

# 日本語のささやき発話における子音有声性の判別

平岡 睦津乃

zero.mh.snow97@gmail.com

キーワード：ささやき ささやき発話 有声性 知覚 実験音声学 日本語

## 要旨

本研究の目的は、日本語のささやき発話における音韻的な有声子音音素と無声子音音素(以下、「有声音素」「無声音素」とする)の聞き分けに関して、次の点を論じていることである。ささやき発話の知覚に関する先行研究では、単一の調音方法のみを扱っているものが多く、複数の調音方法を扱っていても、それらの間で有声・無声音素の聞き分け精度に差があること、およびその要因に言及したものは少ない。そこで、(調音点は同じであるが)調音方法が異なる子音を比較することで、調音方法の違いによって有声・無声音素の聞き分けのしやすさに違いがあるかどうかを検証する。

通常の発話とささやき発話の音声を用いた聴取実験の結果、通常の発話よりは低い程度ではあるものの、ささやき発話においても有声音素と無声音素はある程度の水準で正しく聞き分けられることがわかった。調音方法どうしでの比較では、摩擦音のほうが閉鎖音よりも有声・無声音素の聞き分けの成績が低い傾向にあることが示された。その要因として「有声・無声音素間の開放強度の差異という聞き分けの手がかりが、閉鎖音でのみ利用可能である」とする Mills (2009) の議論を、より精緻化することを試みた。聴取実験の結果はどれも発話者および聴取者による差がかなり大きく、ささやき発話における有声音素と無声音素の出し分け方および聞き分け方は個人差が大きいことも示唆された。この個人差についても、閉鎖音よりも摩擦音のほうが大きいことが観察された。

## 1. 導入

本論では、日本語の語中(母音間)位置に現れる、音韻的な有声・無声の対立を持つ子音音素に関して、ささやき発話においてそれらを正しく聞き分けられる精度について検証する。まず 1.1 節で、本論文の前提知識について示す。1.2 節では具体的な問題提起と先行研究、および本研究の仮説を提示する。2 節で実験の手法と結果を示し、3 節で総合的な考察を展開する。

## 1.1. 予備知識

「ささやき」に関する先行研究においては、様々な用語が使われている。不要な混乱を避けるために、それらの用語を整理した上で、本論での用語の使い方をここで提示しておく。

### 1.1.1. 音韻的有声性と音声的有声性

「有声」「無声」および「有声性」という語は、音韻的な意味と音声的な意味の両方で用いられる。これらの用語遣いについて、初めに説明しておく。

たとえば日本語のタ行の子音音素 /t/ とダ行の子音音素 /d/ は、音韻的な「有声性」という弁別素性によって対立する「無声音素」と「有声音素」である。

通常の発話において「タ」・「ダ」と発話したとき、子音音素はそれぞれ [t]・[d] というように、声帯振動を伴わない閉鎖音とそれを伴う閉鎖音、すなわち音声的な「無声音」と「有声音」として実現する。一方、1.1.2 節で述べるように発話のあいだ一貫して声帯振動を伴わない「ささやき発話」のなかで「ダ」と発する際には、子音音素 /d/ の部分は音声的な「有声音」とはなっていない（「ささやき発話」という用語については直後の 1.1.2 節で解説する）。

さて、本論で「有声音素」および「無声音素」というときは、必ず音韻的な意味であり、実際に声帯が振動している（「声」が出ている）か否かを指すものではない。すなわち「通常の発話中の /z/ 」と「ささやき発話中の /z/ 」はどちらも「有声音素」と呼ぶことになる。

反対に、実際の声帯振動を問題にすると、すなわち音声的な意味を示す場合には、必ず「音声的な有声（無声）音」というように表記する。

### 1.1.2. 「ささやき」とは何か

声門は、声帯間の部分（声帯声門）と、披裂軟骨間の部分（軟骨声門）とに分けられ、これらは独立に開閉する。私たち人間は、声門の状態を変化させ、組み合わせることで様々な発話している。ここでは声門の状態のうち3種類を取り上げ、服部（1984: 19-29）を参考にして説明する。

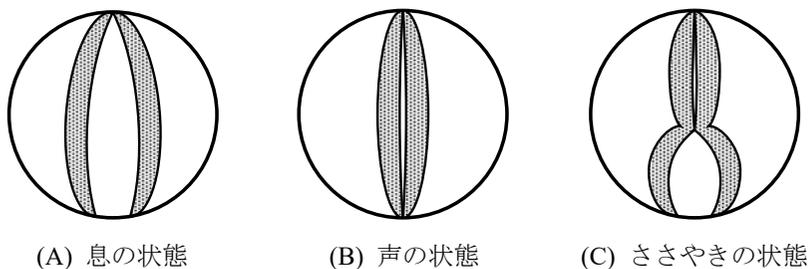


図 1 声門の状態（模式図）

図1の(A)は、喉頭の筋肉が弛緩し、声門が完全に開放された状態であり、日常の呼吸時や、通常の発話で無声子音音素を調音するときの状態である。

(B)は反対に、声門がほぼ完全に閉じた声門閉鎖の状態である。声帯間のわずかな隙間から呼吸が通り抜ける際に声帯が開閉を繰り返し周期的に振動することで、いわゆる「声」(buzzとも呼ばれる)が出る。通常の発話で有声子音音素や母音を調音する際にこの状態となっている。

(C)では、声帯声門は閉じ、軟骨声門のみが開いている。これは主に、内緒話のように声を出さずにヒソヒソと話すなかで見られる状態であるが、ヒソヒソ話<sup>1</sup>のあいだじゅうずっとこの状態というわけではない。たとえば「電気」/deNki/をヒソヒソと話す場合には、声門が(C)の状態であるのは /d/, /e/, /N/, /i/ の部分(音韻的に有声音素である部分)だけである。/k/の部分ではヒソヒソ話のなかでも声門は(C)の状態ではなく、通常の発話における/k/のときと同じく(A)の状態になっている。つまりヒソヒソ話では、通常の発話では「(B)で調音された音」が現れる位置に、代わりに「(C)で調音された音」が現れ、もともと「(A)で調音された音」であるものは、そのまま(A)で調音されるといわれている。

なお、「ささやき」に関する用語にはいくつかの使われ方がある。

服部(1984: 19-29)は、ヒソヒソ話のような話し方のことを「ささやき声」、(C)のような声門の状態のことを「ささやき」、(C)の状態で調音された音声のことを「ささやき音」と呼んでいる。城生(2008: 21-23)も同様、(C)の状態で調音された音声を「ささやき音」と呼んでいる。

一方で、加藤・安藤(2016: 22)は、(C)の状態で調音された音声を「ささやき声」と呼んでいる。

キャットフォード(著)・竹林ほか(訳)(2006: 64-76)は「ささやき」を、声門の状態と調音された音声の両方の意味で用いている(調音された音声については「ささやきの音」とも表現している)。また、ヒソヒソ話のような話し方のことを「ささやきの発話」と呼んでいる。さらに、声帯の振動する音と、軟骨声門の間を呼吸が通るときに発生する乱気流のノイズが同時に出ているような音声(広義の *breathy voice* の一種)を「ささやき声」と呼んでいる。

このように種々の用語遣いが見られるが、混乱を避けるため、本論では次のように使い分けることにする。

<sup>1</sup> ここでいう「ヒソヒソ話」は、声帯振動を伴わないタイプの内緒話の形態を指す。城生(2008: 21-23)は「ひそひそ声(*whisper voice*)」という語を「声門の狭めが(A)より緩いが、声帯の周期的な振動は伴うような音声」を指して用いているようだが、本論文でいう「ヒソヒソ話」とは別物である(ただし城生(2008)の記述は難解で、その意図を正確に理解できていないとは限らない)。さらに、城生(2011: 74-75)では「ささやき声の音量が極めて低いもの」を指して「ひそひそ声」と記述しており、城生(2008)の「ひそひそ声」と城生(2011)の「ひそひそ声」が同じものを指しているかどうか、筆者は判断できなかった。

- ・ 「ささやき発話 (Whispered speech)」: 内緒話のような場面で用いられる (一貫して声帯の周期的な振動を伴わない) ヒソヒソという「話し方の形態」。
- ・ 「ささやき」: (C) のように声帯声門が閉じ軟骨声門が開いているような「声門の状態」、いわゆる「発声のタイプ」(Phonation) のひとつ。
- ・ 「ささやき音」: 声門が (C) の状態で調音された音声。

ここまでを一旦まとめると表 1 のようになる。

表 1 「ささやき」に関わる用語遣い

	話し方の形態	声門の状態	調音された音	その他の意味
服部 (1984)	ささやき声	ささやき	ささやき音	—
城生 (2008)	—	—	ささやき音	ひそひそ声
加藤・安藤 (2016)	—	—	ささやき声	—
キャットフォード (竹林ほか訳) (2006)	ささやきの 発話	ささやき	ささやき (ささやきの音)	ささやき声
<b>本論文</b>	<b>ささやき発話</b>	<b>ささやき</b>	<b>ささやき音</b>	—

たとえば /aza/ と /asa/ に関して、それらが「通常の発話」と「ささやき発話」それぞれにおいて音声的にどのように実現されるかを、本論文の用語遣いに従って説明すると次のようになる。

「通常の発話」中の /aza/ では、/a/, /z/, /a/ はすべて声門が「声」の状態<sup>1</sup>で調音された「有声音」である。それに対し、「通常の発話」中の /asa/ のうち、2つの /a/ は「有声音」であるが、/s/ は声門が「息」の状態<sup>2</sup>で調音された「無声音」である (ここでの「有声音」「無声音」は音声的な意味)。

これを「ささやき発話」に対応させて書くと次のようになる。「ささやき発話」中の /aza/ では、/a/, /z/, /a/ はすべて声門が「ささやき」の状態<sup>3</sup>で調音された「ささやき音」である。それに対し、「ささやき発話」中の /asa/ のうち、2つの /a/ は「ささやき音」であるが、/s/ は声門が「息」の状態<sup>4</sup>で調音された「無声音」である (ここでは音声的な意味)。

ここまでは典型的な「ささやき」の定義を示した。しかし現実的なことを言えば、今回の実験において「この単語を“ささやいて”ください」と指示された発話者が、実際にどのような声門の状態<sup>5</sup>で調音したかを目視で確認することは不可能である。よって、手続き上、筆者自身が聴覚印象から判断して、一貫して声帯振動を伴っていない (いわゆる「声」が出ていない) と思われるものを「ささやき発話」と認定し、そのなかに現れるものを「ささやき音」として取り扱うことになる。

## 1.2. 問題提起・先行研究

通常の発話において、有声音素と無声音素とを聞き分ける際の最も基本的な手がかりは、声帯振動の有無である。一方、通常の発話と異なり声帯の周期的な振動を伴わない特徴を持つささやき発話においても、これらを聞き分けることは可能である。

ささやき発話中の有声音素と無声音素はそれぞれ「ささやき音」と「音声的な無声音」として実現され、両者では声門の状態が異なることを 1.1.2 節で示した。

通常の発話中の有声音素と無声音素（音声的な有声音・無声音）においても声門の状態が違ってくるのだが、単に声門の状態だけが違って、声道におけるその他の調音的特徴は全く同じというわけではない。Kochetov (2014) [日本語：母音間位置]<sup>2</sup>は、通常の発話中の子音音素 /s/, /z/, /t/, /d/ について、舌と口蓋の接触面積が /z/ > /s/ および /t/ > /d/ となっていることを示した。ささやき発話においても、このような舌・口蓋接触面積の差が有意に見られることを Yoshioka (2008) [日本語：母音間位置] が示している<sup>3</sup>（こちらは摩擦音についてのみ）。

このような調音的差異がいくつも組み合わせたり、ささやき発話中の有声音素と無声音素の音響的な差異を生み出していると考えられる。

さて、これら諸々の空気力学的条件は調音方法によって異なるはずである。特にささやき発話においては、有声・無声音素の区別にあたって声帯振動の有無という手がかりを使えないために、こうした調音的差異が有声・無声音素の聞き分けに寄与する度合いがより大きいと考えられる。それに伴い、聞き分けの精度に差が生じることが考えられる。日本語のささやき発話における有声・無声音素の聞き分けに関して、調音方法の違いによる聞き分けの精度に直接的あるいは間接的に言及している先行研究には、吉岡・村瀬 (1992) [母音間位置] と東川 (1994) [語頭位置] がある。

まず吉岡・村瀬 (1992) は、閉鎖音では 100%、摩擦音では 70~80%の精度で有声・無声音素の聞き分けができたとのデータを示している。ただし、この聴取実験は調音的側面の研究の一環としておこなわれており、EPG<sup>4</sup>の器具を付けた状態で発話した音声を刺激として用いている。器具の装着に習熟した発話者であることが望ましいと同論文内でも言及されている通り、実際の発話者は普段通りの発話ができなかった可能性が否定できない。より自然に近い環境での追実験が必要と考える。

<sup>2</sup> 本稿では見やすさのため、それぞれの先行研究がターゲットとした言語と子音音素の出現位置（および子音音素の種類）を、直後に [ ] を付して表した。たとえば「Kochetov (2014) [日本語：母音間位置]」は、日本語の母音間位置における子音音素についての研究であることを表す。

<sup>3</sup> ただし Skarnitzl et al. (2013) [チェコ語：母音間位置の摩擦音] は、舌・口蓋接触面積に関する有声・無声音素間の差は、ささやき発話ではほとんど見られなくなると指摘している。

<sup>4</sup> エレクトロパラストグラフィ (Electropalatography) のこと。複数の電極が配置された人工口蓋床を上顎口腔内に装着して、舌と口蓋の接触パターンなどを観察するものである。

次に東川 (1994) は、日本語の単音節すべてを扱い、ささやき発話中の /s/ と /z/ の聞き分けの精度は極めて良好であったと結論付けている。すなわち、有声・無声音素の聞き分け精度に関して、調音方法の違いによる差がなかったということになるが、次の点に注意が必要である。この東川 (1994) の論考では語頭位置に現れる子音を扱っているが、一般に、日本語の語頭位置に現れる /z/ は破擦音 [dʒ] として実現されることが知られている (詳しくは 2.1.2 節で触れる)。よって語頭位置の /s/ と /z/ の聞き分けは、摩擦音における有声・無声音素の聞き分けとはなっておらず (実際には摩擦音と破擦音の聞き分けである)、閉鎖音とは対比できない。

このように、ささやき発話における調音方法の違いにより聞き分け精度の違いが見られるとする先行研究 (吉岡・村瀬 1992) と、聞き分け精度の違いが見られないとする先行研究 (東川 1994) の両方が存在し、かつその両方とも実験方法に改良の余地がある。そのため、日本語のささやき発話において調音方法の違いによる聞き分け精度の差があるかどうかを知るためには、先行研究とは異なる方法で実験をし直す必要がある。

なお、日本語以外のささやき発話に関する先行研究には次のものがある。

Dannenbring (1980) [英語：語頭位置]、Mills (2009) [英語：語頭位置] などで提示されたデータからは、ささやき発話において同じ調音点の閉鎖音と摩擦音では、摩擦音の方が聞き分けの成績が低いことが読み取れる。Meenakshi & Ghosh (2015) [インド系：母音間位置] も同様、(子音体系は異なるものの) 閉鎖音と摩擦音では聞き分け成績に差があることが読み取れる。このように、日本語以外のささやき発話を扱った先行研究はいずれも、調音方法の違い (閉鎖音か摩擦音か) によって、聞き分け精度に差があることを示唆している。(ただし注意が必要なのは、これら先行研究の多くは「閉鎖音よりも摩擦音のほうが聞き分けの精度が低い」ことを、著者自身がその対比において明示的に考察しているわけではなく、論文内で示されているデータからそうした傾向を (本稿の筆者が) 読み取ることができるに過ぎない点である。)

本論では、上述の有声・無声音素の調音的差異の存在や、日本語以外のささやき発話に関する先行研究での指摘を踏まえて、日本語の母音間位置における子音音素についても、ささやき発話における有声・無声音素の聞き分けの精度は閉鎖音よりも摩擦音のほうが低いと仮説を立て、聴取実験によって検証する。

## 2. 実験

### 2.1. 手法

#### 2.1.1. 発話者

発話者（聴取実験で用いるための録音音声を提供してくれる人）は、東京近辺出身<sup>5</sup>の20代女性3名である（筆者を含む）。筆者以外の2名は、特に音声学の訓練などは受けていない者である。

#### 2.1.2. データ

今回の実験では、「あた」「あだ」「あさ」「あざ」という4つの無意味語を使用した。

日本語の子音音素の実現形で、通常の発話において音声的な有声・無声の対立があるもののうち「同一調音点で調音方法が異なる」ペアには、歯茎閉鎖音 [t], [d] と歯茎摩擦音 [s], [z] の組み合わせがある。これらの子音はタ・ダ行とサ・ザ行に現れるが、子音音素 /t/, /d/ および /s/, /z/ は、表2のように実現されることが知られている（「-」は該当する組み合わせの不在を表す）。

表 2 日本語のタ・ダ行とサ・ザ行の子音音素の実現形

子音の位置	子音音素	後続母音				
		/-a/	/-i/	/-u/	/-e/	/-o/
語頭	/t/, /d/	[t], [d]	[t̠e], -	[t̠s], -	[t], [d]	[t], [d]
	/s/, /z/	[s], [d̠z]	[e], [d̠z]	[s], [d̠z]	[s], [d̠z]	[s], [d̠z]
母音間	/t/, /d/	<b>[t], [d]</b>	[t̠e], -	[t̠s], -	<b>[t], [d]</b>	<b>[t], [d]</b>
	/s/, /z/	<b>[s], [z]</b>	[e], [z]	[s], [z]	<b>[s], [z]</b>	<b>[s], [z]</b>

齋藤 (2006: 89-92) に示されている通り、ザ行の子音音素 /z/ は、丁寧に発音すれば破擦音 [d̠z] および [d̠z̠] となるが、多くの場合、母音間ではそれぞれ摩擦音 [z] および [z̠] となる傾向が知られている。

表 2 からわかるように、先に示した [t], [d], [s], [z] のすべてが現れることができる環境は、「母音間」かつ「/a/, /e/, /o/ のいずれかが後続する」環境である（太字・太枠部分）。

つまり、今回の目的を達成できるような日本語の単語は、/V<sub>1</sub>CV<sub>2</sub>/ (V<sub>1</sub>=/a/, /i/, /u/, /e/, /o/, V<sub>2</sub>=/a/, /e/, /o/) という15種類の環境を持つものに限られる。

今回は分量の都合上、考えられる15種類のうち、先行母音と後続母音がともに /a/ であるものに特に注目した。すなわち /ata/ 「あた」、/ada/ 「あだ」、/asa/ 「あさ」、/aza/ 「あざ」、という4つの無意味語がターゲットとなる。

<sup>5</sup> 本稿で「東京近辺出身」というとき、その者は「0歳から18歳のあいだで最も長く生活していた地域が東京・埼玉・千葉・神奈川のいずれかであり、東京方言の話者であると自覚している」とこととする。

なお、2音節の語を下降調で発話すると、2音節目が弱く不明瞭になることがある。本論文では扱わないが、後続母音のフォルマント構造に注目する実験も同時におこなったため、2音節目が明瞭であることが望ましい。よって今回の実験では、4つの語すべて上昇調で発話した音声を用いることとした。

### 2.1.3. 録音

十分静かな環境で録音がおこなわれた。3名の発話者それぞれ（以下 S1、S2、S3）が、通常の発話とささやき発話の両方で、4つの語を発話した。1つの語につき、S2とS3は6回ずつ、S1は時間の都合上4回ずつ発話した。録音の詳しい流れは巻末の付録に記載する。

合計 128 個の録音サンプルが得られた。内訳は以下の通りである。「発話のタイプ」は「通常の発話」と「ささやき発話」、「単語」は「あた」「あだ」「あさ」「あざ」を指す。

発話者 S1 :	発話のタイプ 2 個	×	単語 4 個	×	4 回	=	32 個
発話者 S2 :	発話のタイプ 2 個	×	単語 4 個	×	6 回	=	48 個
発話者 S3 :	発話のタイプ 2 個	×	単語 4 個	×	6 回	=	48 個
合計 :							128 個

### 2.1.4. 聴取実験に使用する刺激の作成

得られた録音サンプル 128 個すべてを聴取者に聞いてもらうのは負担が大きすぎる。1名の発話者の1個の発話のタイプにおける1単語（たとえば発話者 S1 のささやき発話の「あた」）あたり4ないし6個のサンプルが存在するが、そのうちの1個のみを聞いてもらうことにした。こうすることで、使用するサンプルの数は24個（発話者3人×発話のタイプ2個×単語4個）になる。

4ないし6個のうち1個を選ぶ際には、最も良好と思われるものを選んだ。具体的な選定の過程は2.1.4.1節および2.1.4.2節に示す。

選定した24個のサンプルすべてに対し、Praatを用いて平均dB値を60dBに正規化した。これを聴取実験の刺激として使用した。

#### 2.1.4.1. 通常の発話のサンプル選定

通常の発話については、録音環境の不具合や調音の大きな失敗を除けば、聴取に大きな困難が生じるケースは起こらないと想定した。そのため、次のような基準にしたがって筆者の主観で選定をおこなった。

まず筆者の聴覚印象で判断して、録音環境および発声の不具合（周囲の雑音・過度な反響・喉のかすれなど）があるものを除外した。次に聴覚印象とスペクトログラムの見た目、子音部分が典型的なものであるかを確認した。特にサ行の子音部分が歯茎音ではなく歯音の

ように聞こえるものや、破擦化しているもの、狭めが弱く接近音に近く聞こえるものなどを除外した。さらに、狭帯域スペクトログラムを見て、前後の母音部分のエネルギーが低く、フォルマント構造が不鮮明なものを除外した。以上の点で特に差がない場合には、録音順がより早いほうを選択した。

#### 2.1.4.2. ささやき発話のサンプル選定

基本的にささやき発話は、通常の発話と比べて全体のエネルギーが低いため、録音環境の様々な要因の影響を受けやすい。著しく判別が困難なものについては、録音環境の不備も考えられるため、筆者を含めた聴取者3名による次のような聞き分けの予備調査をおこない、最も成績が高かったものを良好なサンプルとして選定することとした。

2.1.3 節で得られたささやき発話のサンプルは計 64 個である（発話者 S1 が 4 語×4 回、発話者 S2 と S3 が 4 語×6 回）。これら 64 個のサンプルを 1 つずつ出題し、有声音素・無声音素のどちらに聞こえたかを選択してもらった。加えて、その選択にどれくらい自信があるかを 5 段階（5 が最も確信している）で評価してもらった<sup>6</sup>。

予備調査の結果「聴取者 3 名全員が正答」かつ「自信度 4 以上」となったサンプルを、最も良好なものと判断し選択した。該当するサンプルが 1 つもない場合は、最もそれに近いものを選んだ。2 つ以上のサンプルが該当する場合は、2.1.4.1 節で挙げたのと同様の基準（子音部分の特徴・母音部分のフォルマントなど）に照らして、より良いほうを選択した。

#### 2.1.5. 聴取者

2.1.4 節の手順で選定した 24 個の刺激を聞いていただく聴取者として、東京近辺出身の 20 代の男性 6 名、女性 9 名に協力を依頼した<sup>7</sup>。それぞれ L1～L15 と呼ぶ。このなかに 2.1.4.2 節に示した予備調査の協力者は含まれない。

#### 2.1.6. 聴取実験

聴取実験実施時（2020 年 11 月）は新型コロナウイルス感染症の影響で、筆者が所属する大学の実験室などで実験を実施することが困難なため、やむを得ず遠隔での実施となった。ブラウザアプリによる調査フォームを作成し（図 2 参照）、オンラインミーティングシステム Zoom を通じて実験実施者が指示をしながら、実験協力者にアプリを操作してもらう形式をとった。協力者には、雑音の少ない環境の確保と、必ずイヤホンまたはヘッドホンを使用することを依頼した<sup>8</sup>。

<sup>6</sup> 予備調査で使用した調査フォームは、2.1.6 節の本調査（選択式）で使用したのとはほぼ同じものである。

<sup>7</sup> このなかには、発話者 S1～S3 のすべて、またはいずれかと面識がある者も含まれていた。

<sup>8</sup> 本来、聴取実験の環境は聴取者全体でそろえるのが理想的と思われる。今回は情勢的に妥協せざるを得なかったというのは事実だが、必ずしもこれによってデータの信憑性が著しく損なわれたとは考えない。今回の調査は、ある人物が複数の音声を聞き分けることができるかが主眼であり、その人物の聴取環境が複数の音声を聞く間一貫して同じであれば、調査の目的は果たされるからである。

<p><b>本調査(記述式)</b></p> <p><b>1問目</b></p> <p>音声は2回連続で再生されます。 再生ボタンはあと3回押せます。</p> <p><input type="button" value="再生"/></p> <p>クリックから1秒後に再生されます</p> <p>回答: <input type="text"/></p> <p>自信度: ←自信なし ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 自信あり→</p> <p><input type="button" value="次へ"/></p>	<p><b>本調査(選択式)</b></p> <p><b>1問目</b></p> <p>音声は2回連続で再生されます。 再生ボタンはあと3回押せます。</p> <p><input type="button" value="再生"/></p> <p>クリックから1秒後に再生されます</p> <p>回答: ○あた ○あだ</p> <p>自信度: ←自信なし ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 自信あり→</p> <p><input type="button" value="次へ"/></p>
---	--

図 2 調査フォーム (左: 記述式、右: 選択式)

パフォーマンスの信憑性を高めるため、聞こえた単語をひらがなで書き取る記述式と、有声・無声音素の二択による選択式の2つの形式で調査をおこなった。2.1.4 節の手順で作成した刺激 24 個を記述式の聴取実験でランダムに一通り聞いてもらい、さらに選択式の聴取実験でも一通り聞いてもらった。すなわち、聴取者 1 人あたりのべ 48 個の刺激を聞いてもらった<sup>9</sup>。

聴取者が再生ボタンを一度押すと、音声 が 2 回連続で再生される。再生ボタンは計 3 回まで押せるように設定した(すなわち同一の音声 を最高 6 回聴くことができる)。聴取者は、流れた音声 が何と聞こえたかを回答したのち、その回答に対する自信度を 5 段階で評価するよう指示された。「まぎれもなくそれ以外には聞こえない」という場合は「5」、「判断に大きな困難があり、ほとんど当てずっぽうに近い」ような場合は「1」を選んでもらうように説明した。

なお自信度が低い場合には、何かと迷ったのであれば、その対象も併せて入力してもらった(記述式のみ)。たとえば「あさ」だと思ふけれど「あざ」のようにも聞こえて迷った、という場合には「あさ あざ?」のように入力してもらい、正誤判定の際には「あさ」を回答として扱った。

音量の調整と操作方法の確認のため、通常の発話とささやき発話それぞれの調査開始前に、発話者 S1 が発話した /ike/ 「いけ」という語を練習問題として出題した。

調査終了後、特に感想などがあれば口頭で報告してもらった。

<sup>9</sup> 本論文では扱っていないが、実際にはもう一つの実験と併せておこなった。実際の流れは次のとおりである。通常の発話(記述式)、ささやき発話(記述式)、通常の発話(択一式)、ささやき発話(択一式)、という順で計 4 ブロックおこなった。1 ブロックあたり 72 個ずつのトークンを、ランダムに 1 回ずつ出題した。所要時間は 1 ブロックあたり 10~20 分程度、ブロックの間で数分間の休憩を設けた。実験全体の所要時間は聴取者 1 人当たり 50 分~100 分程度であった。

## 2.2. 結果

聴取実験の回答結果を表 3 に示す。ここでは、15 名の聴取者それぞれの、記述式の回答と選択式の回答を合わせたのべ回答数、合計 30 名分として扱う。

表の見方を、5 行目を例に取って説明する。なお表の外側、右端の数字および下端のアルファベットは、論文中でセルの位置を指定するために用いる行番号および列番号である。

セル座標 (5A) および (5B) は、発話者 S1 がささやき発話で発した /asa/ を聞いて「あさ」と回答したのが 30 名で「あざ」と回答したのが 0 名、ということを表す<sup>10</sup>。同様に、発話者 S2 の音声ではそれぞれ 29 名 (5D) と 1 名 (5E)、発話者 S3 の音声ではそれぞれ 23 名 (5G) と 7 名 (5H)、そして発話者 3 名の合計ではそれぞれ 82 名 (5J) と 8 名 (5K)、すなわち 90 名中 8 名が無声音素 /s/ を有声音素 /z/ と聞き間違えたということである (表中の網掛け・太字部分が誤答に当たる)。

列番号 C, F, I の「平均自信度」は、ひとつの刺激に対する回答の自信度を聴取者 30 名分 (のべ) で平均した値を表す。列番号 L の数値は、それらをさらに発話者 3 名分で平均したものである (いずれも小数点第三位以下を四捨五入)。

表 3 通常の発話とささやき発話における、各音素に対する回答数 (のべ) と平均自信度

発話者		S1			S2			S3			S1~S3 合計			
発話のタイプ	出題された語	回答		平均自信度	回答		平均自信度	回答		平均自信度	回答		平均自信度	
		無声	有声		無声	有声		無声	有声		無声	有声		
通常の発話	/asa/	30	<b>0</b>	4.93	30	<b>0</b>	5.00	30	<b>0</b>	5.00	90	<b>0</b>	4.98	(1)
	/aza/	<b>0</b>	30	5.00	<b>0</b>	30	4.97	<b>0</b>	30	5.00	<b>0</b>	90	4.99	(2)
	/ata/	30	<b>0</b>	5.00	30	<b>0</b>	4.93	30	<b>0</b>	4.93	90	<b>0</b>	4.96	(3)
	/ada/	<b>0</b>	30	4.90	<b>0</b>	30	4.87	<b>0</b>	30	4.87	<b>0</b>	90	4.88	(4)
ささやき発話	/asa/	30	<b>0</b>	4.70	29	<b>1</b>	4.93	23	<b>7</b>	4.40	82	<b>8</b>	4.68	(5)
	/aza/	<b>0</b>	30	4.70	<b>14</b>	16	4.43	<b>0</b>	30	4.87	<b>14</b>	76	4.67	(6)
	/ata/	30	<b>0</b>	4.97	30	<b>0</b>	4.90	30	<b>0</b>	4.97	90	<b>0</b>	4.94	(7)
	/ada/	<b>1</b>	29	4.57	<b>3</b>	27	4.30	<b>3</b>	27	4.30	<b>7</b>	83	4.39	(8)
		(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	

<sup>10</sup> 説明を簡潔にするためにこのように書いたが、やや不正確である。ささやき発話の音声で、記述式の回答のなかに「ばさ」のように語頭位置に子音を聞き取る回答が数個あった (録音環境の雑音や音声の切り出し方の関係と思われる) が、母音間の子音部分の有声・無声を正しく回答していれば正答扱いとした (「ばさ」ならば「無声」にカウント)。なお、そのようなケースで自信度が特に低いもの (たとえば「あさ」と「ばさ」で迷って自信度 1、など) は見られなかったため、評価された自信度はすべて母音間の子音部分の聞き分けに関与するものと解釈し、そのまま平均に算入した。

通常の発話（行番号 1～4）では想定通り、有声・無声音素間での聞き間違いはひとつも起こらなかった。誰が発話しても誰が聞いても、有声音素は有声音素として、無声音素は無声音素として正しく認識されたということである。

一方ささやき発話（行番号 5～8）ではいくつかの聞き間違いが発生し、次のような傾向が見られた。

発話者 3 名分の合計で見ると、無声音素を有声音素として聞き取る誤り（5K および 7K）よりも、有声音素を無声音素として聞き取る誤り（6J および 8J）のほうが多かった。

タ・ダ行（行番号 7, 8）に関しては、どの発話者でも、/t/ が /d/ と間違われる誤答は一度もなく、/d/ が /t/ と間違われる誤答（8A, 8D, 8G）のみ見られた。

サ・ザ行（行番号 5, 6）では、発話者によってかなり聞き間違われる傾向が異なることもわかった。ほとんど間違われぬ発話者（S1）もあれば、/z/ が /s/ と間違われることが多い発話者（S2）、反対に /s/ が /z/ と間違われることが多い発話者（S3）もあった。

表 2 のデータを「通常の発話とささやき発話とで成績がどの程度違うか」という視点で整理したものが、下の図 3 である。

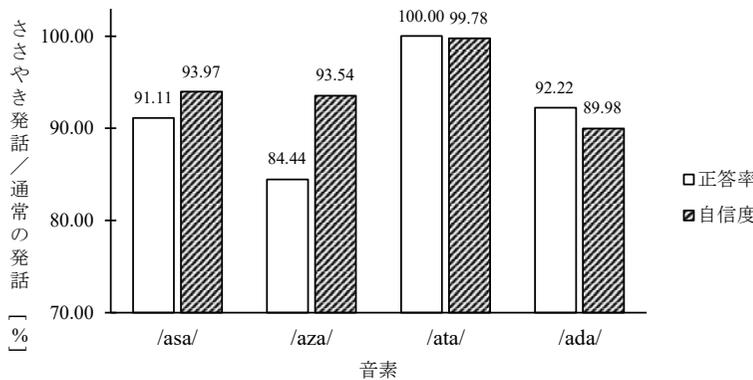


図 3 各音素における正答率と平均自信度の、通常の発話に対するささやき発話の比

ここでは、ひとつの刺激に対して 30 名のうち何名が正答したか（正答率）を用いる。

図中の数値は、ささやき発話に対する正答率・平均自信度を、通常の発話に対する正答率・平均自信度を基準とした百分率で示した値である（小数点第三位以下を四捨五入）。

たとえば一番左のグラフは、ささやき発話の /asa/ の正答率を、通常の発話の /asa/ の正答率で割った値が 91.11%、すなわち、/asa/ はささやき発話のとき通常の発話と比べて 8.9% ほど正答率が低いということを意味する。

ささやき発話のとき、通常の発話と比べて正答率が最も下がるのは /aza/ で約 15%、最も下がらないのは /ata/ で 0%であった。/ata/ では自信度もほとんど下がらず、通常の発話と同様に成績良好だった。

正答率の下がり具合を摩擦音（サ・ザ行）と閉鎖音（タ・ダ行）とで比較すると、摩擦音が約 9%～約 15%、閉鎖音が 0%～約 8%と、摩擦音のほうが大きく下がったと言える。

また、摩擦音と閉鎖音で比較したとき、正答率で見れば摩擦音のほうが大きく下がるのに対し、自信度が最も下がるのは /ada/ であった（約 10%の減少）。つまり、サ・ザ行は誤答であっても自信度は比較的高く、/ada/ は正答であっても自信度が低い傾向にあった。

### 3. 考察

#### 3.1. ささやき発話における聞き分けの正答率を下げる要因

2.2 節で示したように、ささやき発話における有声・無声音素の聞き分けについて、閉鎖音と摩擦音とで比較すると、正答率は摩擦音のほうが平均 10%程度低く、当初の仮説（1.2 節末参照）が立証された。この結果は、Dannenbring (1980) [英語：語頭位置]、吉岡・村瀬 (1992) [日本語：母音間位置]、Mills (2009) [英語：語頭位置] などと一致した。

この節では、こうした聞き分け精度の差が生じる理由を指摘した Mills (2009) [英語：語頭位置] を取り上げ、その説明をより正確なものにすることを試みる。

Mills (2009) は、ささやき発話において摩擦音に比べて閉鎖音のほうが聞き分けの精度が高くなる要因として、開放強度に言及している。閉鎖音では、閉鎖区間において声道内の圧力が上昇するため、より閉鎖時間の長い無声音素のほうが開放時のエネルギーが大きくなり、有声音素と無声音素の音響的な差異が生じる。摩擦音では声道内が完全な閉鎖空間とはならず、このような差異が生じないために、閉鎖音よりは聞き分けの精度が低くなると分析している。

確かに東川 (1994) も、ささやき発話における語頭位置の閉鎖音の調音時の口腔内圧が、無声音素のとき有声音素のときよりも有意に高いことを指摘しており、Mills (2009) の説を調音の面から裏付けている。

ただし Mills (2009) の指摘は説明が不足していると筆者は考える。その理由と、より詳しい分析を次に展開する。

まず川原ほか (2018) が挙げた、通常の発話でよく見られる有声・無声音素間の 7 つの音響的差異を取り上げる。

- ① 声帯振動の有無
- ② 子音区間の長さ（有声<無声）

- ③ 先行母音の長さ（有声＞無声）
- ④ 周囲の母音の F0 の高さ（有声＜無声）
- ⑤ 周囲の母音の F1 の高さ（有声＜無声）
- ⑥ 無声子音の後続母音で見られる F1 のカットバック<sup>11</sup>
- ⑦ 閉鎖音の開放強度（有声＜無声）

以上の7点を、閉鎖音と摩擦音に分けて下表に示す（「差あり」は音響的差異があること、「N/A」は定義上存在しないことを示す）。

表 4 通常の発話で見られる有声子音音素と無声子音音素との間の差異

	通常の発話	
	有声閉鎖 vs. 無声閉鎖	有声摩擦 vs. 無声摩擦
① 声帯振動	差あり	差あり
② 子音の長さ	差あり [英]	差あり [英]
③ 母音の長さ	差あり [英]	差あり [英]
④ F0 の高さ	差あり	差あり
⑤ F1 の高さ	差あり	差あり
⑥ F1 カットバック	差あり	差あり
⑦ 開放強度	差あり	N/A

これらのうち、「ささやき発話において摩擦音のほうが有声・無声音素の聞き分けが困難になる」ことの要因となり得るものを考える。

① 声帯振動の有無

ささやき発話においては有声子音音素が声帯振動を失っているため、この差異を有声・無声音素の聞き分けの手がかりとして用いることはできない。この点は有声閉鎖子音であろうと有声摩擦子音であろうと同様であり、閉鎖音と摩擦音で聞き分け精度の差が生じる要因ではないことが考えられる。

② 子音区間の長さ（有声＜無声）

通常の発話の場合と同様、ささやき発話においても子音区間の長さの差異（有声＜無声）が見られることを Pernell et al. (1977) [英語：母音間位置の阻害音]、Tartter (1989) [英語：

<sup>11</sup> 後述のように、語頭位置の無声子音において、後続母音の F1 のエネルギーの開始が、高次フォルマントのエネルギーの開始より遅れること（ケント&リード著・荒井ほか監訳 1996: 146-147）。

語頭位置の各子音]、Mills (2003) [スコットランド英語：母音前および後の阻害音] などが示した。ささやき発話においてもこの差異を有声・無声音素の聞き分けに用いることができると考えられるが、この差異は閉鎖音と摩擦音のどちらの場合にも観察されるため、閉鎖音と摩擦音で聞き分け精度の差が生じる要因とはなっていないことが考えられる。

### ③ 先行母音の長さ（有声＞無声）

通常の発話の場合と同様、ささやき発話においても先行母音の長さの差異（有声＞無声）が見られることを Mills (2003) が示した。ささやき発話においてもこの差異を有声・無声音素の聞き分けに用いることができると考えられるが、この差異は閉鎖音と摩擦音のどちらの場合にも観察されるため、閉鎖音と摩擦音で聞き分け精度の差が生じる要因とはなっていないことが考えられる。

### ④ 周囲の母音の F0 の高さ（有声＜無声）

通常の発話の場合、有声子音の周囲の母音よりも無声子音の周囲の母音のほうが F0 が高い。一方ささやき発話においては定義上 F0 が存在しないため、この差異を有声・無声音素の聞き分けに用いることはできない。この点は閉鎖音であろうと摩擦音であろうと同様であり、閉鎖音と摩擦音で聞き分け精度の差が生じる要因とはなっていないことが考えられる。

### ⑤ 周囲の母音の F1 の高さ（有声＜無声）

通常の発話の場合、有声子音の周囲の母音よりも無声子音の周囲の母音のほうが F1 が高い。ささやき発話中の有声・無声音素に関して周囲の母音の F1 の高さの差異に言及した先行研究は見つかっていないが、ささやき発話においてはフォルマントの帯域幅が広く、フォルマントの値も通常の発話に比べて不安定である (Mills 2003) ことから、この差異を有声・無声音素の聞き分けの確実な手がかりとして用いることはできないと予想される。したがって、この点が閉鎖音と摩擦音で聞き分け精度の差が生じる要因とはなっていないと考えられる。

### ⑥ 無声子音の後続母音で見られる F1 のカットバック

通常の発話では語頭位置の無声子音において、後続母音の F1 のエネルギーの開始が、高次フォルマントのエネルギーの開始より遅れる。一方ささやき発話においては、この F1 カットバックは閉鎖音ではほとんど見られず、摩擦音では比較的観察されるものの安定的ではない (Tartter 1989)。よって、F1 カットバックに関する差異を（当該子音が閉鎖音であろうと摩擦音であろうと）有声・無声音素の聞き分けの確実な手がかりとして用いることはできず、この点が閉鎖音と摩擦音で聞き分け精度の差が生じる要因とはなっていないことが考えられる。

⑦ 閉鎖音の開放強度（有声<無声）

通常の発話の場合と同様、ささやき発話においても無声閉鎖子音の解放強度の方が有声閉鎖子音（音素）の解放強度より高いことを Tartter (1989) が示した。一方、摩擦音には解放強度が定義上存在しない。よって、この差異は摩擦音の場合には有声・無声音素の聞き分けの手がかりとして用いることができず、閉鎖音の場合にのみ用いることができる。

以上をまとめると下表のようになる。

表 5 通常の発話とささやき発話で見られる有声子音音素と無声子音音素との間の差異

	通常の発話		ささやき発話	
	有声閉鎖 vs. 無声閉鎖	有声摩擦 vs. 無声摩擦	有声閉鎖 vs. 無声閉鎖	有声摩擦 vs. 無声摩擦
① 声帯振動	差あり	差あり	N/A	N/A
② 子音の長さ	差あり [英]	差あり [英]	差あり [英]	差あり [英]
③ 母音の長さ	差あり [英]	差あり [英]	差あり [英]	差あり [英]
④ F0 の高さ	差あり	差あり	N/A	N/A
⑤ F1 の高さ	差あり	差あり	差なし?	差なし?
⑥ F1 カットバック	差あり	差あり	ほぼ差なし [英]	差あり (不安定) [英]
⑦ 開放強度	差あり	N/A	差あり [英]	N/A

通常の発話においては有声・無声音素を聞き分けるために、閉鎖音は①～⑦、摩擦音は①～⑥を手がかりとして用いることができる。一方、ささやき発話においては有声・無声音素を聞き分けるために、閉鎖音は②③⑦、摩擦音は②③を手がかりとして用いることができる。

Mills (2009) の指摘における不足は次の点である。通常の発話では、⑦を手がかりとして使えない摩擦音に関しても、閉鎖音と同程度に有声・無声音素の聞き分けができる。それに対してささやき発話中の摩擦音に関しては、有声・無声音素の聞き分けが困難になる。摩擦音の有声・無声音素の聞き分けに関して「手がかりとして開放強度を使えない」ことは通常の発話でもささやき発話でも同じである。「摩擦音では開放強度の差異という手がかりを使えない」という Mills (2009) の指摘は、ささやき発話中でのみ摩擦音の聞き分け成績が下がる理由としては不十分である。

これを説明するために、通常の発話において①および④が他よりも極めて重要な手がかりとなっている可能性を考える。有声・無声音素の聞き分けの手がかりは、通常の発話では、閉鎖音は7つ、摩擦音は6つである一方、ささやき発話では閉鎖音3つ、摩擦音2つである。ささやき発話のほうが使える手がかりの全体数が少ないため、1つ欠けることの影響は

通常の発話よりも大きい。加えて、例えば①が7つの手がかりのうちかなりの重要性を占めることは十分考えられる。そうだとすれば、ささやき発話ではそれを失う分、さらに⑦の比重が大きくなる<sup>12</sup>。

以上の要因で、摩擦音のほうが閉鎖音よりも有声・無声音素の聞き分け成績が低くなつたと考えられる。

### 3.2. 個人差

本文の図表中には聴取者別のデータは示さなかったが、聞き間違いの傾向には聴取者間の個人差もかなりあった(巻末の付録に示す実験結果の全データを参照すると、聴取者別の回答傾向も見ることができる)。感想を聞いてみても、タ・ダ行のほうが聞き分けに苦勞したという人もいれば、反対にサ・ザ行のほうが苦勞したという人もいた。同じ刺激を聞いているのに傾向が分かれるということは、聴取者によって聞き分けの手がかりとしている要素が異なると言えるだろう。

また全体として、有声音素と無声音素とでは、無声音素のほうが正答率が高く、有声音素を無声音素として聞き取る誤りのほうがその逆よりも多かったが、この傾向はどの発話者から得られた刺激を聞いたかによってかなり差があった。発話者 S3 から得られた刺激に関して言えば、吉岡・村瀬 (1992) [日本語：母音間位置] の結果と同様、無声音素 /s/ を有声音素 /z/ と聞き取られる誤りのみが起こり (2.2 節、表 3 のセル座標 (5H) を参照)、今回の調査結果の全体的傾向とは逆であった。

このような発話者間での傾向の差は摩擦音にのみ見られ、閉鎖音では見られなかった。このことは、ささやき発話中の摩擦音は、人によって有声・無声の出し分け方が異なる可能性を示唆する。

## 4. まとめと今後の課題

本研究では、日本語の語中(母音間)位置の歯茎閉鎖音と歯茎摩擦音を取り上げ、ささやき発話における有声・無声音素の聞き分けについて調査した。その結果、ある程度の精度で

<sup>12</sup> さらに、ささやき発話において⑦の差異が通常の発話のときよりも大きくなっている可能性もある。Osfer (2011) [英語：母音間位置] は、ささやき発話における有声音素と無声音素の調音上の差異について、声門の状態の差異に起因して自動的に生じたものではなく、聞き手にとって判別しやすい音声にするべく、発話者自身が(無意識的に)調音器官に対して積極的な操作を加えている(Hyperarticulation)ということを実証した。すなわち、発話者は「ささやき発話では通常の発話に比べて有声・無声音素の聞き分けが困難になる」ということを経験的に知っていて、聞き手側を指向して「はっきり出し分ける」という選択を無意識のうちにしているということである。これが事実ならば、ささやき発話において有声閉鎖音と無声閉鎖音を調音する際に、開放強度の差異をより大きくしている(無声閉鎖音の開放をより強めるなど)可能性が考えられる。ただし、これはあくまでも音響的・調音的な側面に基づいた仮説である。ささやき発話において有声閉鎖音と無声閉鎖音を調音する際に通常の発話のときよりも開放強度の差異をより大きくしていることが実際に認められたとしても、その差異が聞き分けに寄与していると直ちに判断することはできない。実際に開放強度の差異が聞き分けの手がかりとして使われていることを示すためには、子音区間の情報を奪うような加工を施した音声を用いて聴取実験をおこなう必要がある。

有声・無声音素を聞き分けることができたが、摩擦音のほうが閉鎖音よりも聞き分けの成績が低い傾向にあった。これは、声帯振動を失った状態のささやき発話においては、開放強度の差異という手がかりを使えるか否かによる影響が通常の発話のときよりも大きく、閉鎖音のほうが聞き分けに有利になるとの考察を展開した。

また、聴取実験の結果はどれも発話者および聴取者による差が大きく、ささやき発話における有声音素と無声音素の出し分け方および聞き分け方は個人差が大きいことも示唆された。

今後の課題としては次のことが挙げられる。本研究は「ささやき発話に特有の」聞き分けの手がかりが存在する可能性に触れていない点で不足している。3節では主に通常の発話で用いられる手がかりを前提として考察を展開したが、3.1節の脚注で触れた *hyperarticulation* などによって、通常の発話には見られないがささやき発話には見られるような音響的差異が生じている可能性は否定できない。実際にはこの点についても精査する必要がある。

ただし、調音の仕組みは複合的であり、1つのパラメータのみを動かして他をすべて固定することは不可能である。同様に知覚のメカニズムも、ある1つの音響的特性のみを捉えて聞き分けているわけではないはずである。これが知覚的側面を扱う研究の難しい部分である。

今回、様々な制約のなかで調査を実施したこともあり、あくまで可能性を提示するにとどまった部分もある。聴取者と発話者をさらに増やし、より理想的な環境で実験をおこなえば、個人差についても、より体系的に示唆を与えることができるだろう。発話者間で見られた傾向の差異は、筆者の想像以上であった。今後の課題として、具体的な「出し分け方」の個人差について、調音的な観点からも掘り下げることができるとなおよいと考えている。

また今回は分量の都合上、問題となる子音が母音 /a/ の間に現れる環境だけしか扱えなかった。ささやき発話において、後続母音によって有声・無声音素の聞き分け精度に差があることは、日本語以外のいくつかの先行研究で示されている<sup>13</sup>。また、破擦音を扱うこともできなかった。ささやき発話の有声・無声音素の聞き分けに関して、閉鎖音と摩擦音という対立で比較した研究は先に示した通り複数あるが、破擦音まで含めて比較したものは見つけられなかった。これらについても今後の課題としたい。

## 謝辞

実験に協力してくださった友人のみなさま、参考資料を提供してくださった東京大学大学院総合文化研究科の広瀬友紀先生、実験に関するアドバイスをくださった国際基督教大

<sup>13</sup> たとえば Munro (1990) [英語・語頭位置] では、ささやき発話における両唇閉鎖音の有声・無声音素の聞き分け正答率は、後続母音が /i/, /e/, /æ/, /u/ のとき、それぞれ 54, 60, 66, 71%であった。

学の李勝勲先生、きめ細やかなご指導と温かいサポートをしてくださった東京大学大学院人文社会系研究科の梅谷博之先生、実験フォーム作成のほか全面的に支えてくれた T.I.氏、その他ご支援くださった多くの方々に厚く感謝いたします（所属機関の情報は 2021 年 3 月時点のものを記載）。

## 参考文献

- 加藤重広・安藤智子 (2016) 『基礎から学ぶ 音声学講義』東京：研究社。
- 川原繁人・高田三枝子・松浦年男・松井理直 (2018) 「有声性の研究はなぜ重要なのか」『音声研究』22(2): 56-68.
- キャットフォード, ジョン・カニソン (竹林滋・設楽優子・内田洋子 訳) (2006) 『実践音声学入門』東京：大修館書店。
- ケント, レイ・D., チャールズ・リード (荒井隆行・菅原勉 監訳) (1996) 『音声の音響分析』東京：海文堂出版。
- 斎藤純男 (2006) 『日本語音声学入門 改訂版』東京：三省堂書店。
- 城生佰太郎 (2008) 『一般音声学講義』東京：勉誠出版。
- 城生佰太郎 (2011) 「喉頭原音」城生佰太郎・福盛貴弘・斎藤純男 (編) 『音声学基本事典』74-76. 東京：勉誠出版。
- 服部四郎 (1984) 『音声学 カセットテープ、同テキスト付』東京：岩波書店。
- 東川雅彦 (1994) 「呟語の聴覚的、音響学的ならびに空気力学的研究」『日本耳鼻咽喉科学会会報』97(7): 1268-1280.
- 吉岡博英・村瀬忍 (1992) 「有声、無声の弁別に関する構音動態—声帯振動を伴わない囁語を中心に—」『日本耳鼻咽喉科学会会報』95(10): 1520.
- Dannenbring, L. Gary (1980) Perceptual discrimination of whispered phoneme pairs. *Perceptual and Motor Skills* 51(3): 979-985.
- Kochetov, Alexei (2014) Voicing and tongue-palate contact differences in Japanese obstruents. *Journal of the Phonetic Society of Japan* 18(2): 63-76.
- Meenakshi, G. Nisha & Prasanta Kumar Ghosh (2015) A discriminative analysis within and across voiced and unvoiced consonants in neutral and whispered speech in multiple Indian languages. *INTERSPEECH-2015*: 781-785.
- Mills, Timothy Ian Pandachuck (2003) Cues to voicing contrasts in whispered Scottish obstruents. Master's thesis, University of Edinburgh.
- Mills, Timothy Ian Pandachuck (2009) Speech motor control variables in the production of voicing contrasts and emphatic accent. Doctoral dissertation, University of Edinburgh.
- Munro, J. Murray (1990) Perception of 'voicing' in whispered stops. *Phonetica* 47(3-4): 173-181.

- Osfar, Megan Jo (2011) Articulation of whispered alveolar consonants. Master's thesis, University of Illinois.
- Pernell, Martha, James D. Amerman & G. Beverly Wells (1977) Closure and constriction duration for alveolar consonants during voiced and whispered speaking conditions. *The Journal of the Acoustical Society of America* 61: 612-613.
- Skarnitzl, Radek, Pavel Šturm & Pavel Machač (2013) The phonological voicing contrast in Czech: An EPG study of phonated and whispered fricatives. *INTERSPEECH-2013*: 3191-3195.
- Styler, Will (2020) Using Praat for linguistic research. University of Colorado at Boulder Phonetics Lab. <https://wstyler.ucsd.edu/praat/> [2020年11月アクセス] .
- Tartter, C. Vivien (1989) What's in a whisper? *The Journal of the Acoustical Society of America* 86(5): 1678-1683.
- Yoshioka, Hirohide (2008) The role of tongue articulation for /s/ and /z/ production in whispered speech. *The Journal of the Acoustical Society of America* 123(5): 2335-2338.

## 付録

### イ) 録音機材・環境

マイク : SHURE WH20XLR ヘッドセット  
レコーダー : ZOOM H4nPro ハンディレコーダー (電池駆動)  
データ形式 : 44100Hz・16bit・WAV ファイル  
場所 : 東京大学文学部言語学研究室 演習室

### ロ) 録音の流れ

当初は、後続母音 /a/ のほかに /i/, /u/, /e/, /o/ のもの、さらに子音が語頭位置に現れるものも観察対象にする予定であったため、これらを含めた読み上げリストを作成・使用した。

C = /t/, /d/, /s/, /z/ に対して、/aCa/, /aCi/, /aCu/, /aCe/, /aCo/, /Cai/, /Cii/, /Cui/, /Cei/, /Coi/ という 10 個の母音環境を設定した (実際には分量の都合上、これら 10 個の母音環境のうち C が /aCa/ の環境に現れるサンプルのみを聴取実験に使うことになった)。後半の 5 個 (子音が語頭位置のもの) は下降調<sup>14</sup>のみ、前半の 5 個 (子音が母音間位置のもの) は上昇調・下降調の両方をリストに含めた。なお日本語の /di/ と /du/ は存在しないものとみなしてある (2.1.2 節の表 2 を参照)。

よって、読み上げリストに含まれる語は以下の計 54 個となった (ただし、実際の聴取実験ではこれら 54 個のうち「あた、あだ、あさ、あざ」の 4 個だけに関する刺激を用いた)。

<sup>14</sup> ここでの「下降調」は、第一モーラのほうが第二モーラよりも高い、いわゆる HL のアクセントを指す。

あた、あだ、あさ、あざ、あて、あで、あせ、あぜ、あと、あど、あそ、あぞ、あつ、あす、あず、あち、あし、あじ (18個それぞれ上昇調・下降調で計36個)  
 たい、だい、さい、ざい、てい、でい、せい、ぜい、とい、どい、そい、ぞい、つい、すい、ずい、ちい、しい、じい (下降調のみ18個)

発話者 S2 と S3 に対し、読み上げリスト内のターゲット語をランダムな順番で画面に表示した。画面が表示された後まず通し番号を読み上げてもらい、それに続けてターゲット語を2回発話してもらった。上昇調および下降調は、図Aのように折れ線を描いて示し、事前に説明した。2回の発話の間および次の語に進む際には、それぞれ1秒程度ポーズを取った。



図A 発話者に提示した画面（左：下降調の「あた」、右：上昇調の「あだ」）

こうして一連のリストを読んでもらう工程を、通常の発話とささやき発話で交互に3回ずつおこなった。これにより、すべての語(54個)に対して、通常の発話・ささやき発話の両方で、合計6回分の発音が得られた(合計648個)。一人当たりの録音の所要時間は全体を通して90分程度であった。発話者S1(筆者)は時間の都合上、同様の工程を2回ずつのみおこない、合計4回分のサンプルを得た。

得られた録音データはPraatスクリプトを使用して語ごとに切り出され、2.1.4節の処理にかけられた。

#### ハ) 聴取実験の全データ

聴取実験の結果一覧を次ページ以降に掲載する。

前半：通常の発話の刺激を用いた聴取実験の結果

後半：ささやき発話の刺激を用いた聴取実験の結果

網掛け部分は誤答を示す。



ささやき発話

聴取者	形式	音素	/asa/			/aza/			/ata/			/ada/			
			S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	
L1	記述式	発話者	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	5	5	5	5	5	5	5	3	3	3	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	4	5	4	5	5	5	5	5	5	5	4	3	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
L2	記述式	発話者	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	4	5	5	5	5	5	5	5	5	5	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	5	4	4	5	5	5	5	5	4	5	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
L3	記述式	発話者	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	2	5	4	5	3	5	5	5	5	5	3	5	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	3	5	4	2	3	4	5	5	5	4	5	5	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
L4	記述式	発話者	あさ	あさ	あざ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	4	5	3	5	3	5	5	5	5	5	3	5	
		正誤	0	0	x	0	0	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	3	5	5	5	5	5	5	5	5	5	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
L5	記述式	発話者	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	4	5	5	5	5	5	5	5	5	5	5	4	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	4	4	5	5	5	5	5	5	5	5	5	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
L6	記述式	発話者	あさ	あさ	あざ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	5	5	5	5	4	5	5	5	5	4	
		正誤	0	0	x	0	0	0	0	0	0	0	0	x	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	5	5	5	5	5	5	5	5	5	5	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
L7	記述式	発話者	あさ	あさ	あざ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	5	5	5	5	5	5	5	5	4	5	
		正誤	0	0	x	0	0	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あざ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	4	5	5	5	3	5	5	5	5	4	5	3	
		正誤	0	0	x	0	0	0	0	0	0	0	x	0	
L8	記述式	発話者	あさ	あさ	あさ	あざ	あざ	あざ?	あざ	あた	あた	あた	あだ	あだ	あだ
		回答	5	5	5	4	5	2	4	5	5	5	5	4	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	4	4	1	5	5	5	5	4	4	3	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
L9	記述式	発話者	あさ	あさ	あさ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	5	5	5	5	5	5	5	4	4	3	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	5	5	5	5	5	5	5	2	4	5	
		正誤	0	0	0	0	x	0	0	0	0	0	0	x	
L10	記述式	発話者	あさ	あさ	あさ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	5	5	5	5	5	5	5	5	5	5	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あざ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	5	5	5	5	5	5	5	5	5	5	
		正誤	0	0	x	0	x	0	0	0	0	0	0	0	
L11	記述式	発話者	あさ	あさ	あざ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	4	5	5	5	5	4	5	5	5	5	
		正誤	0	0	x	0	0	0	0	0	0	0	0	0	
	選択式	回答	あさ	あざ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	4	4	5	5	5	5	5	5	5	5	5	
		正誤	0	x	0	0	0	0	0	0	0	0	0	0	
L12	記述式	発話者	あさ	あさ	あさ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	5	5	5	5	5	4	5	5	5	4	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	4	5	5	5	5	5	5	4	4	4	
		正誤	0	0	0	0	x	0	0	0	0	0	0	0	
L13	記述式	発話者	あさ	あさ	あざ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	3	3	5	4	5	5	5	5	4	4	
		正誤	0	0	x	0	x	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	5	5	4	5	5	5	5	5	5	3	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
L14	記述式	発話者	あさ	ばさ	あさ	あざ	あさ	ばざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	5	5	5	4	5	5	5	4	5	4	
		正誤	0	x	0	0	x	0	0	0	0	0	x	x	
	選択式	回答	あさ	あさ	あさ	あざ	あさ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	5	5	5	5	5	4	4	3	5	4	
		正誤	0	0	0	0	x	0	0	0	0	x	x	x	
L15	記述式	発話者	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		回答	5	5	4	5	5	5	5	5	5	5	5	5	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	
	選択式	回答	あさ	あさ	あさ	あざ	あざ	あざ	あた	あた	あた	あだ	あだ	あだ	
		自信度	5	5	2	5	5	5	5	5	5	5	5	4	
		正誤	0	0	0	0	0	0	0	0	0	0	0	0	

# Perceptual Discrimination of Voicing Contrast of Whispered Japanese Consonants

Mutsuno HIRAOKA

zero.mh.snow97@gmail.com

**Keywords:** whisper, whispered speech, voicing, consonant, perception, experimental phonetics

## Abstract

This study discusses the following points regarding the discrimination between phonologically voiced and voiceless consonant phonemes (“voiced” and “voiceless” phonemes) in Japanese whispered speech. Most previous studies on the perception of whispered speech have dealt with only a single manner of articulation; even when they have dealt with multiple manners of articulation, few have specifically examined the differences in the accuracy of voiced vs. voiceless discrimination in terms of manner of articulation. Furthermore, few have discussed the factors that contribute to such differences. This study compares consonants with different manners of articulation (although the place of articulation is the same) to determine if there is a difference in the ease of voiced vs. voiceless discrimination depending on the manner of articulation. The results of the listening experiment using normal and whispered speech revealed that voiced and voiceless phonemes can be correctly distinguished with considerable accuracy in whispered speech, although with a lesser accuracy than in normal speech. Regarding the comparison of performances of voiced vs. voiceless discrimination between two different manners of articulation, the subjects exhibited lower performances in distinguishing between voiced and voiceless fricatives than in differentiating between plosives. We attempted to elaborate on the argument of Mills (2009) that this tendency may be related to the burst intensity of consonants. Our results also indicated considerable differences among speakers and listeners, suggesting that there are large individual differences in how to distinguish between voiced and voiceless phonemes in whispered speech. These individual differences were more conspicuous for fricatives than for plosives.

(ひらおか・むつの 東京大学卒)