

残響環境では無声破裂音がより有声音に知覚される現象について*

○荒井隆行, 岩上純也, 大澤恵里 (上智大・理工)

1 はじめに

残響環境において、音声の知覚がどのように影響を受けるかについては様々な研究が行われている。Nábělek *et al.* は、音声知覚における残響の影響を、次の2つの側面で論じている[1]。1つは分節音内で生じる *self-masking* であり、もう1つは複数分節音にまたがって生じる *overlap-masking* である。前者については、ある分節音内において波形が時間軸方向に「崩れる」などの様子が観測される。しかし、残響時間が短い場合、時間波形の崩れが発話内容の知覚に与える影響は小さく、むしろ音が大きく知覚されて明瞭度が上がる。一方、持続的な分節音はその音響的特徴が時間と共に変化しないため、残響に対する影響が少ない。しかし、破裂音のように分節音内において音響的特徴が時間と共に変化する場合、*self-masking* の結果、長い残響によって大きな影響を受けることが予想される。

そこで本研究では、残響環境における日本語破裂音を対象に知覚実験を行った。破裂音では、口腔内のどこかで閉鎖を作り、口腔内圧が高まった状態で閉鎖が解放されることで *burst* が生じ、その後、摩擦性の雑音や帯気雑音などが生じる区間を経て声帯振動振動が始まる[2]。その *burst* から声帯振動開始までの時間は *VOT* (*voice onset time*) と呼ばれ、破裂音の有声性に対する音響関連量となっている[3]。そのため、*VOT* を変化させた連続体に対する有聲・無声の識別が残響によってどのように影響を受けるかを調べた。

2 実験

2.1 音声サンプル

日本語単音節の /ta/ から /da/ への連続体について、*VOT* を変化させながら音声を合成した。*VOT* は、0 ms から 50 ms の間を 5 ms ステップ間隔で変化させた。音声合成には

フォルマント合成器である XKL [4] を用いた。母音 /a/ については、第1フォルマント (F1) と第2フォルマント (F2) をそれぞれ 750 Hz と 1250 Hz に設定した。定常母音の前には、子音からのフォルマント遷移を設け、有聲開始時間を *VOT* として変化させた。

2.2 参加者

日本語母語話者 13 名が実験に参加した。全員が健聴であり、平均年齢は 21.4 歳であった。

2.3 残響条件

実験は防音室内で実施した。音声サンプルは PC から再生され、オーディオインタフェース (RME, Fireface 802) ならびにデジタルミキサ (Yamaha, DME 24N) を介して、4つのスピーカ (Genelec, 8020A) から音が提示されるようにした。残響にはミキサに接続されたデジタルリバーブ (Roland RSS-303) を使い、4つのスピーカからは疑似的に残響が付加された音声が出力された。4つのスピーカは半径 1.55 m の円周上に等間隔で配置され、聴取者は円の中心に着席した。残響は *Rev0* から *Rev2* まで3条件とし、残響時間は *Rev2* で 2.1 s, *Rev1* で 0.9 s, *Rev0* で 0.1 s (*Rev0* は残響無し条件) であった。

2.4 手続き

それぞれの残響条件で1セッションずつブロックに分け、また難易度が高い条件から実験を行うため、残響時間が長い順に *Rev2*, *Rev1*, *Rev0* と条件を変えながらセッションを進めた。それぞれのセッションでは、連続体を8回の繰り返しでランダムに提示し、各試行において聴取者は「タ」か「ダ」を回答した。ユーザインタフェースとして、Praat [5] を使い、本実験に入る前には練習用の課題を設けた。

* Voiceless plosives tend to be perceived as voiced in reverberant environment, by ARAI, Takayuki, IWAKAMI, Junya, and OSAWA, Eri (Sophia University).

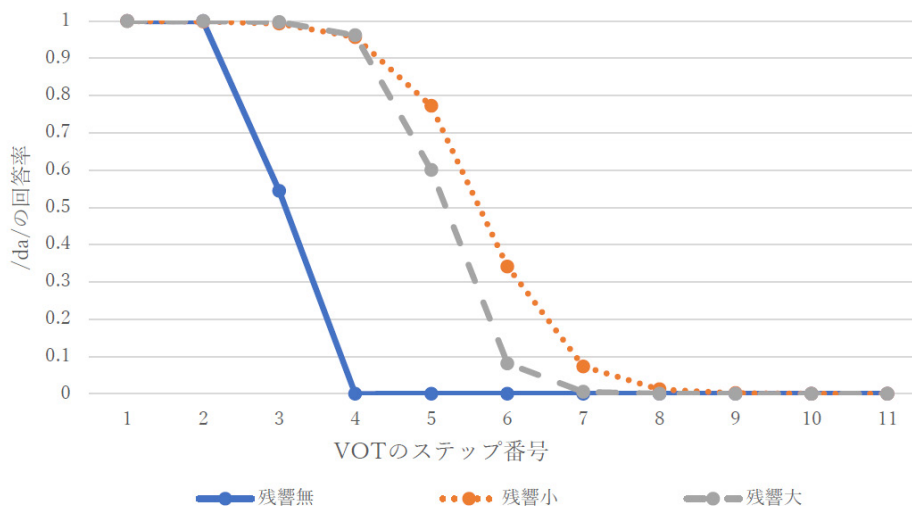


Fig. 1: Experimental results for the three reverberant conditions.

3 結果・考察

Figure 1 に3つの残響条件下における実験結果を示す。この図を見ると、VOTのステップ番号が進むに従って/dɑ/の回答率が減少している様子がわかる。

残響がない Rev0 条件においては、VOTの Step No. 2 から No. 4 にかけて急激に/dɑ/回答率が減少した。これはちょうど VOT が 10 ms 前後で有声・無声の境界となることを意味するが、先行研究ともそれほど変わらない結果であり、カテゴリーカルに知覚されている様子が観測された[6]。

一方、残響条件 Rev1 と Rev2 においては、VOT の Step No. 5 から No. 6 にかけて/dɑ/回答率 50%とクロスしており、境界が VOT = 20-25 ms へとシフトしている様子が観測された。残響が付加された時間波形やスペクトログラムを観測すると、burst は比較的容易にその位置を把握できるのに対し、有声開始の位置を判断することは難しい。より長い残響が付加されると、有声開始の位置に対する予測誤差の範囲も長く広がることが考えられる。その際、聴取者が「より長い VOT」として判断すれば、実験結果は/dɑ/回答率が下がることも予想された。しかし、実際は逆の結果が得られたことから、以前の実験でも考察した hyper-correction が起きていた可能性も考えられる[7]。

4 おわりに

本稿では、残響環境下において日本語破裂音の有声性の知覚について調べた。その結果、残響がある場合に「より有声音」として知覚されることがわかった。今回用いた2つの残響時間 (0.9 s と 2.1 s) において結果に大きな差がなかったことから、残響時間に対して結果がどのように変化するかは引き続き調査したい。

謝辞

内容の一部は、上智大学重点領域研究の助成を得た。

参考文献

- [1] Nábělek, *et al.*, *J. Acoust. Soc. Am.*, 86(4), 1259–1265, 1989.
- [2] Stevens, *Acoustic Phonetics*, MIT Press, Cambridge, 1998.
- [3] Gao and Arai, *J. of Phonetics*, 77, 100932, 2019.
- [4] Klatt, *Speech Communication Group Working Papers IV*, Research Laboratory of Electronics, MIT, Cambridge, 73–82, 1984.
- [5] Boersma, *Glott International*, 5:9/10, 341–345, 2001.
- [6] 清水克正, 音声研究, 22(2), 69–80, 2018.
- [7] Arai *et al.*, *Acoust. Sci. & Tech.*, 39, 252–255, 2018.