

有声音の寸法知覚における高域強調処理の影響

松井 淑恵^{1,a)} 入野 俊夫^{1,b)} 山本 航大¹ 河原 英紀^{2,c)} Roy D. Patterson^{3,d)}

概要: 音声から話者のおおよその寸法(声道長)を知覚することができることがわかっている。これまで、声道長を等価的に伸縮させた音声を用いた寸法弁別実験が行われ、その結果、弁別閾が伸縮率の差で5%程度であることが示されている。また、高域強調のスペクトル傾斜がある場合には、無い場合にくらべて寸法が小さく知覚される傾向があり、その度合いは個人ごとに異なることが、無声音を使った実験で明らかになった。本研究では、有声音を用いて同様の実験を行った。その結果、無声音の場合とほぼ同じ結果が得られた。すなわち、参加者により、高域強調の有無にかかわらずほぼ同一の心理物理曲線になる場合から、高域強調により心理物理曲線がシフトする場合があることがわかった。このことは、声帯振動の有無にかかわらず寸法知覚は行われるが、高域強調の影響は個人ごとに異なることを示している。

キーワード: 寸法知覚, 有声音, 高域強調処理, 心理物理曲線, 主観的等価点, 弁別閾

The enhancing high-frequency components affects size discrimination of voiced speech sounds

MATSUI TOSHIE^{1,a)} IRINO TOSHIO^{1,b)} YAMAMOTO KOUDAI¹ KAWAHARA HIDEKI^{2,c)}
ROY D. PATTERSON^{3,d)}

Abstract: A number of studies, with either voiced or unvoiced speech, have demonstrated that a speaker's geometric mean formant frequency (MFF) has a large effect on the perception of the speaker's size, as would be expected. They revealed that the just noticeable difference (JND) of MFF was approximately 5%. One study with unvoiced speech showed that lifting the slope of the speech spectrum by 6 dB/Oct also led to a reduction in the perceived size of the speaker. This paper reports an analogous experiment to determine whether lifting the slope of the speech spectrum by 6 dB/Oct affects the perception of speaker size with voiced speech (words). The results showed that voiced speech with high-frequency enhancement was perceived to arise from smaller speakers. However, there were large individual differences; some listeners were effectively insensitive to spectral enhancement of 6 dB/Oct; others showed a consistent effect of the same enhancement. The results suggest that models of speaker size perception will need to include a listener specific parameter for the effect of spectral slope.

Keywords: speaker size perception, voiced speech sounds, speech spectrum slope, PSE, JND

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 和歌山大学産学連携イノベーションセンター
Center for Innovative Research and Liaison, Wakayama University

³ CNBH, Department of Physiology, Development, and Neuroscience, University of Cambridge

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

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

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

d) rdp1@cam.ac.uk

1. はじめに

ヒトは音を聞いたとき発音源の寸法と形を分離して知覚することができる。声だけ聞いても相手がどのようなことばを発声しているかがわかり、かつ大人か子供かを推定することも容易である。ヒトの声のみならず、ヴァイオリンとチェロのそれぞれの音を聞いたとき、どちらの楽器が大きいかを間違えることはほとんどない。楽器の大きさの弁別ができるとともに、ヴァ

イオリンとチェロが類似した形状の楽器であることもわかる。このような知覚現象に対して、Iriano and Patterson は音源の寸法情報と形状情報を分離抽出する機能が聴覚系に備わっているとする理論を提案した [1]。

この理論を裏付ける実験的検証の研究は、STRAIGHT[2], [3] が音声を品質良く合成できるようになったことで発展した。有声音と無声音の両方を対象に、スペクトル包絡の伸縮率にもとづく話者の寸法弁別判断の精度が測定された（母音単独、および異なるピッチを持つ母音 4 つからなる音声の寸法弁別 [4], 子音-母音, 母音-子音のシラブルによる寸法弁別 [5], 自然発話された日本語 4 モーラの有聲単語および無聲単語の寸法弁別 [6]）。これらの先行研究によって、通常発声の範囲の内外や有声音無声音を問わず、寸法弁別閾となるスペクトル包絡の伸縮比は約 5% であることが明らかになった。

一方で、同じ寸法のヒトの発話、すなわちフォルマント周波数が同じ音声であっても、有声音よりもささやき声のほうが発話者の寸法が小さく感じられる、という内観報告が有声音と無声音を対象とした寸法弁別実験から得られている [6]。声帯振動を伴わない音声であるささやき声は、波形に周期性を持たない。加えて、有声音に比べて約 +6 dB/Oct のスペクトル傾斜があり、多くの高周波エネルギーを持つ。これはノイズの乱れが声道を励起する際に生じる [7]。ささやき声のほうが小さい寸法の話者に聞こえるという報告は、音声のフォルマント周波数が同じであっても、スペクトル傾斜の違いや周期性の有無によって知覚される話者の寸法が変わる可能性を示唆している。

この内観報告を受けて、Yamamoto らはオリジナルの有声音と同じスペクトル包絡を持つが周期性のない無声化音声 (Unvoiced speech, Unvoiced と表記) と、この Unvoiced を時間微分することにより +6 dB/Oct のスペクトル高域強調処理を行った、ささやき声に似た音声 (Whispered speech, Whispered と表記) の 2 種類を合成し、音声による話者の寸法知覚にスペクトル傾斜の違いが及ぼす影響を調べた [8]。その結果、+6 dB/Oct のスペクトル高域強調処理を施した Whispered は、オリジナルのスペクトル傾斜を持つ Unvoiced よりも平均的に小さい寸法の話者が発したものと知覚された。ただし、スペクトル傾斜の違いが寸法知覚に影響する程度には大きな個人差が見られた。すなわち、比較する音声のスペクトル傾斜の違いが寸法弁別に大きく影響する参加者もいれば、影響しない参加者もいた。これらの個人差は、実験に用いた音声刺激の平均フォルマント周波数の範囲に関係なくほぼ一貫して現れた。この結果から、Yamamoto らはスペクトル傾斜の影響の度合いを個人間で変化させることを許容する寸法知覚モデルを構築し、実験結果を説明した。

Yamamoto らのモデルは無声音の寸法弁別結果をうまく説明できた一方、有声音に対しての有効性は確認されていない。我々が普段聞く音声は、ささやき声ではなく有声音が主である。Yamamoto らの寸法知覚モデルの汎用性を高めるためには、周期性のある有声音でもスペクトル傾斜の違いが話者の寸法知覚に関与するかどうか、ヒトを対象とした寸法弁別実験で確認する必要がある。本研究では Yamamoto らの実験 [8] に倣い、オリジナルの音声のスペクトル傾斜を保持した有声音と、オリジ

ナルの音声に +6 dB/Oct の高域強調処理を付与した有声音の 2 種類の刺激を用いて話者の寸法弁別実験を実施し、有声音による話者の寸法知覚にスペクトル傾斜の違いが及ぼす影響を調べた。

2. 寸法弁別実験

有声音による寸法弁別判断にスペクトル傾斜の違いが与える影響を、ヒトを対象とした実験で測定した。寸法弁別判断の結果に対して心理物理曲線を適合し、主観的等価点 (Point of Subjective Equality, PSE) と弁別閾 (Just Noticeable Difference, JND) を算出した。PSE と JND に対する、高域強調処理の有無の効果、基準音の平均フォルマント周波数比 (mean formant frequency ratio, MFF ratio) と基本周波数に対応する声門パルス比 (Glottal pulse rate ratio, GPR ratio) の組み合わせの効果、これら 2 つの要因の交互作用を検証した。

2.1 刺激

自然に発話された 4 モーラの日本語単語音声 (4 名分収録されている親密度別音声データベース FW03[9]) の音声データを原音声として用いた。この FW03 の単語リストは、単語親密度ごとに 4 段階に分けられ、語頭の音韻バランスおよび語中の音韻バランスも考慮されている。本実験には、男性発話者 (ラベル名 mya) の高親密度 (レベル 4 および 3) の単語リストの音声を用いた。

発話者の相対的寸法と基本周波数、すなわち MFF ratio と GPR ratio を操作する際に目的以外の歪み音が出ないように、高品質な音声分析合成システムである TANDEM-STRAIGHT[3] を用いた。まず、音声から平滑化した STRAIGHT スペクトルと、Glottal pulse rate (GPR) を 5 ms ごとに求めた。STRAIGHT スペクトルの周波数軸に定数を掛けることにより伸縮させ、等価的に声道長の伸縮を行った。GPR も別の定数を掛けることにより基本周波数を変更した。伸縮させた STRAIGHT スペクトルを、変更した GPR のパルス列で駆動することで、実験用音声を得た。

再合成した音声は 2 種類である。ひとつは原音声と同じスペクトル包絡を持つ有声音 (Original Speech, Original と表記) である。もうひとつは、Original を時間微分することにより +6 dB/Oct の高域強調処理を行った強調音声 (Emphasized Speech, Emphasized と表記) である。

2.2 手続き

弁別判断のための基準音と比較音における MFF ratio と GPR ratio の組み合わせを図 1(a) に示す。基準音には、原音の MFF ratio を 1 としたとき $2^{-3/12} (\approx 0.84)$, $2^{2/12} (\approx 1.12)$, $2^{7/12} (\approx 1.50)$ となる 3 種類の MFF ratio と、原音の GPR ratio を 1 としたとき 0.5, 1, 2 となる 3 種類の GPR ratio を用意した。これらの中から MFF ratio と GPR ratio の 5 種類の組み合わせ、(1 ×: $2^{-3/12}$, 0.5), (2 ×: $2^{-3/12}$, 2), (3 ×: $2^{2/12}$, 1), (4 ×: $2^{7/12}$, 0.5), (5 ×: $2^{7/12}$, 2) を基準音の MFF ratio と GPR ratio とした (数字付きの × は、図 1(a) の緑の数字付き × に対応)。これら 5 種類の基準音を 1 とした MFF ratio を

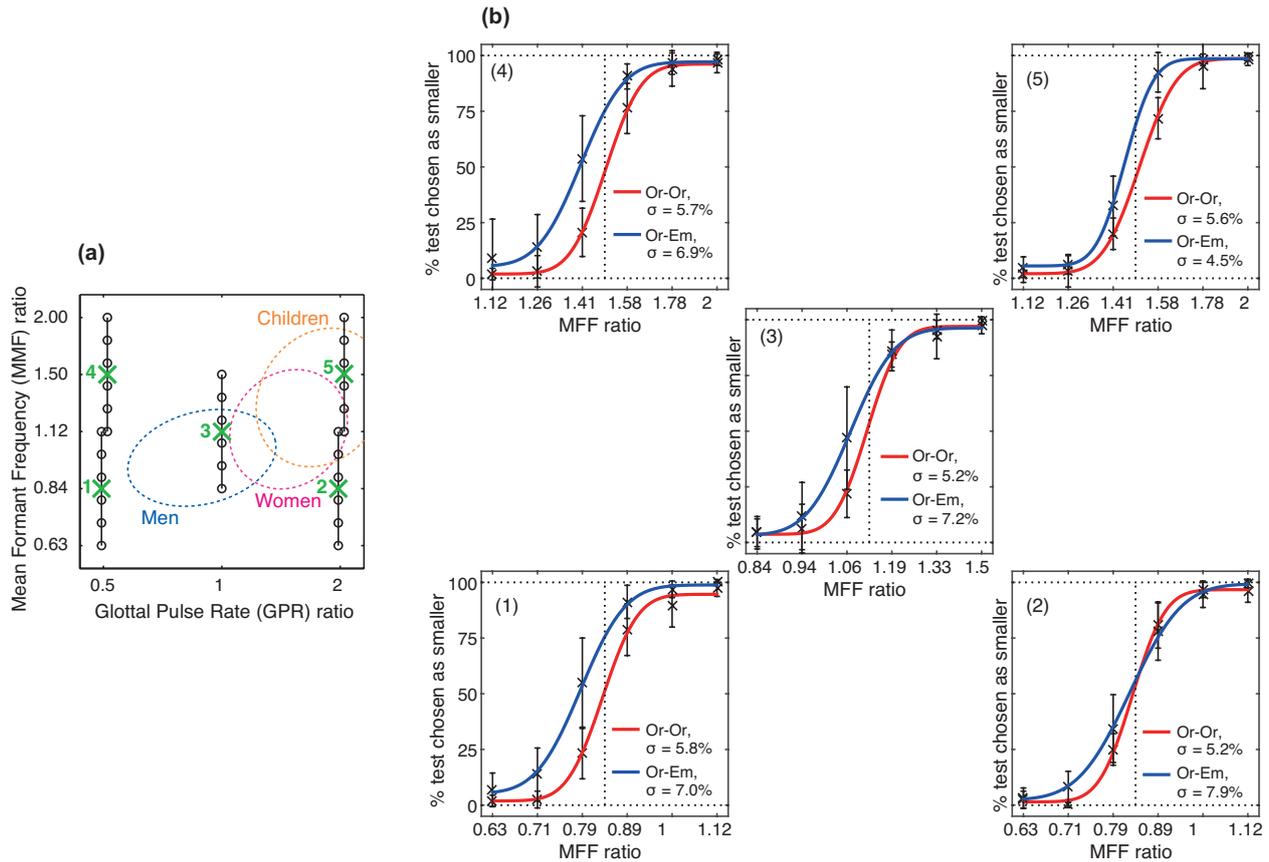


図 1 (a) 刺激の MFF ratio と GPR ratio の組み合わせ。MFF ratio の 1 と GPR ratio の 1 は、再合成前の元の単語の MFF ratio と GPR ratio を示す (話者 mya の MFF は約 1278 Hz, 平均 GPR は約 150 Hz)。番号のついた 5 つの緑の × は基準音を示し、基準音と縦線でつながれた白抜きの × はそれぞれの基準音に対する比較音を示す。破線で示した 3 つの楕円は、成人男性 (Men), 成人女性 (Women), 子供 (Children) の MFF と GPR の平均的分布を示す。(1)(4) ならびに (2)(5) は GPR 軸上でわずかにずれているが、重なりあう部分を可視化するための配置であり、実際の GPR ratio はそれぞれ 0.5 と 2.0 である。(b) 5 つの基準音 ((a) の数字付きの緑の ×) に対して得られた心理物理曲線。8 名の平均に対して累積ガウス分布を適合した。縦軸: 比較音が基準音よりも「小さい話者」と判断された割合 (%)。横軸: 比較音の MFF ratio。各パネル中央に縦方向に走る破線: 基準音の MFF ratio。赤い曲線: 基準音と比較音の両方が Original 刺激である場合 (Or-Or) の弁別結果による心理物理曲線。青い曲線: 基準音が Original, 比較音が Emphasized である場合 (Or-Em) の弁別結果による心理物理曲線。黒い ×: 比較音を「小さい話者」と判断した割合の 8 名の実験参加者の平均。エラーバー: 参加者間の標準偏差 ± 1 。 σ : MFF-GPR 各領域における MFF ratio の JND (%)。

$2^{-5/12}, 2^{-3/12}, 2^{-1/12}, 2^{1/12}, 2^{3/12}, 2^{5/12}$ とした 6 点の比較音を作成した (図 1(a) の白抜きの ×)。

実験は二区間二肢強制選択の恒常法で行った。各試行で、片方の区間にランダムに選択された基準音の 2 単語が呈示され、もう片方の区間には、MFF ratio が基準音より大きい 3 種類と小さい 3 種類、計 6 種類のいずれかからランダムに選択された比較音の 2 単語が呈示された。基準音と比較音の間には 0.5 秒の無音区間を挟んだ。基準音はすべて Original 条件とした。比較音には、基準音とスペクトル傾斜が同じ Original と、+6 dB/Oct の高域強調処理を施した Emphasized がランダムに割り当てられた。基準音、比較音ともに Original の組み合わせを Or-Or、基準音が Original、比較音が Emphasized の組み合わせを Or-Em と表記する。実験参加者は、呈示された 2 区間のどちらが「小さい話者」であったかを GUI 上で回答した。フィードバックは与えられなかった。

試行数は参加者 1 人あたり 1200 試行であった (基準音 (5) ×

比較音 (6) × 比較音の高域強調処理の有無 (2) × 基準音と比較音の呈示順のカウンターバランス (2)=120 試行を 1 セッションとして 10 セッション実施)。1 試行あたり、基準音 2 単語、比較音 2 単語を使用するため、実験で使用した単語数はのべ 4800 単語である。基準音と比較音で全く同じ音声刺激対が呈示されないことがないように、高親密度 (レベル 4 および 3) のリストから単語をランダムに抽出して使用した。

音声のサンプリング周波数は 48 kHz とした。実験は聴力検査室 (リオン, AT62W) で実施し、音声は iMac からヘッドフォンアンプ (FOSTEX, HP-A8) を通して、ヘッドホンスピーカ (Sennheiser, HD-580) で呈示された。呈示音圧レベルは A 特性で平均 70 dB とした。ヘッドホンスピーカはサウンドレベルメータ (Bruël & Kjaer, Type 2250-L) と人工耳 (Bruël & Kjaer, Type 4153) でキャリブレーションを行った。特定の音圧レベルを判断の基準にすることを防ぐため、呈示するすべての音声に対して ± 3 dB の一様乱数で音圧レベルを増減させる

ローピングを行った。

2.3 練習セッション

参加者を寸法弁別判断に慣れさせるため、本実験の前に練習セッションを設けた。本実験と同じ寸法弁別課題で、基準音と比較音が同じスペクトル傾斜の組み合わせ Or-Or のみを用いた。原音声には、本実験とは異なる音声の FW03 の男性発話者 1 名 (ラベル名 mis) の音声を用いた。使用単語は FW03 データベースより高親密度 (レベル 4 および 3) の単語リストとし、各親密度リスト 1000 語からランダムに呈示した。

練習セッションは 3 段階に分けて行った。まず、基準音と MFF ratio の差が大きい比較音 (図 1(a) で、基準音からもっとも遠い白抜きの) を対象としたセッションから開始した。1 セッションは、比較音 (2) × 基準音と比較音の呈示順カウンターバランス (2) × 5 試行=20 試行から構成し、図 1(a) に示す 5 種類の基準音すべてで 2 セッション連続で 90% の正答率に達するまで練習を行った。その次に基準音から遠い比較音でも、同様に 5 種類の基準音すべてで 2 セッション連続で 90% の正答率に達するまで練習を続けた。次の段階では、基準音と MFF ratio の差がもっとも小さい比較音 (図 1(a) で、基準音にもっとも近い白抜きの) を対象としたセッションを行った。1 セッションの試行数は最初の練習セッションと同じ 20 試行とし、5 種類の基準音すべてで 2 セッション連続で 80% の正答率を得るまで続けた。最後に、5 種類の基準音とその基準音に付随する 6 種類の比較音すべてをランダムに呈示するセッションで、2 セッション連続で正答率 85% に達すると、練習セッションを終了した。この最後のセッションでは、基準音 (5) × 比較音 (6) × 基準音と比較音の呈示順カウンターバランス (2) で計 60 試行を 1 セッションとした。練習セッションの間、各試行後に正誤のフィードバックが与えられた。練習セッションの完了に要した時間は 1 人あたり約 5 時間であった。

2.4 参加者

8 名の日本語話者 (男性 2 名、女性 6 名、21-24 歳) が実験に参加した。参加者全員、125 Hz から 8000 Hz の範囲のオーディオグラムで健常聴力レベルであることを確認した。実験に先立ち、インフォームドコンセントを実施し同意を得た。実験はヘルシンキ宣言に基づき、和歌山大学の倫理委員会で承認された。

3. 結果と考察

寸法弁別実験の全参加者の結果の平均に対して、ブートストラップ法 [11] により累積ガウス分布を当てはめて得た心理物理曲線を図 1(b) に示す。5 枚のパネルは、図 1(a) の緑の × で示した 5 つの基準音を中心とする MFF-GPR 領域に対応する。各グラフの横軸は比較音の MFF ratio である。グラフ中央の波線の縦線は基準音の MFF ratio を示す。縦軸は基準音の話者と比べて比較音の話者のほうが「小さい」と参加者が判断した割合の百分率である。エラーバーは参加者間の $\pm 1 \times$ 標準偏差を示す。各パネルには 2 本の心理物理曲線をプロットした。赤い曲線は Or-Or、青い曲線は Or-Em の組み合わせの弁別結果に適合した心理物理曲線である。これらの心理物理曲線から、比

較音が「小さい」と回答された割合が 50% から 76% に増加したときの MFF ratio の増分を弁別閾 (JND, %) として算出した。図 1(b) の各パネル内に σ として示している。この JND は二肢強制選択における $d' = 1$ に対応する。

3.1 主観的等価点 (PSE)

高域強調処理によって寸法の知覚が変化するならば、比較音を高域強調処理した Or-Em 条件は Or-Or 条件に比べて心理物理曲線が左右にシフトするはずである。心理物理曲線上で比較音が「小さい」と回答された割合が 50% となる MFF ratio を PSE として条件間で比較した。基準音と比較音のスペクトル傾斜が同じ条件、すなわち Or-Or (図 1(b) の各パネルの赤線) では、PSE は基準音の MFF ratio (縦方向の破線) と心理物理曲線の交差位置に近い。一方で、基準音と比較音のスペクトル傾斜が異なる条件、すなわち、Or-Em (図 1(b) 各パネルの青線) では、Or-Or に比べて、心理物理曲線が MFF ratio の小さい方向にシフトしている。これは、+6 dB/Oct の高域強調処理が施された Emphasized のほうが、Original より小さい話者に聞こえる傾向があることを示している。

MFF-GPR 領域および比較音の高域強調処理の有無の 2 要因が PSE へ影響するかどうか検証するため、参加者ごとに図 1(b) のような心理物理曲線を適合し PSE を算出した。ここで得た PSE について、基準音の MFF ratio を 1 とする比を取り、分析対象とした (表 1)。PSE に対し、MFF-GPR 領域と比較音の高域強調処理の有無を要因とする二要因分散分析を実施した。基準音の MFF-GPR 領域の効果は認められなかったが ($F(4, 70) = 2.13, p = 0.086$)、高域強調処理の有無に有意な効果があった ($F(1, 70) = 46.58, p < 0.0001$)。Or-Or に比べて Or-Em の PSE は平均で 4.9% 小さいほうにシフトした。すなわち、Emphasized の MFF を 4.9% 小さくすると基準となる Original と同じ寸法に知覚されることを意味する。ただし、MFF-GPR 領域と高域強調処理の有無による交互作用がみられた ($F(4, 70) = 3.10, p = 0.021$)。

Tukey-Kramer の HSD による多重比較を行ったところ、Or-Or 条件では、MFF-GPR 領域による PSE の差はなかった。Or-Em 条件では、MFF-GPR 領域の (2) と (5) で Or-Or との有差がみられなかった ($\alpha = 0.05$)。MFF-GPR 領域の (2) と (5) は、通常聞くことが少ない領域であり、弁別判断がより難しい Or-Em 条件において PSE のシフト量に影響した可能性がある。

3.2 弁別閾 (JND)

基準音と比較音のスペクトル傾斜が同じ条件 Or-Or (図 1(b) の各パネルの赤線) では、JND (図 1(b) 中の σ) は 5.2-5.8% の範囲で、平均は 5.5% であった。有声音と無声音を対象とした先行研究 [6] の同じスペクトル傾斜の有声音同士の比較では、JND は 4.2-5.4% の範囲で、平均は 5.0% であった。この先行研究の結果よりわずかに大きいのが、およそ同じ範囲にあると言える。基準音と比較音のスペクトル傾斜が異なる条件 Or-Em (図 1(b) 各パネルの青線、Or-Em) では、JND は 4.5-7.9% の範囲にあり、平均は 5.9% であった。スペクトル傾斜が同じ条件 Or-Or に比

表 1 全参加者の PSE とその標準偏差および平均

組み合わせ	Or-Or	Or-Or	Or-Or	Or-Or	Or-Or	Or-Em	Or-Em	Or-Em	Or-Em	Or-Em
MF-F-GPR 領域	1	2	3	4	5	1	2	3	4	5
ET	0.96	1.00	1.00	0.99	0.99	0.96	0.98	1.00	0.97	0.99
HI	1.07	0.99	1.01	1.04	1.03	0.94	0.98	0.88	0.87	0.99
HT	0.98	1.00	1.00	1.01	1.03	0.90	0.93	0.87	0.94	0.96
MN	1.01	1.00	0.99	0.99	1.02	0.98	1.01	0.96	0.92	0.95
MT	1.00	1.01	0.98	1.00	1.00	0.84	0.98	0.96	0.82	0.94
MY	1.03	0.97	0.98	1.02	1.00	0.90	1.04	0.96	0.91	0.98
SHY	1.01	1.00	0.99	0.99	1.02	0.97	0.98	0.96	0.97	0.97
SM	1.00	0.98	1.00	1.00	0.99	0.97	0.97	1.00	0.97	0.99
標準偏差	0.03	0.01	0.01	0.02	0.02	0.05	0.03	0.05	0.06	0.02
平均	1.01	0.99	1.00	1.01	1.01	0.93	0.98	0.95	0.92	0.97

表 2 全参加者の JND(%) とその標準偏差および平均

組み合わせ	Or-Or	Or-Or	Or-Or	Or-Or	Or-Or	Or-Em	Or-Em	Or-Em	Or-Em	Or-Em
MF-F-GPR 領域	1	2	3	4	5	1	2	3	4	5
ET	3.91	4.82	4.18	5.17	4.99	4.84	5.13	4.39	5.90	5.06
HI	8.09	5.73	6.61	13.00	5.67	6.25	10.26	10.83	11.54	12.96
HT	5.24	5.26	3.98	4.10	5.35	4.86	5.47	5.57	4.43	3.91
MN	5.88	9.00	5.34	6.34	5.41	6.33	11.18	5.87	5.70	4.47
MT	6.51	5.50	14.74	11.59	5.77	8.42	8.64	8.05	18.43	5.34
MY	12.91	4.35	5.19	5.28	6.25	7.43	7.54	5.42	7.99	3.90
SHY	4.18	4.43	4.35	3.89	5.81	7.19	4.54	5.16	5.16	4.92
SM	5.51	4.73	5.03	4.62	4.17	7.65	6.92	7.27	4.92	4.17
標準偏差	2.89	1.51	3.56	3.53	0.63	1.30	2.43	2.08	4.80	3.03
平均	6.53	5.48	6.17	6.75	5.43	6.62	7.46	6.57	8.01	5.59

べて MF-F-GPR 領域間で JND がばらつく傾向がみられた。

JND に対しても、MF-F-GPR 領域の効果と比較音の高域強調処理の有無の効果があるかどうかを確かめるため、参加者ごとの各条件の JND を算出し (表 2) 二要因分散分析を実施した。その結果、基準音の MF-F-GPR 領域の効果も、比較音の高域強調処理の有無の効果も認められなかった (MF-F-GPR 領域, $F(4, 70) = 0.88, p = 0.48$; 高域強調処理の有無, $F(1, 70) = 1.51, p = 0.22$)。MF-F-GPR 領域によっても、音声間のスペクトル傾斜の異同によっても、弁別閾に系統的な違いは現れないようである。

無声音を用いた先行研究 [8] では、スペクトル傾斜の異なる音声間の比較のほうが、JND が大きい傾向が観察されている (例えば、Whispered と Whispered の比較の平均 JND: 5.5%, Whispered と Unvoiced の比較の平均 JND: 9.6%)。それに対して、本研究ではスペクトル傾斜が同じでも異なっても、JND はおおよそ同じ範囲に収まっている。JND が大きいということは、弁別判断の基準が明確でないということであり、課題として難しいということである。スペクトル傾斜の異なる組み合わせの結果からは、有声音による寸法弁別は無声音よりも容易であったということが伺える。さらに MF-F-GPR 領域による差がなかったことから、寸法弁別の判断基準が安定しているとも言える。普段からよく聞くのは有声音であるため、スペクトル傾斜の異なる音声間の寸法弁別という難しい課題でも安定して判断できたことは考えられるが、実際の原因は不明であり今後調査を行う予定である。

3.3 個人内傾向

先行研究 [8] と同様、Or-Or から Or-Em の心理物理曲線のシフト量 (PSE) や、心理物理曲線の傾き (JND) に大きな個

人差がみられた。比較音の高域強調処理の有無が PSE に影響しなかった参加者と、大きく影響した参加者の心理物理曲線の例を図 2 に示す。参加者 ET の結果では Or-Or も Or-Em でも心理物理曲線がほぼ重なっており、シフトがほとんどない (図 2(a))。一方、参加者 HT の結果では Or-Or に比べて Or-Em の心理物理曲線が MF-F の小さいほうにシフトしている (図 2(b))。図 2(a)(b) からわかるように、心理物理曲線のシフト量は MF-F-GPR 領域でおおよそ一定している。すなわち、高域強調処理が PSE に影響する度合いは、個人内でほぼ一定していると言える。Unvoiced と Whispered を使用して同様の実験を行った先行研究 [8] でも、同様の傾向が指摘されている。

JND の個人差については、どの MF-F-GPR 領域でも 5% 前後の JND を示す参加者もいれば、MF-F-GPR 領域によっては 15% を超える JND を示す参加者もいた。全体の分析では、MF-F-GPR 領域と高域強調処理の有無は JND に影響しなかったが、このような大きな個人差が結果に影響したかもしれない。個人差の傾向やその要因についてはさらなる分析が必要である。

4. 結論

音源の寸法の知覚モデルを構築するため、有声音の高域強調が寸法弁別に与える影響を測定する実験を実施し、心理物理曲線をデータに適合した。その結果、+6 dB/Oct の高域強調処理をした有声音は、高域強調処理をしない有声音に比べて、話者の寸法が小さいと知覚されることが明らかになった。ただし、MMF ratio と GPR ratio の組み合わせによっては、高域強調処理の有無による主観的等価点の差がみられない場合があった。弁別閾に関しては、高域強調処理の有無や MMF ratio と GPR ratio の組み合わせによる系統的な影響はみられなかった。個々人の結果からは、弁別対象である音声同士のスペクトル傾斜の

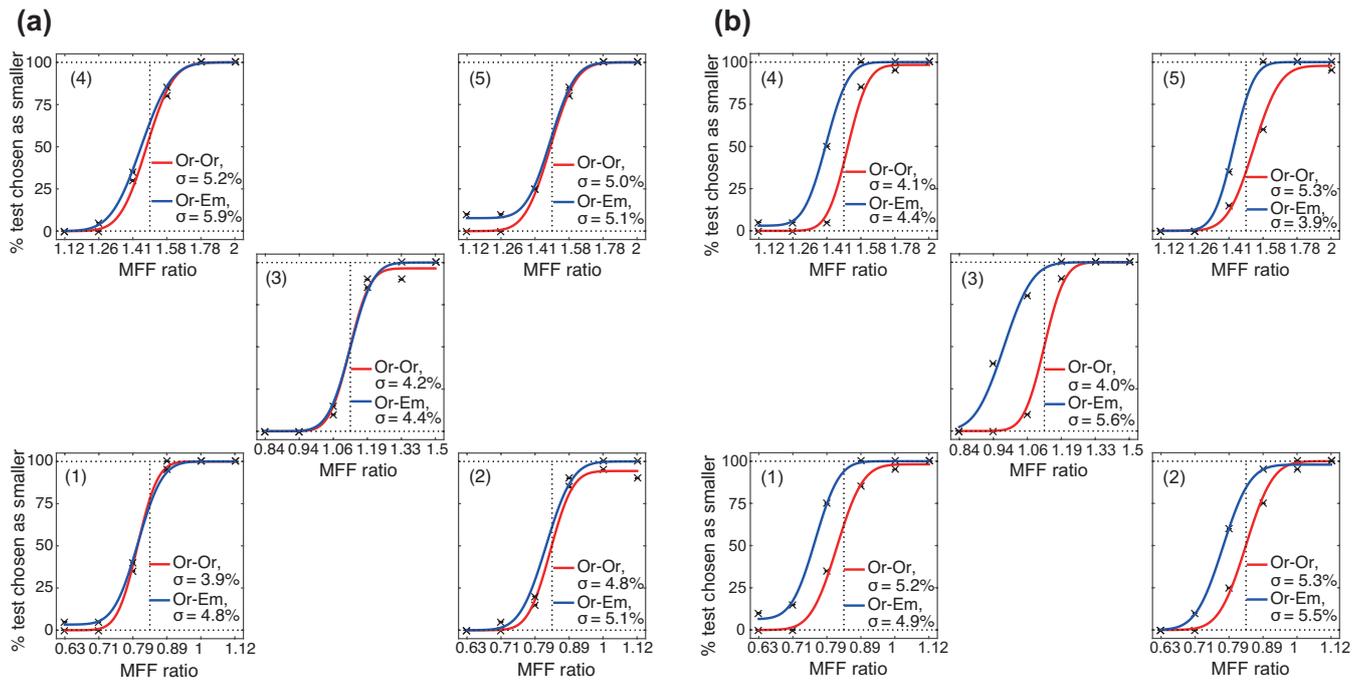


図 2 (a) 参加者 ET の心理物理曲線 . PSE が高域強調処理の有無に影響されない . Or-Or (赤い曲線) と Or-Em (青い曲線) がほぼ重なっている . (b) 参加者 HT の心理物理曲線 . Or-Or (赤い曲線) に比べて , Or-Em (青い曲線) が MFF ratio のより小さいほうにシフトしている . (a)(b) とともに , グラフの読み方は図 1(b) と同じ .

違いが弁別判断に影響しやすい参加者と、影響しにくい参加者がいることがわかった .

謝辞 本研究の実験データの収集をしてくれた尾崎礼氏に感謝する . 本研究は JSPS 科研費 JP25280063 および JP16H01734 の助成を受けた .

参考文献

[1] Irino, T. and Patterson, R. 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. 3-4, 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 f0 extraction: Possible role of a repetitive structure in sounds, *Speech communication*, Vol. 27, No. 3, pp. 187-207 (1999).

[3] 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, F0 and aperiodicity estimation, *Proc. IEEE ICASSP 2008*, pp. 3933-3936 (2008).

[4] Smith, D. R., Patterson, R. D., Turner, R., Kawahara, 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. and Patterson, R. 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] Irino, T., Aoki, Y., Kawahara, H. and Patterson, R. 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).

[7] Fujisaki, H. and Kawashima, T.: The roles of pitch and higher formants in the perception of vowels, *IEEE Transactions, Audio and Electroacoustics*, Vol. 16, No. 1, pp. 73-77 (1968).

[8] Yamamoto, K., Irino, T., Nisimura, R., Kawahara, H. and Patterson, R. D.: How the slope of the speech spectrum affects the perception of speaker size, in *Interspeech 2015 - 16th Annual Conference of the International Speech Communication Association Sept., 6-10, Dresden, Germany, Proceedings*, pp. 1556-1560 (2015).

[9] 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).

[10] Tsujimura, N.: *An Introduction to Japanese Linguistics*, Wiley- Blackwell, Malden, MA (2007).

[11] Wichmann, F. A. and Hill, N. J.: The psychometric function: I. Fitting, sampling and goodness-of-fit, *Perception and Psychophysics*, Vol. 63, No. 8, pp. 1293-1313 (2001).