

雑音・残響が発話に与える影響*

©程島奈緒, 荒井隆行 (上智大), 栗栖清浩 (TOA)

1 はじめに

私達は周囲の音響環境に応じて発話を変化させる。その一例はロンバート効果^[1]であり、雑音下で発話されたロンバート音声は、静かな場所で発話された音声に比べて音響的特徴が変化し(例: 時間長・インテンシティ・F0・F1・F2の増加)^[2,3], 信号対雑音比が10~-5 dBのような雑音下で単語理解度が上昇する^[2,3]。音声・話者認識の一部においてもロンバート効果は応用されている^[4]。

その一方で、残響下の発話では音響的特徴がどのように変化するか、また残響下で明瞭であるかは明らかにされていない。雑音と残響では、音声生成・音声知覚共に与える影響が異なると考えられる。例えば、聴覚障害者と対話することを想定して明瞭に発話されたclear speech^[5]が挙げられる。静かな環境で録音されたclear speechと通常に発話された音声を比較すると、雑音下では通常に発話された音声よりclear speechの正解率の方が上昇したが^[6], 残響時間が長い環境においてはclear speechの効果は確認されていない^[7]。

その理由としては、音声知覚では雑音・残響が音声に与えるマスキングが、音声生成では発話者への聴覚フィードバックが異なる点である。マスキングについては、雑音下では同時マスキングが、残響下では先行する音声区間が後続の音声区間をマスクするoverlap-masking^[8]が発生し、そのパターンは時間・周波数的に異なる。

聴覚フィードバックについては、雑音下では発話中の音声と発話者に提示される雑音には相関がない。ロンバート音声では、提示される雑音の増加によって聴覚フィードバックが減少すると、発話が自発的及び受動的に変化することで明瞭度が上昇すると報告されている^[1]。一方残響下では、発話中の音声と発話者にフィードバックされる音声に相関がある。残響下での聴覚フィードバックには、雑

音下と同様に、提示される残響音の増加に伴って減少するもの(フィードバック1)と、聴覚遅延フィードバック(DAF)のように反射音の遅延によって生じ、残響音の増加に伴って増加するもの(フィードバック2)の2種類がみられる。DAF下の発話ではロンバート音声でみられた特徴^[1-3]のようにインテンシティやF0の増加や話速の低下^[9]がみられるが、発話の間違いや脱落によって明瞭度が低下すると報告されている^[9]。

本研究の最終的な目的は、雑音・残響下で拡声をする場合に、どのような発話が明瞭かを調査することである。この調査には音声生成・音声知覚の両面からの検討が必要であるが、本稿では音声生成に注目し、(1)残響下でどのような音響的特徴の変化が観測されるか、(2)その変化はロンバート音声で観測されたもの^[1-3]と類似しているか、(3)音響的特徴は残響時間によって変化するかを調査した。本稿で分析に用いる音響的特徴は、ロンバート音声^[2,3]で用いられているF0, F1, F2, 発話レベル, 子音と母音のインテンシティ比(CVR), 時間長とした。

2 録音

2.1 発話者

東京方言話者4名(男女各2名, 年齢22-37才)が録音に参加した。発話者へのインタビューから、発話者全員の聴覚に問題はなく、発話障害もないと判断された。

2.2 音声サンプル

原音声として、(1)「今から聞こえてくるのは」に続く親密度が3.1~3.4もしくは4.6~5.0の4モーラ語^[10]を10文、(2)音素バランス1000文^[11]と基本的に同じ5文、の2種類の音声を使用した。音響分析では(1)から各親密度群の5単語ずつ、(2)から3もしくは4モーラ語を5単語選び、計15単語をターゲット語として使用した。原音声(1)と(2)

* The effect of noise and reverberation on speech production, by HODOSHIMA Nao, ARAI Takayuki (Sophia University) and KURISU Kiyohiro (TOA).

では、ターゲットの親密度（1：統制あり，2：統制なし）と，キャリア文からのターゲットの予測度（1：予測度が低い不可，2：予測度が高い）が異なる。

発話条件は，静か（Q），雑音（N），残響（R）2種類の4条件とした。雑音は白色雑音を，R条件には2種類のインパルス応答（R1，R2）を使用した。インパルス応答のオクターブバンドの中心周波数 125-4000 Hz における EDT（初期減衰時間）の平均は R1 で 3.5 s，R2 で 12.3 s であった。なお直接音成分はインパルス応答から削除した。

2.3 手順

Fig. 1 に録音環境を示す。録音は防音室で行い，発話者の音声はマイク（SHURE, Beta-53），アンプ（PreSonus, DIGIMAX FS），サウンドカード（RME, Fireface 800）を介してコンピュータに録音した。雑音は，コンピュータから同サウンドカードを介してヘッドホン（SENNHEISER, HDA200）から提示した。残響音は，同マイクに入力された音声にインパルス応答を実時間で畳み込み，同ヘッドホンから提示した。雑音・残響音の付加は Adobe Audition 3.0 を使用した。雑音・残響音の音圧レベルは，ヘッドホン（SENNHEISER, HDA200）の出力を，人工耳（B&K, Type 4153）を介した騒音計（Ono Sokki, LA-5111）で測定し，発話者の耳元で平均 80 dBA になるようにした。発話レベルとして，発話者の口唇から水平に 20 cm 離れた地点でのターゲットの平均音圧レベルを騒音計（RION, NL-32）で測定した。

練習 8 試行（2 文×4 発話条件）の後に，60 試行（15 文×4 発話条件）を 1 セットとして録音を行った。休憩を挟んで計 2 セットの録音を行った。各試行では，発話者は 1.5 m 離れたモニタに表示された文を連続して 3 回繰り返して読み上げた。N と R 条件では，発話者はできるだけ明瞭に発話するよう指示された。録音は Q 条件の後に N・R 条件を行った。各発話条件内で 15 文を録音する順番は，発話者ごとにカウンタバランスをとった。また N・R 条件の順番も，発話者ごとにカウンタバランスをとった。

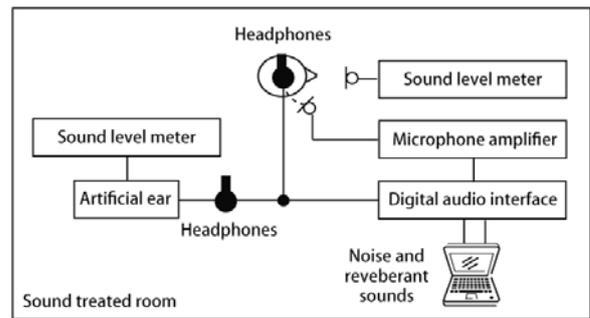


Fig. 1 Recording settings

3 音響分析

Fig. 2 に各発話者（T1-4）と発話者の平均（mean）におけるターゲットの $F_0 \cdot F_1 \cdot F_2 \cdot CVR \cdot$ 音圧レベル・長さ，文の長さを示す。なお，原音声（1）（2）のターゲットでは音響的特徴に似た傾向が観測されたため，本稿では両原音声のターゲットの平均値を示した。音響分析には，雑音等が入力されていない限りは 2 セット目の 3 回目の発話を使用した。Fig. 2 中のアスタリスクは，音圧レベル以外の各音響的特徴の平均値において，Q-N, Q-R1, Q-R2, R1-R2 条件間のうち話者・発話条件に対する繰り返しのある分散分析の結果が 1% 水準で有意なものを示す。

N 条件では，Q 条件と比較して， F_0 と F_1 の増加， CVR の減少，音圧レベルの増加が見られたが，これは先行研究^[1-3]と一致した。一方で， F_2 ，ターゲット・文長の増加は先行研究^[2,3]とは異なる結果となった。

R 条件では Q 条件と比較して次のような結果が得られた。まず N 条件と同様に F_0 と F_1 は予測通り増加した。これは，ロンバート音声や DAF 下で観測された特徴^[1-3,9]と同様にフィードバック 1 による発話の強調によって増加したためと考えられる。次に CVR は N 条件と同様に減少した。これは，ロンバート音声で観測された特徴^[1-3]と同様にフィードバック 1 により母音が強調されたためと考えられる。ただし，子音に対する母音のエネルギー比が増加すると残響による子音へのマスキング量がより増加するため，音声知覚実験を行った場合，子音の聴取は困難になると考えられる。ただし T3 のみ CVR が増加しており，以下で述べるように R 条件では発話の方法が発話者によって異なることが考えられる。さらに発話レベルは増加し，予測通り N 条件下

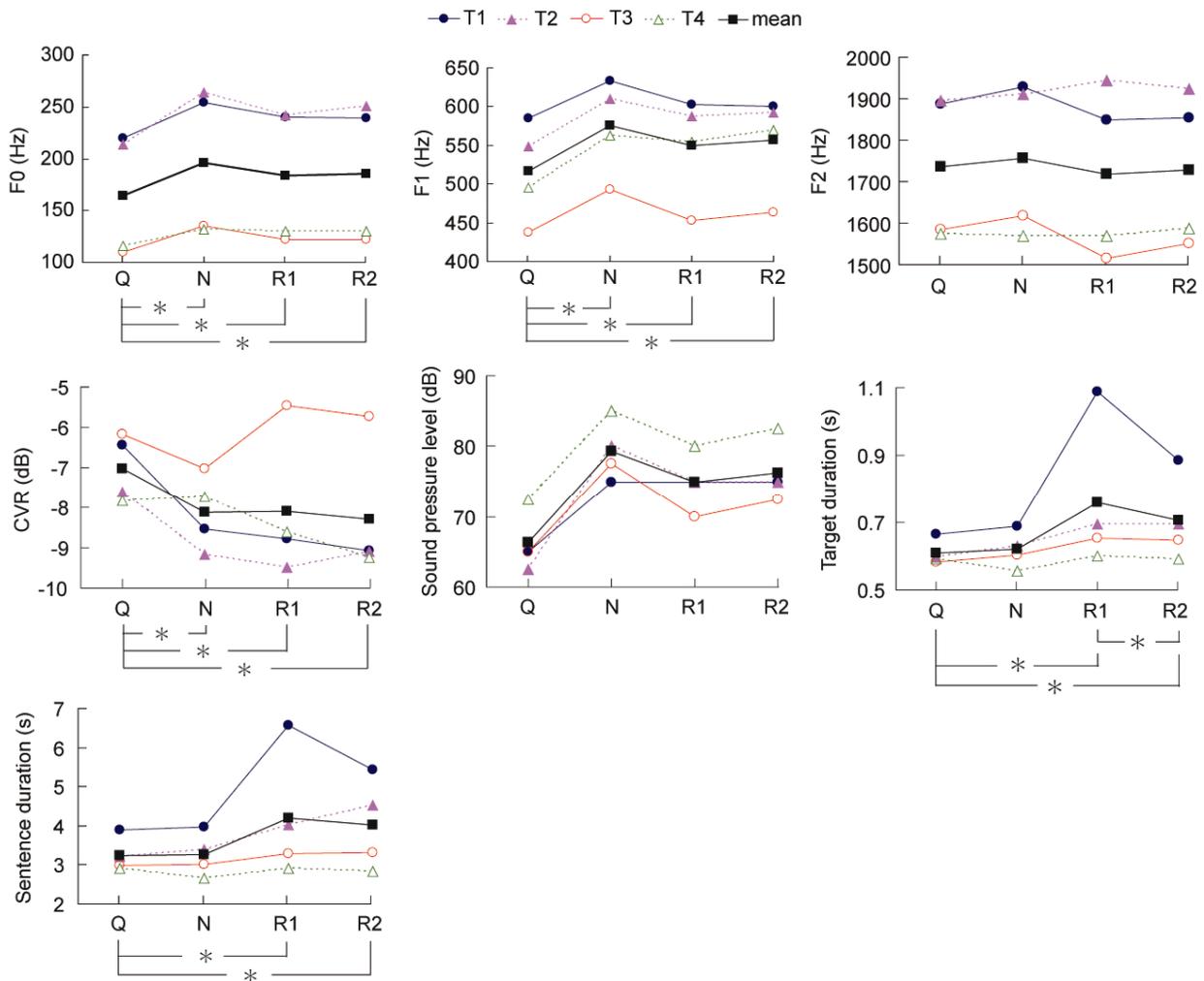


Fig. 2 Mean F0, F1, F2, CVR, sound pressure level, target duration and sentence duration in each talker and recording condition (* shows a significant difference at $p < 0.01$)

の増加量よりも少なくなった。これはまず聴覚フィードバック 1 と 2 から、音圧レベルが Q 条件下よりも N 条件下で増加する。しかし、音圧レベルが増加し過ぎると、**overlap-masking** が増加する。それを防ぐために、発話者の音圧レベルの増加が抑えられたと考えられる。そしてターゲット及び文の長さは予測通り長くなった。これは話速を低下することで **overlap-masking** を減少させ、その結果主にフィードバック 2 が大きく減少されたためだと考えられる。

今回調査した音響的特徴においては F0、発話レベル、CVR のみが N・R 条件共に Q 条件と比べて同様の変化を示した (F0 と発話レベルは増加、CVR は減少)。このことから、本稿で用いた R 条件と N 条件下での音声生成は、発話が強調されるという観点では同一の音響的特徴で評価できる場合と、マスキングパターンや聴覚フィードバックなどの違い等によ

り異なる音響的特徴を示す場合があることが示された。

ターゲット長以外は、R1・R2 間で Q 条件と比較して音響的特徴の変化に大きな差はみられなかった。従って、先行研究^[2,3]で提示雑音レベルが音響的特徴の変化に大きく関与しなかったのと同様に、残響量も大きく関与しない可能性が示唆された。

発話者間の音響的特徴の変動は、N 条件よりも R 条件の方が増加した。その理由としては、一定の白色雑音が提示される N 条件と比べて、R 条件では発話者に与えられる聴覚フィードバックに対して発話者が異なった方法で発話を行っていると考えられる。発話者へのインタビューから、ある発話者は R 条件での発話が困難であると回答し、ある発話者は聴覚フィードバックを減少させようと発話を調整したと回答した。すなわち後者の発話者は話速の低下、発話レベルの低下、子音を強

調した発話などにより発話者及び聴取者へ付加される overlap-masking を減少させたと考えられる。

さらに著者の聴覚印象では N・R 条件共に発話者全員に母音の鼻音化が確認され、鼻音化の度合いは R 条件の方が強くなった。聴覚障害者においては、聴覚フィードバックが阻害されることによる口蓋咽頭の異常のために鼻音化が発生する可能性があるとして述べられている^[12]。このことから、雑音・残響音の提示レベルが高く、かつ R 条件では反射音による遅延が存在する環境下において、発話中の鼻咽腔閉鎖のタイミングが通常の発話とは異なっていたことが考えられる。その結果聴覚障害を持っていない若年者の発話が鼻音化した可能性が考えられる。

4 おわりに

本稿は静かな環境 (Q)・雑音 (N)・2 種類の残響 (R) 下で発話された単語の音響的特徴を比較した。その結果、(1) R 条件で Q 条件に比べて音声生成に変化がみられた (F0, F1, 音圧レベル, 単語・文の長さの増加, CVR の減少), (2) F0, 音圧レベル, CVR のみ R・N 条件共に Q 条件と比べて同様の変化を示した (F0 と音圧レベルは増加, CVR は減少), (3) ターゲット長以外は, R 条件間で Q 条件と比較して音響的特徴の変化に差はみられなかった, (4) 発話者間の音響的特徴の変動は, N 条件よりも R 条件の方が増加した。以上より, 発話者は周囲の音響環境に応じて発話を自発的及び受動的に変化させ, そのうち聴覚フィードバックの条件に応じて, 音響的特徴が N 条件と R 条件で似る場合と異なる場合の両面が観測された。

本報告では音声生成側の検討を行ったが, 今後は観測された音響的特徴の変化が雑音・残響下の音声明瞭度にどう寄与するかを調査したい。さらに, 比較的高い音声明瞭度が求められる公共空間において, 利用者に効果的に情報を伝える音声も検討したい。

謝辞

本研究は文部科学省私立大学学術研究高度化推進事業上智大学オープンリサーチセンター「人間情報科学プロジェクト」の支援を受

けて行われた。インパルス応答を提供して下さった橘秀樹先生, 上野佳奈子先生, 横山栄先生, 録音に参加して下さった発話者に感謝致します。

参考文献

- [1] Lane and Tranel, J. Speech Hear. Res., 14, 677-709, 1971.
- [2] Summers *et al.*, J. Acoust. Soc. Am., 84, 917-928, 1988.
- [3] Junqua, J. Acoust. Soc. Am., 93, 510-524, 1993.
- [4] Hansen, Speech Commun., 20, 151-173, 1996.
- [5] Picheny *et al.*, J. Speech and Hear. Res., 28, 96-103, 1985.
- [6] Payton *et al.*, J. Acoust. Soc. Am., 95, 1581-1592, 1994.
- [7] Hodoshima *et al.*, Proc. Acoustics 08, 2393-2397, 2008.
- [8] Nabelek *et al.*, J. Acoust. Soc. Am., 86, 1259-1265, 1989.
- [9] Yates, Psychol. Bull., 60, 213-232, 1963.
- [10] 天野ら, “親密度別単語了解度試験用音声データセット 2003 (FW03)”, 音声資源コンソーシアム, 2006.
- [11] “音素バランス 1000 文”, NTT アドバンステクノロジー株式会社, 1999.
- [12] Fletcher and Daly, J. Commun. Disorders, 9, 63-73, 1976.