)*, pp. 1691–1694 (2001).