残響環境下における合成音声/ba/-/wa/の識別
―フォルマント遷移に焦点を当てて―

☆川井友子（上智大），大澤恵里（昭和大），溝口愛（前橋工科大），荒井隆行（上智大）

1 はじめに

音声知覚ではスペクトルや振幅情報などの音響的特徴を手がかりとし，音声に含まれる音素を識別あるいは弁別している。特定の音響的特徴が連続的に変化する刺激に対し，同じ音素内であれば音響的特徴の変化に鈍感に反応し，異なる音素にまたがる刺激間ではその音響的特徴の変化に鋭敏に反応するようなカテゴリー知覚を示すことがある。例えば合成音声を用いて音響的物理量を連続的に変化させた刺激を作成し提示すると，ある音声カテゴリーから違うカテゴリーに属する音へと音声知覚が変化し，カテゴリー境界で判断が50%に分かれる[1]。音声におけるカテゴリー知覚の研究はこれまでに数多く存在する。閉鎖音/b/と接近音/w/では, Liberman らのフォルマント遷移の持続時間を連続的に変化させた合成音声/bɛ/-/wɛ/の刺激を用いた識別実験から，フォルマント遷移の持続時間が閉鎖音/b/と接近音/w/の識別における音響的手がかりの 1 つであるとわかった[2]。その際フォルマント遷移の持続時間が 40 ms 付近で/b/から/w/へと判断が変化した[2]。

日本語母語話者においても阿部らの研究から同様の知覚傾向が報告されている[3]。フォルマント遷移の持続時間を連続的に変化させた合成音声/ba/-/wa/刺激連続体を用いた識別と弁別課題において，若年者群では平均 41.2 ms にカテゴリー境界がみられた[3]。そのことから日本語母語話者においても，フォルマント遷移の持続時間の変化が/b/と/w/の識別あるいは弁別における音響的手がかりの 1 つになっていることがわかった[3]。

残響環境下では音声の明瞭度が低下する。残響時間(RT)が長い空間では直接音に対し遅れて到達する反射音によって，先行する子音や母音のエネルギーが後続する子音や母音に重なる“overlap-masking”や，各音素内で時間波形が崩れてエネルギー分布が不鮮明になる“self-masking”が生じる[4]。このような残響によって生じる影響は，特に高齢者や聴覚障害者，非母語話者の音声知覚に影響を与えることがわかっている[5][6][7]。

残響が子音や母音の音声知覚に与える影響については多くの調査が行われてきた[8][9]。日本語閉鎖音においては，有声開始時間(VOT)を連続的に変化させた合成音声/ta/-/da/に対し，残響が有声・無声の識別にどのような影響を与えるのかの調査がある。残響と非残響環境の両条件で VOT が長いと/ta/の回答が増加する知覚傾向があり，また非残響環境下と比較し残響環境下では/da/の回答率が50%になるカテゴリー境界が, より VOT が長い刺激ステップにシフトした[8]。

そして日本語の特殊拍においては，残響環境下では非残響環境下と比較して音声知覚が変化することがわかっている[9]。残響環境下にて子音や母音の時間長を連続的に変化させた音声刺激を用いてモーラ数の長短を識別させる課題から、長短母音と鼻音においては非残響環境下と比較して残響環境下では短い時間長の刺激ステップに長いとの回答が増加する知覚傾向が生じた[9]。つまり，音声に含まれる音響的特徴の時間的変化は残響の付加によって大きく影響を受けることがあり，それによって音声知覚にも変化が生じることがある。そのことから，フォルマント遷移の持続時間においても残響の影響を受けることで，知覚に変化が生じる可能性がある。

英語においては，フォルマント遷移の持続時間を変化させた合成音声による連続体に対して単母音/a/と二重母音/aɪ/のどちらであるかを問う識別実験の結果によって，健聴者群では残響環境条件と非残響環境条件を比較し

---

*Identification of /ba/-/wa/ in reverberant environments with synthesized continuum by changing formant transitions, by KAWAI, Tomoko (Sophia Univ.), OSAWA, Eri (Showa Univ.), MIZOGUCHI, Ai (Maebashi Institute of Technology), ARAI, Takayuki (Sophia Univ.).

て知覚に変化が生じたと報告されている[6]。

先行研究では，閉鎖音/b/と接近音/w/間での残響によるフォルマント遷移の持続時間の変化に伴う音声知覚への影響の調査はまだ行われていない。よって本研究では，フォルマント遷移の持続時間を連続的に変化させた合成音声刺激/ba/-/wa/を用いて，残響環境下においてフォルマント遷移の持続時間を変化させたことに伴い知覚にどのような影響が生じるかを調査した。

2　方法

2．1　実験参加者

実験参加者は 19 歳から 27 歳の健聴な 24 名の日本語母語話者であった。

2．2　刺激

フォルマント遷移の持続時間が 30 ms ～ 70 ms の時間長を 5 ms 刻みで変化させた 9 ステップの合成音声刺激を用いた。合成音声刺激はフォルマント合成ソフトウェア XKL[10]を用いて作成した。刺激音声はフォルマント遷移部＋定常母音/a/（持続時間 290 ms）で構成されている。フォルマント遷移は F1 と F2 を変化させることで表現した。各フォルマント周波数は阿部らを参考にした[3]。

F1 は開始点で 200 Hz から 800 Hz に上昇，F2 も開始点で 922 Hz から 1342 Hz に上昇させた。この上昇時間をフォルマント遷移部とした。その他のフォルマント周波数について，F3 は 2468 Hz，F4 は 3700 Hz，F5 は 4500 Hz であり，音声の開始点から終了点まで一定とした。Fig.1 は合成音声刺激のステップ 1 とステップ 7 におけるフォルマント周波数の変化を示している。

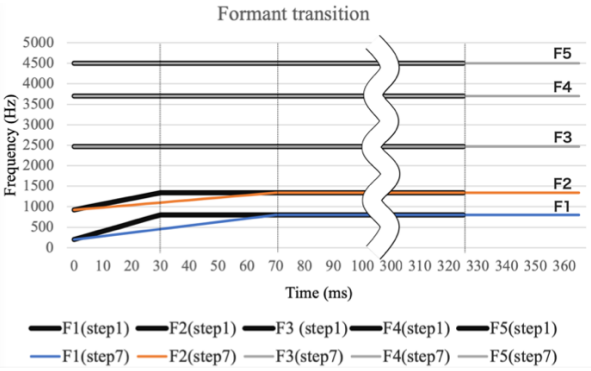

Fig. 1 F1-F5 Frequencies in stimuli

基本周波数(F0)は音声の開始点からフォルマント遷移部は 130 Hz で一定，定常母音開始から 180 ms 間かけて 110 ms に低下，その後さらに 110 ms 間かけて 70 Hz に低下させた。

振幅(AV)については，音声の開始点から 10 ms 間かけて 0 dB から 60 dB に上昇，その後 60 dB で一定とした。音声の終了点の 50 ms 前から終了点にかけて 50 ms かけて 0 dB に低下させた。

2．3　実験環境

実験は上智大学荒井研究室の防音室にて実施した。4 つのスピーカ(Genelec 8020A)を実験参加者が中心となるように配置した。残響環境条件では 4 つ全てのスピーカ，非残響条件では正面 2 つのみのスピーカから刺激音声を提示した。これらの配置は Fig. 2 に示す。

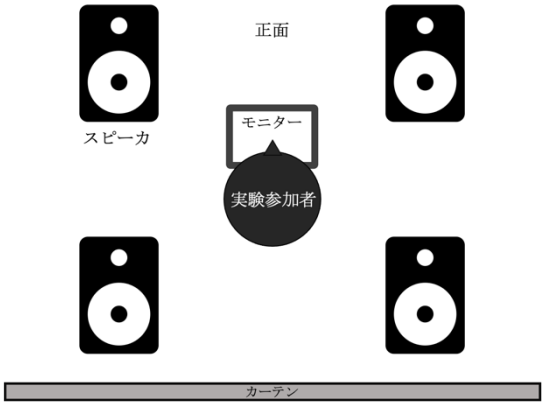

Fig. 2　Experimental setup

実験参加者刺激音声は全て等価騒音レベル(LAeq) 50 dB で提示した。残響環境条件ではリバーブレーター (Roland-RSS-303)によって残響を付加した。付加した残響の残響時間は周波数が 125 Hz，　500 Hz ，1 kHz，　2 kHz，4 kHz における平均値で 2.08 s であった。

PC モニターに表示した Praat[11]の GUI を介して実験参加者に回答を入力してもらうことで刺激に対する実験参加者の反応を求めた。

2．4　手順

実験参加者には教示により日本語音声の知覚実験であること，そしてスピーカから聞こえた音声に対して提示された 2 つの選択肢である「ば」と「わ」においてどちらに聞こえたかを回答する課題であることを伝えた。合成音声刺激はキャリア文を使用せず提示した。選択肢は PC 画面上に Praat の GUI によって 2 つのボックスが表示され，実験参加者はそれらのいずれかをマウスでクリックして回答した。

実験での回答における操作に慣れてもらう

ために練習課題を行った。一連の動作などを確認した後に実験へと移行した。

実験は最初に残響環境条件，次に非残響環境条件の順で実施した。各条件で試行回数は144 回(9 ステップ×選択肢の左右配置の入れ替えで 2 回×繰り返しで 8 回)となり，残響の有無による 2 条件で合計 288 試行であった。

実験の最後に年齢，性別，本人の出身地，養育者の出身地，音楽の経験についてのアンケート調査を行った。

3　分析

実験参加者による各刺激への応答に対し，一般化線形混合分析(GLMM)を用いて分析を行った[12]。モデリングでは R Studio[13]の *lme4* パッケージの *glmer* 関数を使用した。

実験参加者をランダム効果とし，各刺激と残響の有無を固定効果とした。従属変数は刺激に対する応答である（実験参加者による/ba/あるいは/wa/いずれかの回答）。その後，R Studio の *emmeans* パッケージによる Post-hoc test を行った。

4　結果と考察

Fig.3 は残響環境下と非残響環境下における，合成音声/ba/-/wa/の連続体に対するの/ba/応答の予測確率を示している。

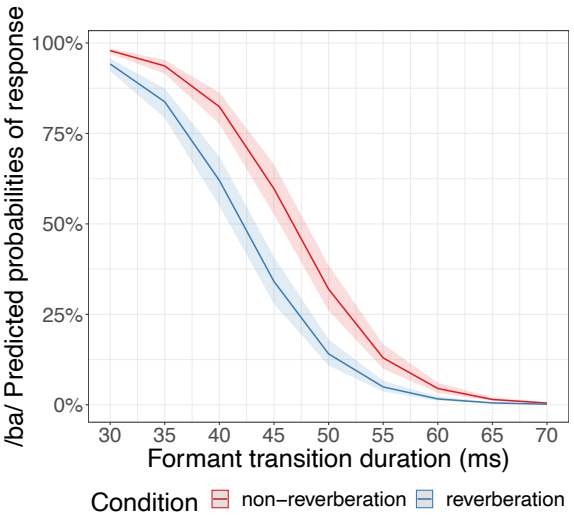

Fig.3　Predicted probability of /ba/ in the continuum of /ba/-/wa/

残響環境下と非残響環境下の 2 条件によるコンディションの違いには有意差がみられた($\chi 2(1) = 176.7, p < 0.01$)。コンディションと各刺激ステップとの間には有意差がみられなかった($\chi 2(1) = 1.7563, p < 0.1851$)。

実験結果から，残響の付加が閉鎖音/b/と接近音/w/間におけるフォルマント遷移の持続時間を変化させた合成音声刺激の知覚に影響を与えることがわかった。残響環境下では非残響環境下と比較して，より短い時間長の刺激ステップに/ba/と/wa/のカテゴリー境界があった。そして同じ時間長の刺激ステップに注目すると，非残響条件と比較して残響条件では/wa/の回答が増加した。この結果から，音素内の時間波形が崩れてエネルギー分布が不鮮明になることで，非残響環境下と比較してフォルマント遷移の持続時間が長くなったように知覚する傾向にあることが考えられる。

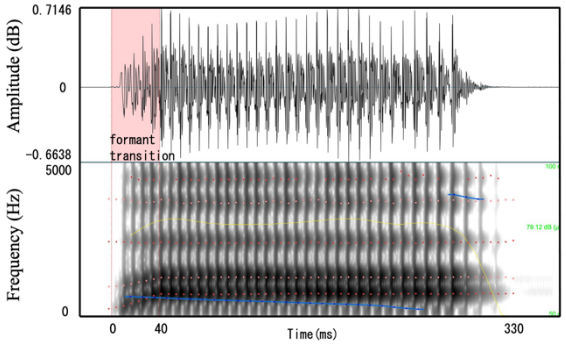

Fig. 4 Spectrogram of stimulus step 3 (formant transition of 40 ms) in non-reverberant environment

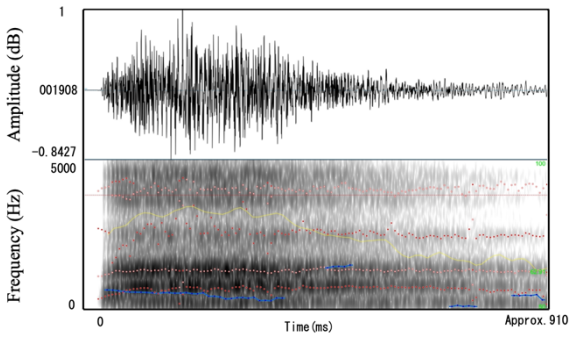

Fig. 5 Spectrogram of stimulus step 3 (formant transition of 40 ms) in reverberant environment

Fig. 4 と Fig. 5 残響環境下におけるカテゴリー境界付近の音声刺激であるステップ 3（フォルマント遷移の持続時間 40 ms）のスペクトログラムをそれぞれ示す（Fig. 4 での残響は本実験に使用した残響とは異なる）。Fig.4 の赤い範囲はフォルマント遷移部を表している。このように残響を付加すると，非残響下でのスペクトログラムと比較してフォルマントの軌跡が遷移部において不鮮明になっていることがわかる。

今回の実験で使用した残響は約 2.2 s と長い

RT を持っていたにも関わらず, 合成音声刺激/ba/と/wa/における残響環境下と非残響環境でのカテゴリー境界の変化は約 5 ms ほどであり，ヒトの音声知覚における頑健性を表している。つまり，self-masking などの残響による影響が生じることによってエネルギー分布が不鮮明になった音声刺激に対し，残響が無い状態でのフォルマント遷移の持続時間を推定する知覚補正がそれなりに働いていた可能性がある。そのような補正があった可能性はあるが，それでもなお，フォルマント遷移の持続時間が長くなったように知覚し，非残響環境下と比較して短いフォルマント遷移の持続時間で/wa/であると判断したのではないかと考えられる。

Nábělek らの先行研究では残響環境下における単母音/a/と二重母音/aɪ/の合成音声刺激の連続体に対する識別では，残響環境下では非残響環境下と比較してフォルマント遷移の持続時間がより長い時間長の刺激ステップにて二重母音であるという回答が増えた[6]。この結果は本研究と同様，残響環境下において残響が無い状態のフォルマント遷移を推定する知覚補正が十分でなかったことにより，残響環境下と非残響環境で音声知覚に変化が生じたと考えられる。

今回は単音節のオンセット内で生じる残響の影響を調査するためにキャリア文は使用しなかったが，先行する母音や子音がある場合は今回の実験結果の知覚傾向と異なる可能性がある。例えば，母音＋子音(フォルマント遷移部)＋母音の場合は, 先行する定常母音のエネルギーが後続する子音よりも大きいため[14]，今回の実験結果よりも残響が大きく影響を与える可能性がある。そのことから，残響環境下と非残響環境下で知覚が大きく変化することが予測できる。この点については今後検証したい。また，本研究における結果は合成音声による単音節に対する音声知覚であるため，自然音声などの異なる条件に対する音声知覚では，本研究とは知覚傾向が異なる可能性がある。そのような条件下での音声知覚も調査する必要がある。

5　結論

本研究では，閉鎖音/b/と接近音/w/において合成音声/ba/−/wa/の刺激連続体に対する識別課題から，フォルマント遷移の持続時間の変化に伴い残響環境下での知覚にどのような影響が生じるかを調査した。その結果，非残響環境下と比較して残響環境下ではフォルマント遷移の持続時間が短い時間長の刺激ステップで/wa/の回答が増えた。そのことから残響が合成音声刺激/ba/−/wa/の知覚に影響を与えることがわかった。

謝辞

本研究は上智大学重点領域研究の一部として助成を得た。

上智大学「人を対象とする研究」に関する倫理委員会の承認を受けた（承認番号 : 2021-53）。

参考文献

[1] ジャック･ライアルズ 著, 今富摂子, 荒井隆行, 菅原勉 監訳, *音声知覚の基礎,* 海文堂, 2003.

[2] A. M. Liberman e*t al.*, Journal of Experimental Psychology, 52(2), 127-137,1956.

[3] 阿部晶子他, 特殊教育学研究, 40(1), 13-23, 2002.

[4] A. K. Nábělek *et al.*, J. Acoust. Soc. Am., 86(4), 1259-1265, 1989.

[5] T.Arai *et al*., IEEE Trans. on Audio, Speech, and Language Processing, 18(7), 1775-1780, 2010.

[6] A. K. Nábělek *et al.*, J. Acoust. Soc. Am., 95(5), 2681-2693, 1994.

[7] E. Osawa *et al*., Acoust. Sci. & Tech., 39(6), 369-378, 2018.

[8] 荒井隆行他, 日本音響学会講演論文集, 733-734, 2021.

[9] T. Arai *et al.*, Acoust. Sci. & Tech., 39(3), 252-255, 2018.

[10] D. H. Klatt, J. Acoust. Soc. Am., 67(3), 971-995, 1980.

[11] P. Boersma *et al.*, *Praat, a system for doing phonetic by computer,* Glot International, 2019.

[12] E. Osawa *et al.*, Speech Communication, 134, 1–11, 2021.

[13] R Core Team., *R Foundation for Statistical Computing*, Vienna, Austria, 2019.

[14] T. Arai *et al.*, Acoust. Sci. & Tech., 23(4), 2002.