

閉鎖子音における有声性の知覚メカニズム： 聴覚時間分解能との関係性

田村俊介^{†1} 伊藤一仁^{†2} 廣瀬信之^{†3} 森周司^{†4}

概要：閉鎖子音における有声性の知覚と聴覚時間分解能の関係性を探るために、/da/-/ta/系列（有声開始時間の長さを段階的に変化）の合成音声を用いて音声識別課題及び同時性判断課題を行った。同時性判断課題では、閉鎖子音における雑音部の開始と母音部の開始が同時か、非同時かを判断した。その結果、有声音と無声音の知覚が切り替わる VOT 境界と同時性判断閾値は一致しなかったが、それらの値の間に有意な正の相関が見られた。

キーワード：閉鎖子音, 有声性知覚, 有声性開始時間, 聴覚時間分解能, 同時性判断

Mechanisms underlying voicing perception in stop consonants: Relationship with auditory temporal resolution

SHUNSUKE TAMURA^{†1} KAZUHITO ITO^{†2}
NOBUYUKI HIROSE^{†3} SHUJI MORI^{†4}

Abstract: To investigate the relationship between perception of voicing for stop consonants and auditory temporal resolution, we conducted a speech identification task using /da/-/ta/ continuum (synthesized by varying the duration of voice onset time systematically) and a simultaneity judgment task in which the participant judged whether the onsets of noise part and vowel part in stop consonants were simultaneous or non-simultaneous. Although VOT boundaries did not correspond to simultaneity judgment thresholds, there was a statistically significant positive correlation between them.

Keywords: Stop consonants, Voicing perception, Voice onset time, Auditory temporal resolution, Simultaneity judgment

1. はじめに

閉鎖子音における有声音と無声音（/d/-/t/, /b/-/p/, /g/-/k/ など）は、有声開始時間（Voice Onset Time, 以下 VOT）の長さによって識別される[1]。VOT とは、声道の閉鎖開放によって生じる雑音成分と、声帯振動によって生じる母音部の開始時間差である(Fig. 1)。この VOT を連続的に変化させた合成音声/da/-/ta/を用いて聴取実験を行うと、日本語話者の場合、VOT 15~20 ms を境にその知覚は/da/から/ta/へと急激に切り替わる（カテゴリー知覚）[2]。VOT の知覚については約半世紀に渡って膨大な数の研究が行われてきたが、有声音と無声音の知覚が急激に切り替わる VOT 境界はどのようにして決定されるのか、その具体的なメカニズムは未だ明らかになっていない。

Kuhl & Miller[3]は、言語機能を有さないチンチラであっても、ヒトと同様に閉鎖子音を VOT によってカテゴリー的に識別できることを発見した。この結果から、ヒトが音素を識別する際には、聴覚系の生得的な機能を反映した知覚的不連続(natural psychophysical boundary)を手がかりに利用することを示唆した。閉鎖子音における有声音と無声音は VOT という時間的な情報によって識別されることから、幾つかの研究で、聴覚時間分解能との関わりが示唆されて

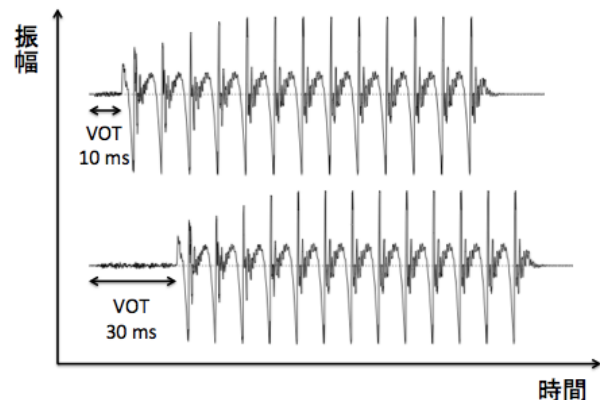


Fig. 1 /da/ (上) と/ta/ (下) の音声波形

いる[4,5]。

Pisoni[4]は、2つの純音（500, 1500 Hz）の開始時間差をずらしていき、その開始の同時性を判断する課題を行った。その結果、20 ms 程の時間差でその判断がカテゴリー的に切り替わることが示された。これは、有声音と無声音の知覚がある VOT 長で急激に切り替わることに類似している。このことから、2音の開始が同時と知覚されるか、逐次的と知覚されるかが有声音と無声音の識別に関与することを示唆している。しかし、どのように関与するのか、その具体的な部分については明確に示されていない。また、閉鎖子音を聴取する際にも、2つの純音と同様に同時性判断が行

^{†1,2,3,4}九州大学 Kyushu University

われるかどうかは定かではない。

Steinschneider et al.[5]は、閉鎖子音に対する閉鎖子音のカテゴリー的な反応が2音の開始の同時性判断に類似していることを示した。彼らは、VOT 0, 20, 40, 60, 80 ms の閉鎖子音/da/-/ta/の聴取時に、英語話者の一次聴覚野から聴覚誘発電位を測定した。その結果、VOT 40 ms 以上の刺激では、潜時 100 ms 付近に雑音成分及び母音部それぞれの開始に同期したオンセット反応が見られたが、VOT 20 ms 以下の刺激ではオンセット反応は1つしか得られなかった。つまり、VOT が長い場合には、雑音成分と母音部の開始を別々に捉えられるが、VOT が短い場合には、2つの音を別々に捉えられないのである。一般的に、音韻レベルの知覚処理は、言語音の情報が、大脳皮質の一次聴覚野を経て上側頭溝まで伝達されることで初めて行われるとされている[6]。このことから、Steinschneider et al.[5]の研究で得られた閉鎖子音に対する一次聴覚野の反応は、雑音成分と母音部の開始の同時性判断の処理を表すと考えられている[7]。

上述した2つの研究[4,5]では、2音の開始の同時性判断を行う処理と、閉鎖子音における有声性知覚の類似性が示されたのみである。よって両者の関係性については更なる検討が必要である。そこで、本研究ではまず、閉鎖子音を刺激に用い、VOT 区間に存在する雑音成分の開始と、それに続く母音部の開始の同時性を判断する課題を行い、その判断にカテゴリー性が生じるかどうかを確かめた。また、本研究では、同時性判断課題と同じ刺激を用いて音声識別課題を行い、有声音と無声音の VOT 境界と同時性判断閾値の関連性についても検討を行った。ここでは、雑音部と母音部の同時性判断閾値に差が生じるような刺激設定を行うために、雑音部の振幅が大きい刺激と小さい刺激の2つの条件で実験を行った。雑音部が大きい条件の方が、短い開始時間差のずれで、同時性の判断が切り替わると考えられるが、有声音と無声音の VOT 境界が雑音部の振幅の大小によってどのように変化するのか検討を行った。

2. 実験

2.1 実験参加者

参加者は日本語話者8名(女性3名、男性5名)で、いずれの参加者も聴覚障害を有しておらず、本実験で用いた刺激の聴き取りに困難を訴えるものはいなかった。この実験は九州大学システム情報科学研究所実験倫理委員会の承認の下で行われ、全ての実験参加者は、実験前にその内容の説明を受けた上で同意書に署名した。

2.2 実験装置

全ての実験は、防音室内で行われた。実験の刺激呈示には、パーソナルコンピュータ (DELL, Latitude E5530)、聴覚実験システム (Tucker-Davis Technologies, TDT System 3)、

ヘッドフォンアンプ (STAX, SRM-006tS)、ヘッドフォン (STAX, SR-407) を使用した。実験参加者の反応収集は、TDTに接続された反応ボックス(Tucker-Davis Technologies, RBOX)を用いて行った。刺激の呈示音圧の調整には、騒音計 (B&K, 2260)、人工内耳 (B&K, 4153)、マイクロフォン (B&K, 4192) を用いた。

2.3 刺激

同時性判断課題及び音声識別課題で用いた合成音声刺激の作成は、音声分析用のソフトウェア Praat 及び Matlab を用いて行った。合成音声は、Fig. 1 に示すように VOT 区間にあたる雑音部とそれに後続する 100 ms の母音部からなる。雑音部には、振幅が小さい刺激と大きい刺激の2種類を用いた (振幅の小さい刺激と大きい刺激の振幅比は 1:10 となるように設定した、音圧については後述する)。母音部は 100 Hz の基本周波数と 5 つのフォルマントを持ち、5 つのフォルマントは母音部の開始と同時に生起させた。第 1, 4, 5 フォルマントの周波数は固定値とし、それぞれ 600 Hz, 3300 Hz, 3750 Hz に設定した。第 2, 3 フォルマントの周波数は、生起から 40 ms の間に 1600 Hz から 1200 Hz, 3000 Hz から 2500 Hz とそれぞれ変化させ、その後は固定値であった。帯気雑音部には 1 ms の立ち上がり/立ち下がり、母音部には 1 ms の立ち上がり、10 ms の立ち下がりを付加した。

同時性判断課題では、実験前に行った予備実験の結果をもとに、実験参加者ごとに7つの刺激の VOT 長 (10~50 ms の間の刺激を 5 ms 刻みで用いた) を決定した。また、音声識別課題では、VOT 0~30 ms (5 ms 刻み) の刺激を用いた。

刺激の呈示音圧は、同時性判断課題及び音声識別課題ともに、最大で 85 dB SPL(A)になるように調整し、ヘッドフォンから左耳に呈示した。これと同じ設定で雑音部に用いる2種類の刺激の音圧を測ったところ、その音圧差は約 15 dB(A)(雑音が大きい刺激は 60.5 dB(A)、小さい刺激は 44.6 dB(A))であった。

2.4 手続き

同時性判断課題は、1区間2肢強制選択課題で行われた。VOT長の異なる7つの刺激をランダムな順序で呈示していき、「雑音成分の開始と母音部の開始が同時」、あるいは「雑音成分の方が先に始まっている」の2択で答えてもらった。本セッションでは、1つの刺激につき25試行、合計で175試行を行った。また、本セッションを行う前に練習セッションを行った。練習セッションでは、7つの刺激のうち VOT が最も短い刺激を「雑音成分の開始と母音部の開始が同時」、最も長い刺激を「雑音成分の方が先に始まっている」と教示した後、この2つの刺激をランダムな順序で計20回呈示し、正答率が90%以上になるまで課題を行ってもらった。練習セッションでは正誤のフィードバックを与えた。

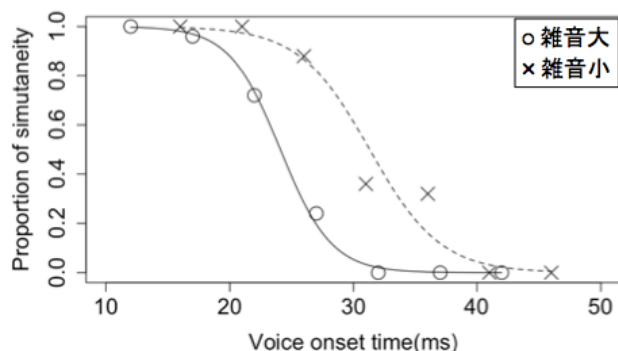


Fig.2 参加者 1 名(IH)の同時性判断課題の心理測定関数

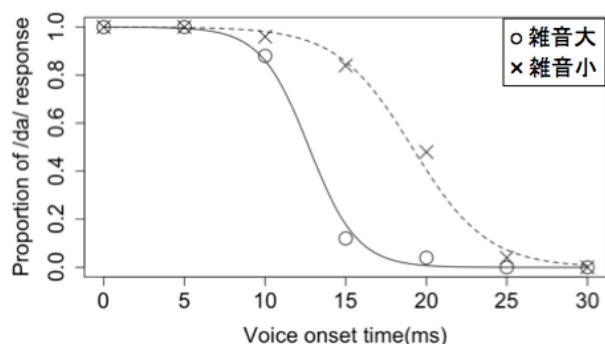


Fig.3 参加者 1 名(IH)の音声同定曲線

音声識別課題は、1 区間 2 肢強制選択課題で行った。VOT 長の異なる 7 つの刺激をランダムな順序で呈示していき、/da/あるいは/ta/の 2 択で答えてもらった。本セッションでは、1 つの刺激につき 25 試行、合計で 175 試行を行った。また、本セッションを行う前に練習セッションを行った。練習セッションでは、VOT 0 ms の刺激を /da/、30 ms の刺激を /ta/ と教示を行い、この 2 つの刺激をランダムな順序で計 20 回呈示し、正答率が 90% 以上になるまで課題を行ってもらった。この課題では正誤のフィードバックを与えた。

実験は、同時性判断課題、音声識別課題の順に行われた。また、両課題とも雑音成分の大きな刺激と小さな刺激の 2 条件で課題を行ったが、その実施順序についてはカウンターバランスを取って実験を行った。

2.5 結果と考察

Fig. 2 に、同時性判断課題で得られた実験参加者 1 名の心理測定関数を示す (7 つの刺激に対する同時性判断率のデータにロジスティック関数を当てはめた)。この心理測定関数から 50% の確率で「雑音成分の開始と母音部の開始が同時」と判断される VOT 長を求め、それを同時性判断閾値とした。同時性判断閾値は、参加者 8 名の結果を平均すると、雑音部の振幅が大きい刺激では 24.9 ms、雑音部の振幅が小さい刺激では 32 ms となった ($t(14)=4.36, p<0.01$)。

Fig. 3 に音声識別課題で得られた実験参加者 1 名の音声

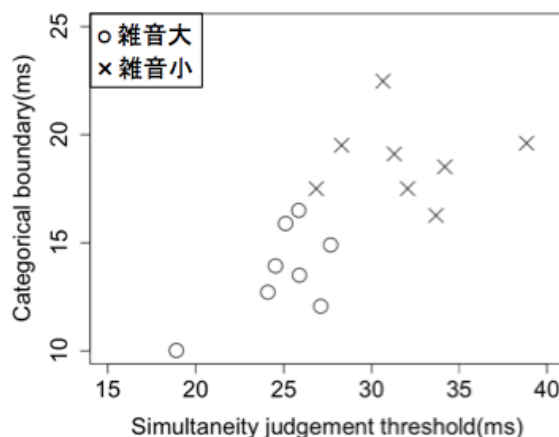


Fig. 4 同時性判断閾値と有声音と無声音の VOT 境界の散布図

同定曲線を示す (7 つの刺激に対する /da/ の反応率のデータにロジスティック関数を当てはめた)。この同定曲線から 50% の確率で /da/ と答えられる VOT 長を求め、それを /da/ と /ta/ の VOT 境界とした。VOT 境界は、参加者 8 名の結果を平均すると雑音部の振幅が大きい刺激では 13.7 ms、雑音部の振幅が小さい刺激では 18.8 ms となった ($t(14)=5.12, p<0.01$)。

Fig. 4 に同時性判断閾値と /da/ と /ta/ の VOT 境界の散布図を示す。ここでは、両者の値の間に有意な正の相関が確かめられた ($r=0.718, p<0.01$)。これは、閉鎖子音における雑音部の開始と母音部の開始を別々に捉えやすい程、無声音を知覚しやすくなることを示しており、2 音の同時性判断(聴覚時間分解能)と閉鎖子音における有声性知覚の関連性を示す結果である。Hayashi et al.[9]によって、聴覚時間分解能の低下が見られる難聴者は、健聴者に比べて、無声音を知覚するために長い VOT 長を必要とする傾向があると報告されている。本研究では、これを刺激の設定で再現し、示している。

一方で、閉鎖子音における有声性知覚のメカニズムを探る上で、同時性判断閾値(25~35 ms)が、/da/ と /ta/ の VOT 境界(10~20 ms)と一致していないことにも注目しなければならない。このことは、同時性判断閾値という知覚的不連続点がある有声音と無声音の VOT 境界を決定するものではないことを示している。

それでは、どのような手がかりを用いることで有声音と無声音の VOT 境界は決定されるのだろうか。閉鎖子音における有声性知覚の本質的なメカニズムについて考えるためには、議論を振り出しに戻す必要がある。そもそもヒトは、Kuhl & Miller[3]が言うように、知覚的不連続を有声音と無声音の識別の手がかりにしているのだろうか。先に結論から述べると、この説は支持し難いと考えられる。

閉鎖子音における有声音と無声音の VOT 境界は、日本語ではおおよそ 15~20 ms[2]、英語では 30 ms[1]、フランス

語では 0 ms[10]と、言語によって様々である。それでは、言語によって異なる知覚的不連続の手がかりが存在するのだろうか、またそれがカテゴリー性を生む手がかりに成り得るのだろうか。もし、言語ごとで手がかりが存在するとしても、本研究では、日本語話者が手がかりとする知覚的不連続を見出せていない。また、他の多くの言語でもその手がかりを見出せていないのが現状である。本研究で、同時性判断閾値と有声音と無声音の VOT 境界に相関が見られたのは、おそらく、同時性判断閾値が VOT という時間情報の伝達精度を表しているからであろう。あくまで、有声音と無声音の VOT 境界を決めるものではないと考えられる。

また、Hoonhorst et al.[10]による乳幼児の音声知覚の研究では、生後半年頃までは、どの言語圏の乳幼児も 30 ms 前後で閉鎖子音の有声音と無声音を識別するが、生後半年から 1 年の間に言語圏に依存した (成人と同じ) VOT 境界を獲得するとされている。このような乳幼児の VOT 境界が変化する時期は、喃語の発話や、音声聴取による下側頭回 (音声の生成を司る脳領域)の賦活が始まる時期と一致する [11,12]。このことから有声音と無声音の VOT 境界の決定には、音声生成機構の働きが何らかの形で関与していると考えられる。おそらく、言語機能が発達していない生後数ヶ月の乳幼児は、言語機能を持たないチンチラと同様に何らかの知覚的不連続を手がかりに閉鎖子音を識別していると考えられる。しかし、言語機能を獲得することで異なるメカニズムで閉鎖子音を識別するようになったと考えられる。近年の研究で、音韻レベルの知覚処理を行う脳領域である上側頭溝が音声聴取時に下前頭回(ブローカ領域)や運動前野という音声の生成を司る脳領域と情報のやりとりを行っていることも明らかになっている [13]。そのため今後は、閉鎖子音における有声音の知覚メカニズムが Liberman & Mattingly [14] で提唱された音声知覚の運動理論等の発話系との関わりを考慮した音声知覚モデルで説明される可能性について考えなくてはならないだろう。

謝辞 本研究は、科学研究費補助金基盤研究 (A) (課題番号: 25240023, 研究代表者: 森周司) の補助を受けました。ここに記して感謝の意を表します。

参考文献

- [1] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422.
- [2] 田村俊介, 森田大樹, 伊藤一仁, 廣瀬信之, 森周司 (2016). 有声音・無声音のカテゴリー境界と無音検出閾値の関連性. 日本音響学会春季研究会講演論文集, 485-488.
- [3] Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of Acoustical Society of America*, 63(3), 905-917.
- [4] Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (1997). Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory

- gap detection. *Journal of Acoustical Society of America*, 101(6), 3694-3705.
- [5] Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of Acoustical Society of America*, 61(5), 1352-1361.
- [6] Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of neurophysiology*, 82(5), 2346-2357.
- [7] Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in neurosciences*, 26(2), 100-107.
- [8] Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C., & Howard, M. A. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cerebral Cortex*, 15(2), 170-186.
- [9] Hayashi, A., Imaizumi, S., Harada, T., Seki, H., & Hosoi, H. (1990). Relationships between the ear's temporal window and VOT perception for normal and hearing-impaired listeners. *Ann. Bull. RILP*, 24, 145-156.
- [10] Hoonhorst, I., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2009). French native speakers in the making: From language-general to language-specific voicing boundaries. *Journal of experimental child psychology*, 104(4), 353-366.
- [11] Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, 5(11), 831-843.
- [12] Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, 17(10), 957-962.
- [13] Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
- [14] Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.