Vol.2018-MUS-119 No.57
Vol.2018-SLP-122 No.57
2018/6/17

通常発声とささやき声を比較した時の寸法知覚 -どちらが小さい話者に聞こえる？-

上村 怜央[1]　入野 俊夫[1,a)]　Roy D. Patterson[2,b)]

概要：音声から話者のおおよその寸法（声道長）を推定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, 有声音と無声音を使った実験で明らかになった．しかし, 発話様式が異なる有声音とささやき声を比較した場合の寸法知覚は調べられていないため, 本研究ではこの条件での寸法弁別実験を行った．その結果, 発話様式が同じ場合に比べ, ささやき声を基準音とした場合に寸法弁別の精度が低下するのに対し, 有声音を基準音とした場合には従来知られている精度とさほど変わらないことがわかった．これは, ささやき声に対して話者の寸法に対応する “自然な” 声帯振動を想定して寸法弁別を行なっていると仮定すると説明できる可能性がある．

キーワード：寸法知覚, 有声音, ささやき声, 心理物理曲線, 主観的等価点, 弁別閾

Size perception when comparing voiced and whispered speech Does whispering make you smaller?

Ryo Uemura[1]　Toshio Irino[1,a)]　Roy D. Patterson[2,b)]

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 showed that lifting the slope of the speech spectrum by 6 dB/Oct led to a reduction in the perceived size of the speaker. The comparisons were limited to the comparison between voiced and voiced speech or between whispered and whispered speech. This study is performed to know the performance of size perception when comparing voiced and whispered speech directly. Results show that the JND is much greater when using whispered sounds as the reference than when using voiced sound as the reference. This asymmetry could be explain by an assumption that the listeners assumes “natural” glottal pulse rate in whispered sounds when comparing with the voiced sounds.

Keywords: speaker size perception, voiced speech, whispered speech, psychometric function, PSE, JND

1. はじめに

ヒトは音を聞いたとき発音源の寸法と形状を分離して知覚することができる．声だけ聞いて声道長の異なる大人か子供かを推定でき, 同時に声道形状の異なる発音を正確に把握できる．また, ヒトの声のみならず, ヴァイオリンとチェロの音を聞き比べた時, どちらの楽器が大きいかを間違えることはほとんどなく,

[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)] irino@sys.wakayama-u.ac.jp
[b)] rdp1@cam.ac.uk

同時に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, Irino 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計算理論を提案した [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 (CV), 母音-子音 (VC) の音節を用いた寸法弁別 [5], 自然発話された日本語 4 モーラの有声単語および無声単語の寸法弁別 [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 [7] があり, 駆動音源が高周波成分の大きい雑音のためである．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被験者もいれば, 影響しない被験者もいた．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.

さらに, この無声音の実験を受けて, 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 [10], [11]. 原音声のスペクトル傾斜を保持した有声音と, +6 dB/Oct の高域強調処理を付与して合成した有声音の 2 種類の刺激を用いて話者の寸法弁別実験を実施した．その結果, Yamamoto らの実験と同様に個人差はあるが, +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傾向があることが明らかになった．このように先行研究では, 有声音と無声音のそれぞれで高域強調処理の有無によるスペクトル傾斜の違いが寸法知覚に与える影響について調べられた.

以上の実験では, 同じ発話様式 (有声音あるいは無声音) どうしの対比であった. これは, 高域強調処理自体の影響を明確にするためであった. 本研究ではもともとの質問に戻って, 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どの程度適確かを調べるための実験を行った．すなわち, 有声音とささやき声に近い無声音を音声刺激対として寸法弁別実験を行った.

2. 寸法弁別実験

原音声のスペクトル傾斜を保持した有声音と, ささやき声に聞こえる+6 dB/Oct の高域強調処理をした無声音を音声刺激対として寸法弁別実験をおこなった．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, 基準音と比較音の組合わせの効果, 平均フォルマント周波数比 (mean formant frequency ratio, MFF ratio) の効果, これらの要因の作用を検証した.

2.1 刺激音合成

自然に発声された 4 モーラの日本語単語音声が 4 名分収録されている親密度別音声データベース FW03[12] 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, 声帯の振動周波数 (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比例的に伸縮させ, 平均フォルマント周波数 (Mean Formant Frequency, MFF) を変化させた．声道長 (VTL) と MFF はほぼ反比例の関係 (VTL$\propto$ 1/MFF) なので, 等価的に声道長を伸縮したことになる．伸縮させた STRAIGHT スペクトルを, GPR のパルス列で駆動することにより有声音 (Voiced Speech, 以下 Voiced と表記) を得た．また, STRAIGHT スペクトルを, 雑音で駆動することにより無声化音声を作成した．さらに, ささやき声 (Whispered Speech, 以下 Whispered と表記) に似た音声を得るため, この無声音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.

2.2 刺激音の組み合わせと提示

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．まず, 領域が異なる 3 種類の基準音を用意した (緑色 $\times$)．このために, 原音声の MFF ratio を 1 として, MFF ratio が $2^{-3/12}(\simeq 0.84)$, $2^{2/12}(\simeq 1.12)$, $2^{7/12}(\simeq 1.50)$ となるようにした．その上で, 3

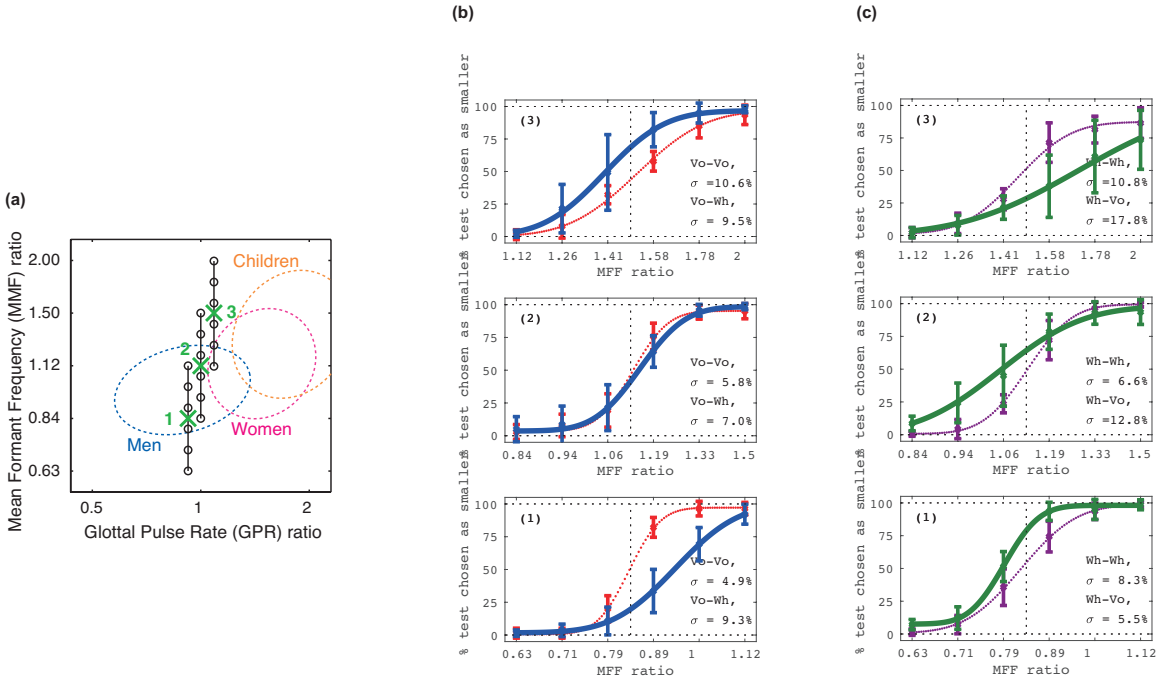

図 1: (a) 刺激音声の MFF ratio と GPR ratio の組み合わせ．MFF ratio の 1 と GPR ratio の 1 は, 再合成前の元の単語の MFF ratio と GPR ratio を示す (話者 mya の MFF は約 1278 Hz, 平均 GPR は約 150 Hz)．番号のついた 3 つの緑の×は基準音を示し, 基準音と縦線でつながれた白抜きの◯はそれぞれの基準音に対する比較音を示す．破線で示した 3 つの楕円は, 成人男性 (Men), 成人女性 (Women), 子供 (Children) の MFF と GPR の平均的分布を示す．(1)(2)(3) 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 はすべて 1.0 である．(b)(c) 3 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．7 名の結果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」だと判断された割合 (%)．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．赤色の曲線: 基準音と比較音の両方が Voiced 刺激である場合 (Vo-Vo) の弁別結果による心理物理曲線．青色の曲線: 基準音が Voiced, 比較音が Whispered である場合 (Vo-Wh) の弁別結果による心理物理曲線．紫色の曲線: 基準音と比較音の両方が Whispered 刺激である場合 (Wh-Wh) の弁別結果による心理物理曲線．緑色の曲線: 基準音が Whispered, 比較音が Voiced である場合 (Wh-Vo) 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 7 名の被験者の平均．エラーバー: 被験者間の標準偏差 ±1．σ: MFF-GPR 各領域における MFF ratio の JND(%)．

種類のそれぞれの基準音を中心として, 相対的に MFF ratio を $2^{-5/12}, 2^{-3/12}, 2^{-1/12}, 2^{1/12}, 2^{3/12}, 2^{5/12}$ だけ変化させた 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, 基準音と同一領域内で 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, Voiced および Whispered の 2 条件とした．比較音には, 基準音と同じ GPR とスペクトル傾斜をもつ Voiced と, 無声化処理を行い, +6 dB/Oct の高域強調処理を施した Whispered がランダムに割り当てられた．基準音, 比較音ともに Voiced の組み合わせを Vo-Vo, 基準音が Voiced, 比較音が Whispered の組み合わせを Vo-Wh, 基準音, 比較音ともに Whispered の組み合わせを Wh-Wh, 基準音が Whispered, 比較音が Voiced の組み合わせを Wh-Vo と表記する．被験者は, 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．本実験では, 有声音とささやき声の対比で正誤を刺激音から判定できないため, フィードバックは与えられなかった．

試行数は被験者 1 人あたり 1440 試行であった (基準音 (3) × 比較音 (6) ×基準音と比較音の組み合わせ (4) × 1 つの組み合わせに対する測定回数 (基準音と比較音の呈示順のカウンターバランス (2) × 10 = 20)．1 セッション 72 試行として 20 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 5760 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ることがないよう, 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 (OPPO HA-1) を通して, ヘッドフォンスピーカ (OPPO PM-1) で呈示された．呈示音圧レベルは A 特性で平均 65 dB とした．ヘッドフォンスピーカはサウンドレベルメータ (Bruël & Kjær, Type 2250-L) と人工耳 (Bruël & Kjær, Type 4153) でキャリブレーションを行った．また, 音圧レベルによって判断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ビングを行った．

2.3 被験者

7 名の日本語話者 (男性 3 名, 女性 4 名. 21-23 歳.) が実験に参加した. 被験者は全員, 125Hz から 8000 Hz の範囲で聴力検査により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. 実験は和歌山大学の倫理委員会で承認されている.

2.4 練習セッション

被験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 Vo-Vo, Wh-Wh 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 3) の単語リストとし, 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 18 試行から構成し, 図 1(a) に示す 3 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 3 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 18 試行とし, 3 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.

さらに次の段階では, 3 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を行った. 36 試行を 1 セッションとし, 2 セッション連続で正答率 85%の正答率を得るまで続けた. 最後に, 本実験と同様の組み合わせ (Vo-Vo, Vo-Wh, Wh-Wh, Wh-Vo) を含んだセッションを行い, 練習セッションを終了した. この最後のセッションでは 36 試行を 1 セッションとし, 正答率ノルマは設けなかった. 最後のセッションを除き, 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.

3. 実験結果と考察

寸法弁別実験の全被験者の結果に対して, ブートストラップ法 [13] 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) および図 1(c) に示す. 6 枚のパネルのうち, 図 1(b) は基準音が Voiced, 図 1(c) は基準音が Whispered であり, 下から順に図 1(a) の緑の×で示した 1〜3 の基準音の領域にそれぞれ対応する. 各パネルの横軸は比較音の MFF ratio である. グラフ中央の縦点線は基準音の 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」と被験者が判断した割合の百分率である. エラーバーは被験者間の ±1× 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. 図 1(b) の赤曲線は Vo-Vo, 青曲線は Vo-Wh, 図 1(c) の紫曲線は Wh-Wh, 緑曲線は Wh-Vo 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. 各被験者ごとの結果を表 1 に示す. また, 全被験者のデータから求めたものは, 図 1(b) および図 1(c) 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.

さらに, 心理物理曲線が判断率 50%と交差する MFF ratio の値を主観的等価点（PSE）として計算した. 各領域間で比較できるように, PSE の値を基準音からのずれの相対的な割合 (%) を算出した結果を表 2 に示す. この値が負であれば左，正ならば右に，図 1(b) および図 1(c) で示した心理物理曲線がシフトしていることを示す.

3.1 心理物理曲線・弁別閾（JND）・主観的等価点 (PSE)

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, 有声音を用いた先行研究 [10], [11] では, スペクトル傾斜が同じでも異なっていても, JND はおおよそ同じ範囲に収まっている (例えば, VoicedVoiced の比較の平均 JND: 5.5%, Original と Emphasized の比較の平均 JND: 5.9%).

今回の実験によって得られた心理物理曲線を図 1(b) および図 1(c) に示す. また, JND と PSE に関して, 被験者・領域・刺激対ごとにまとめたものを表 1 と表 2 にそれぞれ示す.

基準音 Voiced, 比較音 Voiced の組み合わせ (Vo-Vo)(図 1(b) 各パネルの赤線) では, JND(図 1(b) 各パネル中の σ) は 4.9-10.6%の範囲で, 平均は 7.1%であった. また, 心理物理曲線の中心はほぼ基準音のところにあり, 判断にバイアスが無いことがわかる. これらの結果は, 発話様式の一致する従来研究の実験結果とほぼ同じで, この実験で十分精度良 く再現できていることがわかる.

基準音 Voiced, 比較音 Whispered の組み合わせ (Vo-Wh)(図 1(b) 各パネルの青線) では, JND は 7.0-9.5%の範囲にあり, 平均は 8.6%となった. JND に関してはばらつきも考慮すると Vo-Vo 条件とほぼ同じ程度である. 心理物理曲線の中心が領域ごとに異なり, 領域 1(下段) では右シフト, 領域 2(中段) ではシフト無し, 領域 3(上段) では左シフトしている. このことは領域 1 では Vo の方が小さい話者, 領域 3 では Wh の方が小さい話者と判断される傾向があることを示す. このように領域によって異なることは, MFF の違いだけでは説明できない要因があることを示唆している. これに関しては,3.2 で検討する.

さらに, 基準音 Whispered, 比較音 Whispered の組み合わせ (Wh-Wh)(図 1(c) 各パネルの紫線) では, JND(図 1(c) 各パネル中の σ) は 6.6-10.8%の範囲で, 平均は 8.5%であった. これは, Vo-Vo 条件や Vo-Wh 条件とほぼ同程度で, 心理物理曲線にバイアスが無い点も含めると従来研究と一致する結果である.

これに対して, 基準音 Whispered, 比較音 Voiced の組み合わ

表 1: 全被験者の弁別閾 (JND %) とその平均および標準偏差

組み合わせ	Vo-Vo	Vo-Vo	Vo-Vo	Vo-Wh	Vo-Wh	Vo-Wh	Wh-Wh	Wh-Wh	Wh-Wh	Wh-Vo	Wh-Vo	Wh-Vo
MFF-GPR 領域	1	2	3	1	2	3	1	2	3	1	2	3
NO	4.37	6.72	8.24	9.38	3.67	6.08	6.97	4.86	12.81	3.99	12.84	13.68
MY	4.51	5.19	7.74	9.28	7.57	3.94	8.12	4.52	6.49	5.76	5.52	8.87
SH	3.92	4.02	13.43	6.79	6.24	9.13	8.69	7.36	-	7.13	8.38	13.66
YT	6.31	10.10	10.20	11.08	15.45	13.16	8.74	5.80	11.23	8.50	14.29	17.86
KY	4.48	3.76	7.57	4.74	4.91	5.65	7.68	5.21	5.85	4.37	8.52	-
NY	4.20	5.26	9.73	8.27	5.45	5.45	3.58	3.97	7.54	4.98	9.03	13.14
RS	7.78	7.61	18.80	15.51	6.05	11.89	13.76	8.76	18.26	6.99	21.38	19.61
平均	5.08	6.10	10.82	9.29	7.05	7.90	8.22	5.78	10.36	5.96	11.42	14.47
標準偏差	1.42	2.23	4.05	3.41	3.89	3.53	3.01	1.70	4.74	1.64	5.28	3.80

表 2: 全被験者の主観的等価点 (PSE) の基準音からのずれの割合 (%) とその平均および標準偏差

組み合わせ	Vo-Vo	Vo-Vo	Vo-Vo	Vo-Wh	Vo-Wh	Vo-Wh	Wh-Wh	Wh-Wh	Wh-Wh	Wh-Vo	Wh-Vo	Wh-Vo
MFF-GPR 領域	1	2	3	1	2	3	1	2	3	1	2	3
NO	-1.76	-3.12	1.35	19.74	5.92	-8.98	-1.48	-1.64	1.28	-8.12	-14.02	16.51
MY	3.06	2.84	0.56	14.48	2.66	2.35	0.53	2.29	3.47	-6.70	-5.97	0.22
SH	0.59	1.85	4.35	9.99	5.41	-8.51	-2.43	3.04	-0.53	-8.24	-6.07	15.38
YT	-0.44	-1.80	4.81	3.67	-11.11	1.16	-1.96	0.76	2.18	-2.63	0.00	4.00
KY	-1.95	-0.04	2.66	12.24	0.30	-14.94	-1.21	0.07	-5.56	-6.72	-1.93	-
NY	-0.02	2.55	-1.29	11.74	2.81	-1.21	-1.23	-1.91	-2.20	-3.60	-6.08	2.72
RS	-0.76	1.01	2.07	7.84	-1.24	-7.63	-0.54	2.78	2.02	-7.22	3.78	12.67
平均	-0.18	0.47	2.07	11.39	0.67	-5.39	-1.19	0.77	0.09	-6.18	-4.32	8.58
標準偏差	1.69	2.25	2.13	5.07	5.79	6.31	0.96	2.04	3.12	2.19	5.67	7.08

せ (Wh-Vo) (図 1(c) 各パネルの青線) では, JND は 5.5-17.8%の範囲で, 平均は 12.0%となっており, Wh-Wh 条件よりも JND もばらつきも平均値も大きくなっている. このことから, ささやき声 (Wh) を基準音として有声音 (Vo) を比較する場合は, 他の条件よりも弁別が難しくなる傾向があることがわかる. また, 心理物理曲線のシフトに関して, 領域 1, 3 では Vo 基準音 (図 1(b)) の場合と逆転しているのは整合性のある結果である. しかし領域 2 に関しても左シフトしており, 単純には解釈できないこともわかる.

実験の当初, 発話様式が異なる有声音とささやき声での比較はかなり難しくなると予測していた. また, 有声音とささやき声のどちらを基準音としても, 心理物理曲線のシフトが反転するだけでほぼ対称の結果が得られると考えていた. 今回の結果から, Vo を基準音とした Vo-Wh では JND もそれほど変わらずに判断できた. これに対して, Wh を基準音とした Wh-Vo ではこの状況と異なり, 難しさが増した. さらに Vo-Wh と Wh-Vo の心理物理曲線は対称的になっていない. これは興味深い結果で, なぜそのようになるか考察する価値がある.

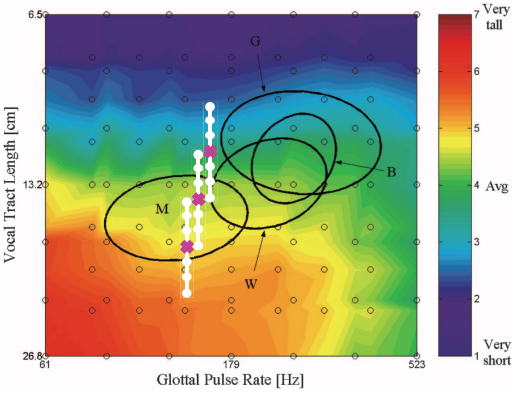

図 2: 話者の身長判断への声道長と声帯振動周波数の影響. 縦軸: Vocal Tract Length(VTL∝ 1/MFF). 横軸:Glottal Pulse Rate(GPR). 濃い赤色に近いほど寸法が大きく, 深青色に近づくほど寸法が小さい話者の音声として知覚している。各円は, 男性 (M), 女性 (W), 男の子 (B), 女の子 (G) の通常の音声の範囲を表す. 実験で使用した音声のおおよその位置 (×が基準音, その上下の○が比較音). (文献 [9] を参考に作成)

3.2 考察

まず, 図 1(b) で領域 1 と 3 で心理物理曲線のシフトの方向が反対であることに関して考える. 上記のとおり領域 1 では Vo の方が小さい話者, 領域 3 では Wh の方が小さい話者と判断される傾向がある. もし, 寸法知覚が VTL だけに頼って行われていれば, このような心理物理曲線のシフトは起こらないはずである. そこで音声からのピッチ感が判断に影響するのではないかと考えた. Smith と Patterson[9] は, VTL と GPR が話者の身長推定に与える影響を調べている. そこでは寸法弁別実験ではなく, 寸法判断のスケーリング実験を行なった. その結果を図 2 に示す. 寸法推定には VTL の影響（縦方向の変化) が GPR の影響 (横方向の変化) に比べて大きいが, GPR の影響もある程度あることが示されている.

ささやき声には音からピッチ情報は与えられないが, 聞いた時に自然なイントネーションの印象を持つことは経験することである. また, 音声合成で間違ったイントネーションを与えるよりも, 雑音駆動にした方が自然な音声になることも知られている. そこでここでの仮説として, ささやき声から得られる音声の自然なピッチ感が脳内に生成され, この情報も必要に応じ寸法判断で用いられると仮定する.

図 1(b) のどの領域でも GPR の値は原音声と同じで固定されている. したがって領域 1 の基準音の MFF を小さくした (VTL

を大きくした) 場合は, 自然な VTL-GPR の組み合わせの音声に比べ, VTL の大きさの割にピッチが高く聞こえる少し不自然な音声になっている. もしささやき声が自然なピッチ感を持っていたとすると, 有声音の方が甲高い声に聞こえるはずである. すなわち小さい話者と判断しやすいこととなる. これが領域 3 では逆で, 有声音は VTL の小ささの割にピッチが低く, 大きい話者と判断しやすくなる. VTL-GPR の組み合わせが原音声とほぼ同じで自然な領域 2 では, 両者のピッチ感の違いはなくちょうどバランスが取れていることになる. 以上で図 1(b) の 3 領域は矛盾なく説明できる.

一方, 図 1(c) の, ささやき声 (Wh) を基準音とした場合の領域 1, 3 における心理物理曲線のシフトは上記とほぼ同様に説明できる. しかし, 領域 2 に関してもシフトがあり, JND も大きい. また, 領域 3 においては被験者間のばらつきが非常に大きく, JND も大きい. この原因は, タスクの難しさが増加したためとして考えられる. この場合, 基準音がささやき声 (Wh) で常にピッチ感が一定である. しかし比較する有声音 (Vo) の MFF ratio によって, VTL に対して不自然なピッチが毎回変わることになる. 寸法知覚にピッチの影響があると考えると, 判断に迷いが生じやすい状況になっていることがわかる. これは 図 1(b) の, 比較音がささやき声 (Wh) の場合は GPR 情報がないため, 比較音の MFF ratio によって自然性はまったく変わらない. これにより, 図 1(c) の領域 2, 3 における被験者のばらつきと JND の大きさを矛盾なく説明できる. ただし, 逆に領域 1 において, シフトがあるものの小さい JND がなぜ出たのかの説明とはならない. これに関してはまだ良くわからず今後の課題である.

最後に図 2 の結果 [9] と今回の結果を比べる. 図 2 におおまかに今回の実験の 3 領域の刺激音の配置を示した. VTL 方向 (縦方向) の変化は明確である. これに対して GPR の変化 (横方向) に対する判断の変化はあまり大きくない. これは, 弁別実験ではなく有声音だけを使って寸法をスケーリングする実験のため, VTL の影響だけが強く出たと考えられる. これに対し今回の有声音とささやき声の実験では, VTL だけではなくピッチ感 (有声音の場合は GPR から. ささやき声の場合は自然な想定から) の影響も大きく出ているようである. この寸法弁別のパラダイムは脳内のピッチ感を調べるためにも有効かもしれない.

4. まとめ

有声音とささやき声を音声刺激対として寸法弁別実験を行った. その結果, 基準音をささやき声とした場合, 基準音を有声音とした場合や従来研究での結果に比べ, 寸法弁別の精度が低下することが分かった. また, 心理物理曲線のシフト方向から, VTL が大きい領域 1 では Whispered よりも Voiced の方を小さい寸法の話者だと判断する傾向があり, VTL が小さい領域 3 では Voiced よりも Whispered の方を小さい寸法の話者だと判断する傾向があることがわかった. 実験では声帯振動の影響を避けるために各領域で一定の GPR としたが, 領域 1 と領域 3 で使用した音声が, 寸法とピッチに矛盾があるやや “不自然な” 音声となってしまい, 影響が出た可能性がある. この結果は, 本来声帯振動が無いささやき声に “自然な”GPR を被験者が想定し, 有声音の GPR と比較して寸法判断をおこなったと仮定すると説明できる可能性がある. この仮説が正しいか検証するため, 現状の領域 1 と領域 3 を変更して, VTL と GPR の組み合わせが “自然な” 音声を使用した寸法弁別実験を計画中である.

謝辞 本研究は，科研費基盤 A 16H01734 の支援を一部受けた.

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, 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, J. Acoust. Soc. Am., 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, J. Acoust. Soc. Am., 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 Trans., Audio 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, Interspeech 2015, pp. 1556–1560, Dresden, Germany (2015).

[9] Smith, D. R. and Patterson, R. D.: The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age, J. Acoust. Soc. Am., Vol. 118, No. 5 (2005).

[10] 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, 情報処理学会, 音学シンポジウム 2017, 情報処理学会研究報告,Vol.2017-MUS-115, No.44,pp.1-6 (2017).

[11] Matsui, T., Irino, T., Yamamoto, K., Kawahara,H. and Patterson, R.D. : The effect of spectral tilt on size discrimination of voiced speech sounds, Interspeech 2017, pp.2949-2953, Stockholm, Sweden (2017).

[12] 坂本修一, 鈴木陽一, 天野成昭, 小澤賢司, 近藤公久, 曽根敏夫: 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] 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).