

雑音・残響下の発話による音声の明瞭度改善*

◎程島奈緒（東海大），荒井隆行（上智大），栗栖清浩（TOA）

1 はじめに

駅などの雑音や残響が存在する公共空間では，音声案内の聴取が難しい場合がある。この問題に対して，建築的・電気音響的な対応や，拡声システムの一部としての信号処理が提案されている^[1,2]。これらの対応では，公共空間ならびに電気音響設備の設計や改修が必要となる。

明瞭度は音声案内を発話する環境によっても変化する。これは，Lombard 効果^[3-5]のように人が周囲の雑音に応じて発話を変化させ，その結果として音声明瞭度が改善することが挙げられる。例えば雑音下での発話は，静かな場所での発話に比べて音響的特徴が変化する（例：時間長・インテンシティ・F0・F1・F2 の増加）^[3-5]。残響下でも同様の音響的特徴量の変化が観測されている^[6]。

音声知覚においても，信号対雑音比（SNR）が-5～15 dB 程度の雑音下で，雑音下の発話の方が単語理解度は上昇する^[3,4]。一方残響下では，残響下での発話が静かな環境での発話よりも明瞭であるかは分からない。これは雑音と残響では音声へのマスクングのパターンが異なるためである。つまり，雑音下では同時マスクングが，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^[7]が発生する。従って雑音下とは異なり，残響下では聴覚遅延フィードバック（DAF）のように発話と発話者が聴取する音声に相関がみられる。DAF 下の発話には Lombard 音声にみられる音響的特徴^[3-5]のようにインテンシティの増加等^[8]がみられるが，発話の間違い等によって明瞭度が低下する^[8]。よって，残響下の発話の明瞭性は，音響的特徴量や雑音下での知見^[3-6]と同様の結果になるかは分からない。

本研究の目的は，雑音や残響が存在する公共空間で明瞭な音声案内を提供することである。本稿では発話を行う際の環境に着目し，

Table 1. Speaking conditions

Speaking conditions	Sounds presented over headphones
quiet (Q)	no sounds
noise (N)	white noise
reverberation (R1)	reverberant speakers' utterance

音声案内を発話する環境が静かな環境よりも雑音や残響下の方が，同じ雑音や残響下で聴取した時に明瞭になるかを調査した。さらに，聴取場所とは異なる残響環境下で発話を行った場合においても，明瞭性は向上するかを調査した。また，残響下で聴取をする場合，キーワードとなる単語と先行する音声区間に付加された残響による overlap-masking がキーワードの単語理解度に与える影響を調査した。

2 録音

2.1 発話者

東京方言話者 2 名（男女各 1 名，平均年齢 21 才）が録音に参加した。発話者へのインタビューから，両発話者とも聴覚に問題はなく，発話に障害もないと判断された。

2.2 音声サンプル

発話する文章として，36 種のターゲット語（親密度が 2.5～4.0 の 4 モーラ語）が 1 つずつキャリア文に挿入された 36 文を使用した。発話条件は Table 1 に示す通り，静か（Q），雑音（N），残響（R1）の 3 条件とした。N は白色雑音を，R1 は教会で測定されたインパルス応答を使用した。残響時間（RT）は 125-4000 Hz の平均で 3.6 s であった。

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），

* Speech spoken in noise and reverberation improves word intelligibility, by HODOSHIMA Nao (Tokai University), ARAI Takayuki (Sophia University) and KURISU Kiyohiro (TOA).

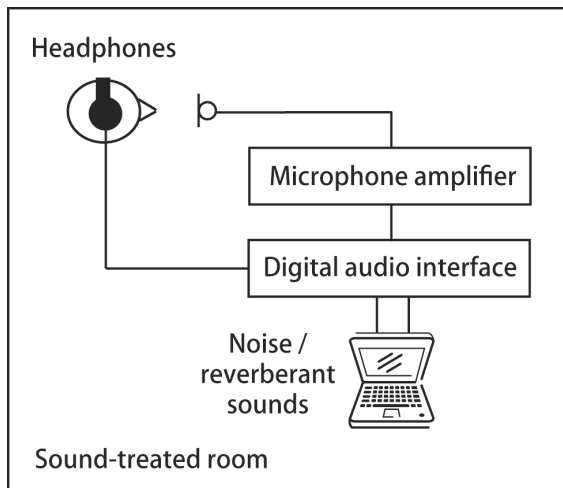


Fig. 1. Recording settings

オーディオインタフェース (RME, Fireface 800) を介してコンピュータに録音した。雑音は、同コンピュータから同オーディオインタフェースを介してヘッドホン (SENNHEISER, HDA200; dynamic, closed circumaural type) 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遅延時間が数 ms 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、発話者の耳元における提示音圧レベルは、発話者の口元の発話レベルに対して -22 dB とした。録音した音声には、雑音・残響のいずれも含まれていない。

各発話者は練習の後に 108 文 (36 文×3 発話条件) の録音を行った。録音に際して、発話者は約 1.0 m 離れたモニタに表示された文を 2 回繰り返して読み上げた。発話者には、発話した音声ヘッドホンから聞こえる雑音や響きを持つ公共空間内の聴取者に拡声されることを想像するよう教示を与え、できるだけ明瞭に発話するよう指示した。録音は 3 つの発話条件のうち Q 条件をまず行い、その後 N・R 条件を行った。各発話条件で 36 文を録音する順番は、発話者ごとにランダム化した。

3 聴取実験

3.1 参加者

参加者は、日本語を母語とする若年者 32 名 (男性 4 名, 女性 28 名, 平均年齢 23 才) であった。参加者の気導聴力レベルは 125-8 kHz において両耳で 25 dBHL 以下であった。

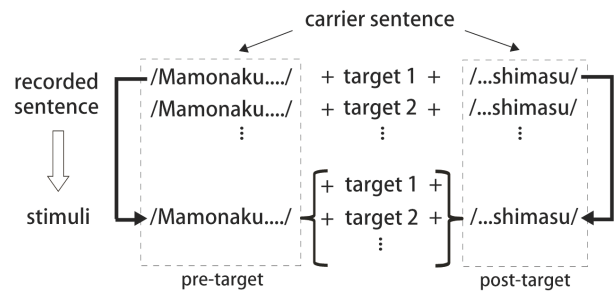


Fig. 2. Stimuli preparation

Table 2. Listening conditions

	Recording condition			
	Q	N	R1	QR1
N (SNR=-2 dB)	Q_N	N_N		
R1 (RT=3.6 s)	Q_R1		R1_R1	QR1_R1
R2 (RT=2.6 s)	Q_R2		R1_R2	QR1_R2

3.2 刺激

2.2 節の通り、録音された音声 (Fig. 2 上) はターゲット語がキャリア文に挿入されたものであり、キャリア文はターゲット語の前の区間であるプリターゲットとターゲット語の後の区間であるポストターゲットに分けることができる。録音において発話者は同じキャリア文を発話しているが、各発話は音響的に異なる。残響下では overlap-masking によってプリターゲットがターゲット語の了解度に影響を与える。その統制をとるため、各発話条件におけるターゲット語に対して、プリターゲットとポストターゲットの対を 1 つ選び、36 個のターゲット語を同じ対の間に挿入した (Fig. 2 下)。キャリア文とターゲット語の組み合わせは、Q (キャリア文) と Q (ターゲット語)、N と N、R1 と R1 の他 Q と R1 (QR1) とし、その組み合わせは同一発話者の発話を使用した。また、キャリア文とターゲット語のインテンシティ比やキャリア文とターゲット語間の無音区間も統一した。

聴取実験では、上記の連結した音声に白色雑音の付加もしくはインパルス応答の畳み込みを行った。Table 2 に聴取条件を示す。SNR は -2 dB とした。聴取実験で使用したインパルス応答は、録音で使用した R1 と、R1 に指数

包絡をかけてRTを2.6 sに短くしたもの(R2)である。刺激文のインテンシティは聴取条件・発話者内で正規化した。

3.3 手順

聴取実験は参加者毎に無響室内で行った。刺激はコンピュータにオーディオインターフェース (Onkyo, MA-500U) を介したヘッドホン (STAX, SR-303; electrostatic, open circumaural type) から diotic で提示した。刺激の提示レベルは参加者の聞きやすいレベルに調整した。各試行では刺激を一度提示し、参加者は聞こえたターゲット語を回答用紙に記入した。各参加者に対して、練習を2回行った後、計32刺激(8聴取条件×4単語)をランダムな順で提示した。全34刺激のターゲット語には、各参加者で全て異なる単語を使用した。ターゲット語と聴取条件の組み合わせは、参加者間でカウンタバランスをとった。

4 結果と考察

Fig. 3に各発話者(S1, S2)及び発話者の平均(average)における、各聴取条件のモーラ正解率を示す。発話者の主効果は有意ではなく、両発話者間で正解率に差がないことが示された。聴取条件の主効果は有意であり($p < 0.01$)、雑音・残響別の分析結果を以下に示す。

雑音条件では、N_NはQ_Nよりも正解率が有意に高くなった($p = 0.003$)。この結果はLombard効果に関する先行研究[3,4]と一致するものであり、雑音下での発話は静かな場所での発話よりも明瞭度が高くなることが示された。

残響条件では、R1_R1はQ_R1($p = 0.019$)とQR1_R1($p = 0.001$)よりも正解率が有意に高くなった。また、R1_R2はQ_R2($p = 0.027$)とQR1_R2($p = 0.018$)よりも正解率が有意に高くなった。これらの結果から、雑音と残響ではマスキングのメカニズムは異なるが、雑音下に加え、残響下の発話も静かな場所での発話よりも明瞭になることが示された。

また、発話と聴取が同じ残響環境(Q_R1に比べてR1_R1)のみならず、異なる場合(Q_R2に比べてR1_R2)においても、残響下の発話は静かな場所での発話よりも明瞭に

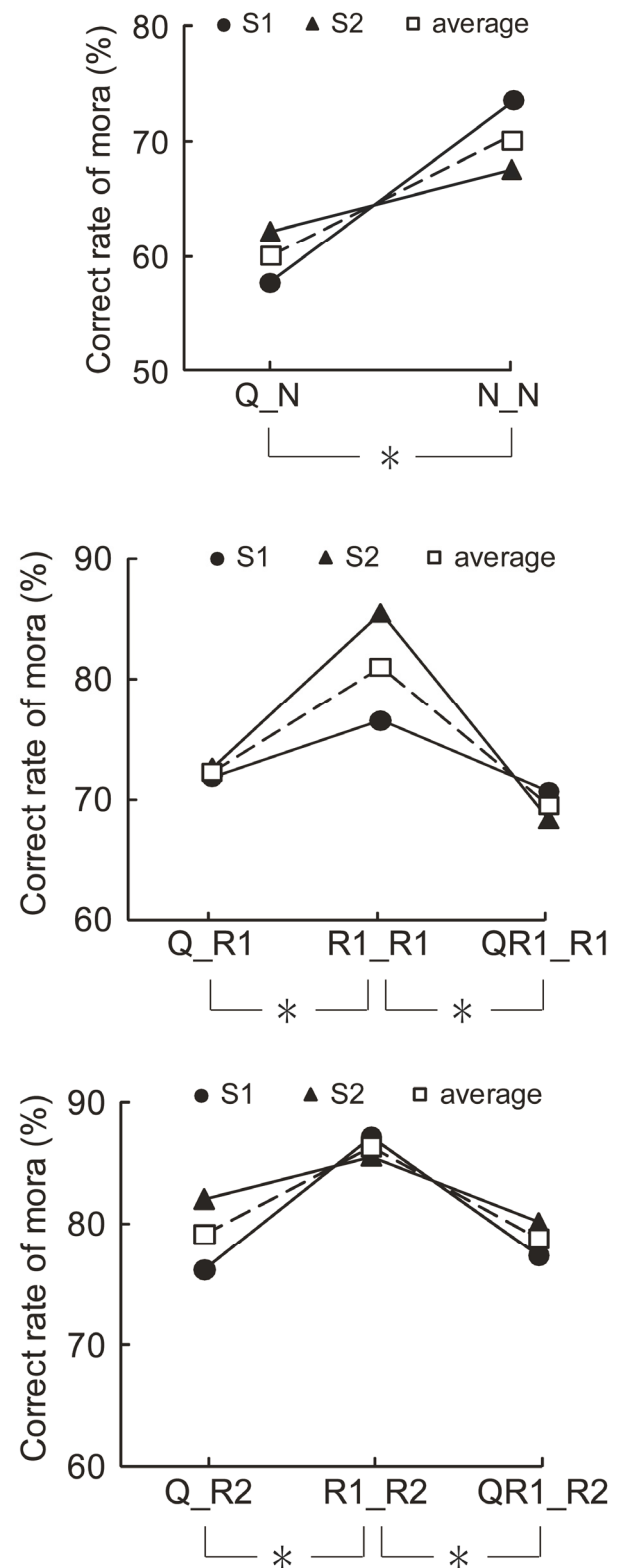


Fig. 3. Mean percent correct of mora in noise (top), reverberation R1 (middle) and reverberation R2 (bottom) for each speaker (circles and triangles) as well as for the average of speakers (squares). An asterisk shows a significant difference at $p < 0.01$.

なることが示された。この結果から、残響音を聞きながら発話することにより、音声案内をより明瞭にすることができる可能性が示された。さらに、その際に使用する残響条件は、音声案内を拡声する公共空間と必ずしも同じである必要はないことも示された。

Q_R1 と QR1_R1 及び Q_R2 と QR1_R2 間の正解率に有意差は認められなかった。このことから、ターゲット語が Q から R1 もしくは R2 に変化しても、ターゲット語の了解度は有意に改善しないことが示された。一方、R1_R1 と R1_R2 は QR1_R1 と QR1_R2 よりも各々正解率が有意に高くなった。よって、プリターゲットが Q から R1 もしくは R2 に変化すると、ターゲット語の了解度は有意に改善することが示された。これは、静かな環境での発話と比べて残響下の発話の音響的特徴が変化する一例に、無音区間が長くなること、基本周波数や第一フォルマントの増加、子音対母音のインテンシティ比の減少^[6]などが関連していると考えられる。以上より、残響下での音声案内を作成する際には、ターゲット語に加えて、プリターゲットも考慮する必要があることが示された。

5 結論

本稿は静かな環境、雑音、残響下での発話を録音し、雑音や残響を付加した録音音声を用いて若年者に対する聴取実験を行った。その結果、Lombard 効果^[3-5]で報告されている雑音下での発話だけではなく、残響下での発話も、雑音や残響下で聴取した場合に静かな環境での発話より明瞭になることが示された。また、この残響下での発話による明瞭度改善は、音声案内を作成する環境と拡声する残響環境が同じである場合でも、異なる場合でも観測された。さらに、残響下で明瞭な音声を作成するためには、発話者は主要な単語だけではなく、先行する語句にも注意を払う必要

があることが示唆された。

今後の検討として、今回の聴取実験で使用した雑音・残響下での発話の音響分析を行い、音声明瞭度との対応を調査したい。また、プリターゲットのうちどの音響的特徴（無音区間、インテンシティ、時間長、子音対母音のインテンシティ比など）が雑音・残響下の明瞭度改善に対応しているのかも調査したい。

謝辞

本研究は文部科学省科学研究費補助金（若手研究 B, 21700203）、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。実験デザインと統計に関してご助言を頂いた道又爾先生、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いました発話者に感謝します。

参考文献

- [1] Arai *et al.*, *Acoust. Sci. Tech.*, 23(4), 229-232, 2002.
- [2] Hodoshima *et al.*, *J. Acoust. Soc. Am.*, 119(6), 4055-4064, 2006.
- [3] Lane and Tranel, *J. Speech Hear. Res.*, 14, 677-709, 1971.
- [4] Summers *et al.*, *J. Acoust. Soc. Am.*, 84, 917-928, 1988.
- [5] Junqua, *J. Acoust. Soc. Am.*, 93, 510-524, 1993.
- [6] 程島他, 電子情報通信学会技術報告, SP2009-69, 43-48, 2009.
- [7] Nabelek *et al.*, *J. Acoust. Soc. Am.*, 86, 1259-1265, 1989.
- [8] Yates, *Psychol. Bull.*, 60, 213-232, 1963.
- [9] 天野他, “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.