母語話者に対する残響下の日本語単音節・特殊拍の聴取能低下*

○荒井隆行，大澤恵里，井下田貴子（上智大・理工），
程島奈緒（東海大・情報通信）

1　はじめに

小さい室では残響によって音声明瞭度が上がる場合がある一方，大きい室になると音声明瞭度は一般的に低下する．その要因として，例えば self-masking ならびに overlap-masking の影響が指摘されている[1]．特に後者では，先行する音声に伴う残響の「尾」が後続の音声をマスキングすることによって，音声が聞き取りにくくなる．

単音節明瞭度試験を行った研究では，母音よりも子音においてより顕著な異聴が起きやすいことが英語[2,3]や日本語[4]などで報告されている．その理由として，一般的に母音はもともとインテンシティが強い傾向にあり，さらに持続性があるという性質に起因する．インテンシティが強いために，直前までの音声の残響の尾による overlap-masking に対して，対象とする母音のレベル比（speech to overlap-masking ratio, SOR [5]）が相対的に高くなる．さらに，時間変化が比較的遅く持続性が強いために，音声信号の時間構造が損なわれても母音識別への影響が小さいものと考えられる．一方，子音の場合，インテンシティが弱いものも多く，結果的に SOR は相対的に下がる．また，子音は時間とともに変化する性質を持ったものも多い．その結果，overlap-masking の影響を受けやすいのと同時に，残響によって時間構造が崩れやすい．

ところで，日本語には例えば「おばさん」と「おばあさん」，「型」と「買った」などのように，長短の対立によって語の意味が変わる長音や促音といった特殊拍が存在する．また，別の特殊拍の種別である撥音も，例えば「尼（あま）」と「按摩（あんま）」のような場合，「ん」が後続子音を長く調音することによって実現されるために，結果として長短の対立ができる．このように，日本語の特殊拍を含む音声の知覚において音声の時間構造が重要となるが，そのような場合において残響がどのような影響を及ぼすかは十分にわかっていない．

そこで，本研究では日本語を対象とし，１）単音節ならびに２）特殊拍に関する聴取実験を残響の有無の２条件で実施した．１）については，先行研究[4]と同様に単音節明瞭度試験を行った．なお，本研究では対象とする単音節の種類は増やした．２）については，長音・促音・撥音に関して連続体を対象に識別実験を行った．

2　聴取実験

2.1 音声資料

本研究では，２つのセットについて刺激音声を準備した．刺激セット１は単音節に関するもので，20 子音と 5 母音を組み合わせ，日本語の音韻規則に従って 62 モーラを音声資料として録音した（拗音は除いた）．刺激セット２は特殊拍に関するもので，長音・促音・撥音に関して連続体を作成するため，それらの特殊拍を含む無意味語を音声資料として録音した．具体的には，以下の通りである．

長音：/babaa/
促音：/appa/, /atta/, /akka/
撥音：/amma/

なお，これらの無意味語についてはいずれもアクセント型は頭高型で統一した．

音声資料として，20 代の東京方言話者（男性）の発話を使用した．キャリア文である「これから聞こえてくるのは____です」に対し，上記の単音節および特殊拍を含む無意味語を埋め込んだ形で発話してもらった．録音は，リニア PCM レコーダー（SONY PCM-D50）とマイクロフォン（SONY

* Reverberation degrades native listener's perception of Japanese mono-syllables / special-morae, by ARAI, Takayuki, OSAWA, Eri, IGETA, Takako (Sophia Univ.), and HODOSHIMA, Nao (Tokai Univ.).

ECM-MC957）を用いた（サンプリング周波数 44.1kHz，量子化 16 bit）.

2.2 刺激音声

単音節に関する刺激セット１については，キャリア文である「これから聞こえてくるのは____です」に全 62 モーラを埋め込んだ（「____」の部分がターゲット）．キャリア文には，同一の発話を用い，キャリア文前半終了時点からターゲット開始時点までの間隔を 150 ms に統一した．

特殊拍に関する刺激セット２については，まず連続体を作成した．長音については，長母音の波形を 1 周期ずつ切り取ることで短母音へと変化させた．促音については，閉鎖区間を/p/は 10 ms ずつ，/t/は 12 ms ずつ，/k/は 15 ms ずつ切り取った．撥音についても同様に，10 ms ずつ短くした．その後，単音節の場合と同じキャリア文の発話を用い，ターゲットとして連続体の 1 語 1 語を埋め込んだ．

刺激セット１・セット２ともに，同一の刺激音声が 3 回ずつランダムに提示され，その順番は実験参加者ごとに異なるランダム配列となるようにした．

2.3 参加者

聴取実験の参加者は，日本語母語話者 20 代の健聴者 11 名であった. 健聴か否かは参加者の自己申告とした．また，参加者はみな東京，及び東京近郊の出身であった．

2.4 実験方法

実験は上智大学荒井研究室の防音室で行った．防音室に設置された PC 上で，音声研究用ソフトウェア Praat [6]によりスクリプトを実行し，刺激音声の再生と実験参加者からの反応も記録した．単音節に関するセッションでは, PC 画面上にターゲットの 62 モーラ表を表示し, 聞こえたと思うモーラを 62 の選択肢の中からクリックして選んでもらった．特殊拍に関するセッションでは, PC 画面上に特殊拍に関する長短の対立をもつ無意味語を対で表示し，2 つの選択肢の中から二者択一で答えてもらった．また，長短の回答と合わせて，その回答に対する自信度も 5 段階スケール（5: 自信あり，1: 自信なし）で評価してもらった．なお，本実験に入る前に特殊拍課題のみ練習を 10 試行, 残響あり条件で行ってもらった.

刺激音声は同 PC 上に格納され，再生時には PC からオーディオインタフェース（RME Fireface 800）を介してディジタルミキシングエンジン（Yamaha DME24N）に送られた. 同機にはスピーカ（Genelec 8020A）が 4 台接続されており，実験参加者は４台のスピーカの中央に着席した．スピーカは前方左右 45 度の方向と後方左右 45 度に同心円状に配置され，各スピーカから受聴位置までの距離は約 1.3 m であった．ディジタルミキシングエンジンにはディジタルリバーブ（Roland RSS-303）が接続されており，このリバーブによって擬似的に残響を付加した．残響は予めプリセットされた長めのものを用いたが, swept sine 信号を用いて音場における受聴位置でのインパルス応答の計測の結果，残響時間は 2.7 秒であった（500-2k Hz における 1/1 オクターブ毎の分析結果の平均値）. なお, 残響なし条件はディジタルミキシングエンジンの設定によりディジタルリバーブをバイパスすることにより実現し，前方の２つのスピーカからの提示とした．この場合，受聴位置でのインパルス応答の計測を行った結果，残響時間は 0.1 秒であった．

3 結果および考察

3.1 単音節課題

まず 11 名全員に対する 62 モーラ全体の正答率は，残響なし条件で 98.0%，残響あり条件で 84.7%となり，母語話者であっても残響時間が比較的長い今回のような条件下では，単音節明瞭度が下がることが再確認された．また，残響あり条件においてどのような誤りのパターンをしていたかを調べるため, 特に有声性と調音方法に関して, Table 1 と Table 2 に，その confusion matrix を示す．

Table 1: 有声性に関する異聴（%）.
（括弧内は文献[4]の付録 B による.）

		Response	
		voiceless	voiced
Stimuli	voiceless	99.8	0.2
		(93.3)	(6.7)
	voiced	3.9	96.1
		(20.2)	(79.8)

Table 2: 調音方法に関する異聴（%）.
S: stops, F: fricatives, A: affricates, N: nasals.
（括弧内は文献[4]の付録 B による.）

		Response			
		S	F	A	N
Stimuli	S	96.7	3.0	0.0	0.0
		(75.1)	(8.3)	(2.2)	(14.5)
	F	10.3	89.1	0.0	0.6
		(15.1)	(81.3)	(3.6)	(0.0)
	A	0.6	0.0	99.4	0.0
		(0.2)	(1.1)	(98.7)	(0.0)
	N	13.6	0.6	0.0	77.0
		(21.6)	(3.0)	(0.0)	(75.5)

なお，表中の括弧内には日本語の先行研究として，文献[4]の付録 B にあるデータを参考値として載せた.

これらの表を見ると，先行研究[4]に比べて大まかに傾向が類似していた．例えば，有声性についてはほとんど誤りがなかった．一方，例えば破裂音と摩擦音の間に多少の異聴が見られた．さらに，鼻音が破裂音に間違えられるケースが散見されたが，これらはいずれも先行研究[4]と同じ傾向を示していた.

3.2 特殊拍課題

長音，促音，撥音の連続体に対して，「長いほう」の選択肢を選んだ回答率を縦軸にプロットした結果の図を，Fig. 1～Fig. 3 に示す．これらの図を見ると，すべてに共通する傾向として，以下のことが観測された:

1）残響なし条件では，連続体のステップ番号 5 ないしは 6 前後を境界としてカテゴリカルな知覚結果を示していた.
2）残響あり条件では，残響なし条件と比べて境界点が「短」回答へシフトしていた.
3）促音の場合に，残響のあり・なしであまり結果に大きな差はなかった.
4）長音の場合に，残響あり条件において回答が全般的に「長」になる傾向が強かった.

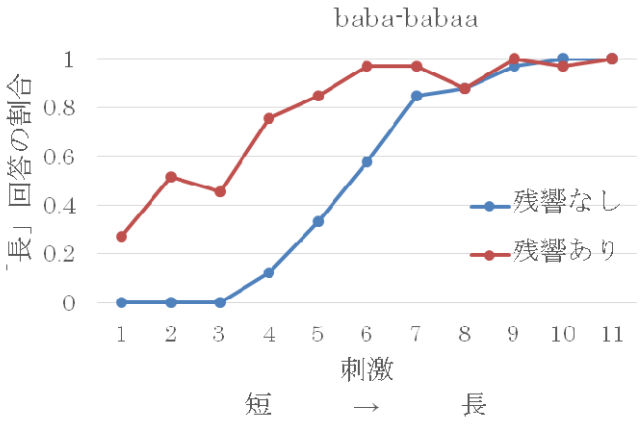

Fig. 1: 長音に関する/baba/-/babaa/ 連続体に対する「長」回答の割合.

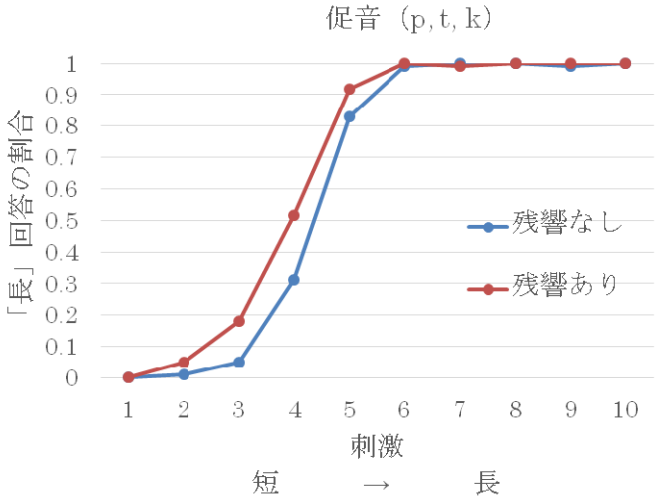

Fig. 2: 促音に関する /apa/-/appa/, /ata/-/atta/, /aka/-/akka/ 連続体に対する「長」回答の割合（3 子音に渡って平均）.

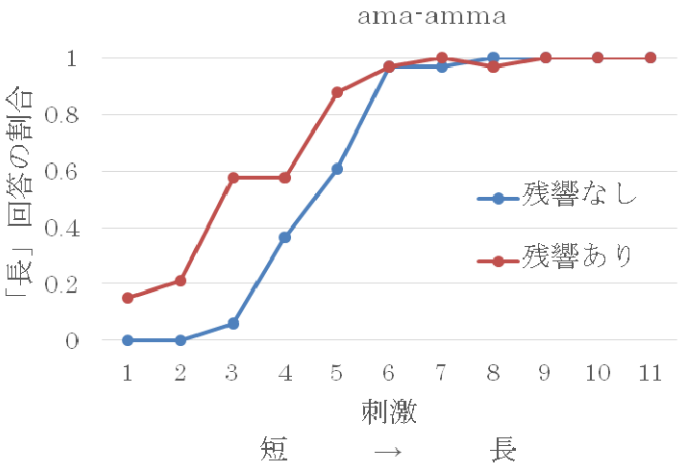

Fig. 3: 撥音に関する/ama/-/amma/ 連続体に対する「長」回答の割合.

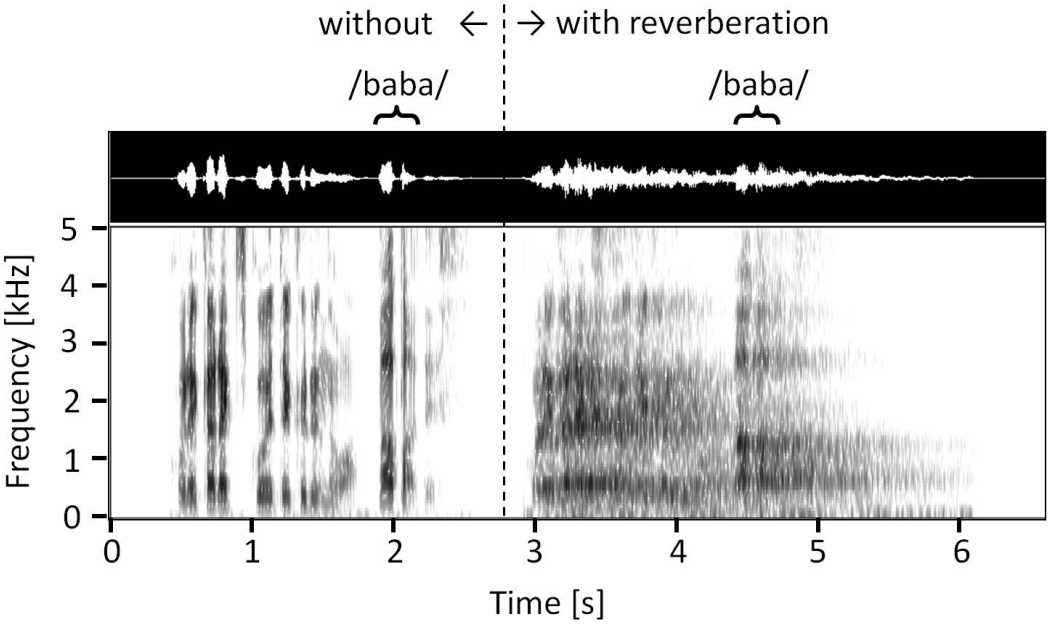

Fig. 4: キャリア文に埋め込まれた/baba/の時間波形（上段）とスペクトログラム（下段）．左半分は残響なし，右半分は残響あり．

特に今回，長音に関しては語末に位置するものを対象とした．その結果，残響によってその時間構造が崩れ，母語話者であっても短母音を長母音として聞き誤る傾向が強くなったものと考えられる．

Fig. 4 に残響の有無で，/baba/がどのように変化するかを時間波形とスペクトログラムにて示した．図の左側のように残響がない場合には，音声のセグメンテーションが比較的容易であるのに対し，残響がかかると（右半分），/baba/であるか/babaa/であるかの判断が難しくなることが視覚的にもわかる．

4　おわりに

本稿では，日本語単音節と特殊拍に関して，残響のあり・なしによる音声知覚を比較をした．単音節明瞭度試験の結果，比較的長い残響がかかると，母語話者であっても異聴を起こすことが再確認された．また，特殊拍を含む無意味語を用いた連続体に対して，識別実験を行った．その結果，持続時間が手がかりになる識別の場合，残響の影響を受けやすいことがわかった．

参考文献

[1] A. K. Nábělek, T. R. Letowski and F. M. Tucker, “Reverberant overlap- and self-masking in consonant identification,” *J. Acoust. Soc. Am.*, 86(4), 1259-1265, 1989.

[2] A. K. Nábělek and A. M. Donahue, “Perception of consonants in reverberation by native and non-native listeners,” *J. Acoust. Soc. Am.*, 75(2), 632-634, 1984.

[3] Y. Takata and A. K. Nábělek, “English consonant recognition in noise and in reverberation by Japanese and American listeners,” *J. Acoust. Soc. Am.*, 88(2), 663-666, 1990.

[4] N. Hodoshima, “Effects of listening environments, listeners and preprocessing parameters on steady-state suppression for improving speech intelligibility in reverberation,” Ph.D. Dissertation, Sophia University, 2007.

[5] T. Arai, Y. Murakami, N. Hayashi, N. Hodoshima and K. Kurisu, “Inverse correlation of intelligibility of speech in reverberation with the amount of overlap-masking,” *Acoust. Sci. & Tech.*, 28(6), 438-441, 2007.

[6] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glot International*, 5(9/10), 341-345, 2001.