

時間・周波数特性が異なる残響下での発話音声の明瞭度に与える影響

神林 千尋[†] 程島 奈緒[‡]

[†] [‡] 東海大学情報通信学部 〒108-8619 東京都港区高輪 2-3-23

E-mail: [†] 8bjnm008@mail.u-tokai.ac.jp, [‡] hodoshima@tokai-u.jp

あらまし 人は周囲の雑音や残響環境に応じて無意識的に発話を変化させることがある。具体的に雑音下の発話音声は静音下の発話音声よりも音声明瞭度が高くなることが分かっており、この現象を Lombard 効果という。そして残響下の発話においてもこの効果が同様に得られ、残響下の発話は静音下の発話と比較し音声明瞭度が高くなることが分かっている。本研究はこの残響下での発話特性を応用し、時間・周波数特性の異なる残響下での発話が明瞭度に与える影響を調査した。静音下(Q), 残響下[R:インパルス応答の D 値が 48%、RS: D 値が 52%、RA: D 値が 22%、残響時間は全て 2.2 s]の合計 4 条件で 4 名の話者が発話した音声を 16 名の若齢者に、残響下で聴取させる実験を行った。その結果 R・RA 条件の方が Q 条件よりもモーラ正解率が高くなった。この結果から、残響時間が先行研究より短く、公共空間(駅や空港)の環境に近い残響下においても音声明瞭度の高い発話が生成されることが確認された。そして D 値が低い環境での発話は話者への直接音のフィードバックが、D 値の高い環境での発話よりも少ないと考えられるため、話者が明瞭な発話をしようとする事が予測される。今後は明瞭度の高くなった音声の特徴を調べるために発話レベルやフォルマント周波数、母音対子音のインテンシティ比などの音響分析を行いたい。

キーワード 残響, Lombard 効果, 周波数特性, 音声明瞭度

Improving intelligibility of speech spoken under reverberant environment conditions: Effect of time and frequency characteristics of reverberation on speech intelligibility

Chihiro KAMBAYASHI[†] and Nao HODOSHIMA[‡]

[†] [‡] Faculty of Information and Telecommunication Engineering, Tokai University 2-3-23 Takanawa, Minato-ku, Tokyo, 108-8619 Japan

E-mail: [†] 8bjnm008@mail.u-tokai.ac.jp, [‡] hodoshima@tokai-u.jp

Abstract Humans instinctively adapt their speech to overcome their surrounding noisy or reverberant environment. Previous studies reported speech spoken in noise environments was more intelligible than speech spoken in a quiet environment. This phenomenon, which is commonly referred to as the Lombard effect, has been confirmed in speech spoken in reverberant environments. Previous studies reported that speech spoken in reverberant environments is more intelligible than speech spoken in a quiet environment. This study investigated whether speech spoken in reverberant environments with different time and frequency characteristics were intelligible compared with speech spoken in a quiet environment. Four speakers recorded speech in a quiet environment (Q) and in reverberant environments [Deutlichkeit is 48 % for R, 52% for RS and 22% for RA respectively. The reverberation times of these impulse responses are 2.2 s.]. Sixteen native speakers of Japanese participated in listening test in reverberant environments. The results showed that R and RA conditions were more intelligible than Q condition, especially under reverberant environments where reverberation times were close to those in public spaces (The previous studies used longer reverberation times.). The results also suggest that the reverberant environments with lower Deutlichkeit has less feedback of direct sound to the speakers than the reverberant environments with higher Deutlichkeit, so the speakers may attempt to speak more intelligible. Future study will carry out acoustic analysis (e.g., speech level, formant frequencies and consonant-vowel intensity ratio) of speech in order to examine which acoustic characteristics correlate with higher mora correct rate.

Keywords Reverberation, Lombard effect, Frequency characteristic and Speech intelligibility

表 1 発話条件

発話条件	使用したインパルス応答	D 値[%]
Q (静音下)		
R (残響下)		48
RS (残響下)	R の 1k Hz~2k Hz の帯域の振幅を 1.5 倍にしたもの	52
RA (残響下)	R の直接音から 50 ms 以内をカットしたもの	22

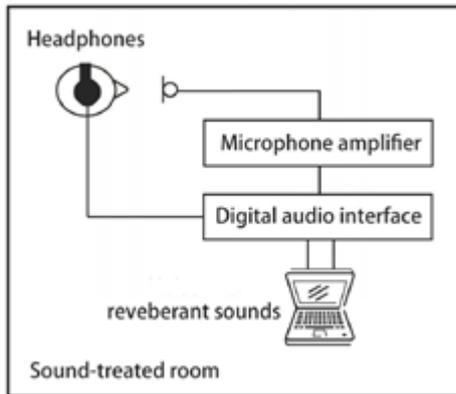


図 1 録音環境[3]

1. はじめに

現在、室内公共空間では様々な音声案内が存在する。そして緊急時にはこの音声案内が重要な存在になり、情報を正確に伝えるための手段となっている。しかし混雑時の放送は室内公共空間の残響によって、情報の伝達がうまくいかないことがあり、その原因として残響による overlap-masking や self-masking が挙げられている。そして先行研究[1]から残響下での情報伝達に影響を与えているのは残響によって後続の音声区間をマスクする overlap-masking と考えられている。この overlap-masking による音声明瞭度の低下を軽減させる方法のひとつとして、Lombard 効果を応用した研究[2]が行われている。Lombard 効果というのは人の発話が雑音下と静音下で音響的特徴が異なる現象である[3]。具体的には発話レベルや基本周波数などの上昇が確認されている[3]。またこの雑音下での発話は静音下での発話と比較し、音声明瞭度が高くなることが分かっている[4]。

さらに先行研究[5]では残響時間の異なる 2 種類の残響下(残響時間 R1: 3.5 s, R2: 12.3 s)での発話を使用して音響的特徴を調査した結果、Lombard 効果の特徴である基本周波数と発話レベル、第一フォルマント、インテンシティの増加が残響下での発話にも確認された。そして先行研究[2]では、静音下と雑音下(白色雑音)、残響下(残響時間: 3.6 s)の 3 条件の発話を使用し聴取実験を行った。その結果、雑音下での発話は静音

下での発話と比較し有意に正答率が高くなり、約 10% 正解率が向上した。そして残響下での発話の聴取は静音下での発話と比較し有意に高くなり、約 10% 正解率が向上した。そこで先行研究[6]では残響時間 3.6 s で残響のフィードバック量が異なる 2 種類の残響下(R1: 残響のフィードバック量が -22dB, R2: -10dB)の発話を使用して聴取実験と音響分析を行った。その結果、R1・R2 の残響下の発話音声の方が静音下の発話音声より有意に正答率が高くなり、それぞれ 6% と 10% 正答率が向上した。音響分析の結果、聴取実験で他の話者よりも明瞭度が向上した話者の音声は、静音下の音声よりも時間長と基本周波数が上昇する傾向が確認された[6]。

しかし先行研究[2, 6]で残響環境の録音時に使用されたインパルス応答の周波数特性や残響時間は、公共空間(駅や空港)の環境とは異なるものである。先行研究[7, 8]では公共空間における音環境の実態を把握するために、首都圏の鉄道駅や空港ロビーなどの音環境の実測調査を行った。その結果、先行研究で得られたいくつかのデータでは駅の残響の周波数特性は、250 Hz~2k Hz の周波数帯域のうち 1k Hz~2k Hz の残響時間が他の帯域よりも長く[7]、空港は 500Hz~1kHz の残響時間が他の帯域よりも長くなった[8]。しかしこのような環境下での発話がどのように明瞭度に影響を及ぼすかは明らかになっていない。

本研究ではこのような公共空間で明瞭な音声案内を作成するために、駅や空港に近い時間と周波数特性の残響下での発話がどのように音声明瞭度に影響するかを調査すべきと考えた。具体的には静音下と周波数帯域ごとに D 値と残響時間が異なるインパルス応答を 3 種類使用した残響下の合計 4 条件で録音を行い、録音した発話音声を刺激として聴取実験を行った。また本研究の最終目標は駅や空港などの雑音残響下で人々に明瞭な音声案内を提供することである。

2. 聴取実験

2.1. 原音声

原音声は先行研究[6]から、「まもなく 3 番線に新宿行きの (ターゲット語) が参ります」というキャリア

文に4モーラのターゲット語を挿入したものを使用した。親密度は2.5~4.0[9]のものとし、使用した単語数は48個である。

2.2. 録音

発話者は日本語母語話者4名(男性2名、女性2名、年齢22~24歳)とした。録音条件は表1の4条件である。RS、RAは実際の公共空間に近い環境(駅や空港)の発話を録音するために使用した。発話条件Qは静音下、その他の3条件(R、RS、RA)は残響下で録音を行った。

Rのインパルス応答はコンサートホールで録音したものを使用した。そしてRSのインパルス応答はRを中心周波数125 Hz、250 Hz、500 Hz、1k Hz、2k Hz、4k HzのオクターブバンドにMatlabでピーキングフィルタを施し、Rの1k Hz~2k Hzの帯域の振幅を1.5倍にしたRAを使用した。インパルス応答は、Rのインパルス応答の直接音から50 ms以内をカットしたものである。残響下3条件の残響時間の平均は約2.2 sであり、これらのインパルス応答は500 Hz~1k Hzの残響時間が他の帯域よりも長くなったが、RSだけ250 Hzの帯域も残響時間が長くなった。そしてそれぞれのインパルス応答のD値はRが48%、RSが52%、RAが22%である。この残響時間と周波数特性は、先行研究[7, 8]の駅と空港の音響特性のデータを参考にして決定した。

録音は防音室で行った。そして残響下の録音の際、残響音の付加はAdobe Auditionを使用し、マイク(SHURE KSM141)から入力された音声に録音に使用したインパルス応答を実時間で畳み込んだものをヘッドフォン(SENNHEISER HAD 200)を介して発話者に提示した(図1)。発話者には実験者が提示した文章のターゲット語の部分を0型アクセントで発話させた。録音の順番は発話条件Qを最初に行いその後、発話条件R、RS、RAの順で行った。

2.3. 刺激

刺激の作成手順として最初に、録音した原音声のキャリア文の前半(ターゲット語に先行する部分)からターゲット語へのoverlap-maskingの量を一定にするため、録音した音声を音声解析ソフトPraatでキャリア文とターゲット語に分離した。その後、切り取ったキャリア文の中から各話者の発話条件ごとにキャリア文をひとつ選び、数値解析ソフトMatlabを使用して、選んだキャリア文の()の部分にターゲット語を挿入した。そしてその音声の振幅を正規化し、この4種類の音声に対してRAで使用したインパルス応答を畳み込んだ。作成した刺激は計768(発話4条件×4話者×48単語)である。

2.4. 実験手順

実験参加者は、日本語を母語とする20代の16名(男

性6名、女性10名、年齢21~23歳)であった。実験は防音室で行った。実験参加者には防音室内に設置されているパソコン上のインターフェイスを操作させ、オーディオインターフェイス(TASCAM US-144MK2)を通じてヘッドフォン(SENNHEISER HAD 200)から刺激を提示した。なおパソコン上のインターフェイスは音声分析ソフトPraatで作成した。また参加者に提示する話者は、各実験参加者につき2名し、計52刺激提示した。その際、提示する話者ならびに使用する単語、条件が重複しないようカウンターバランスを作成した。

実験参加者には最初に練習として4刺激提示し、その後本番として計48刺激(発話条件4条件×6刺激×2話者)をランダムな順番で提示した。実験参加者には聞こえたターゲット語を、回答用紙に平仮名4文字で記入させた。

3. 結果・考察

図2は発話条件ごとの話者別モーラ正解率の平均である。FMは女性話者を、Mは男性話者を表している。統計解析ソフトSPSSを使用し、発話条件間に対応のない因子(発話者)と対応のある因子(発話条件)の2元配置分散分析を行った。結果、発話条件による主効果が確認された($p < 0.05$)。その後、発話条件ごとでSidakによる多重比較を行ったところ、条件Rは条件Qと比較しモーラ正解率が有意に高くなった($p = 0.012$)。さらに条件RAは条件Qと比較し、モーラ正解率が有意に高くなるという結果になった($p = 0.007$)。

そして話者ごとで分散分析を行ったところ、FM2($p = 0.010$)とM2($p = 0.014$)の発話条件による主効果が得られた。その後、発話条件ごとでSidakによる多重比較を行った結果、FM2の条件RAは条件Qと比較しモーラ正解率が有意に高くなった($p = 0.015$)。そしてM2の条件Rは条件Qと比較し、モーラ正解率が有意に高くなった($p = 0.022$)。

実験結果で話者の正解率を平均化したとき、条件Rが条件Qと比較してモーラ正解率が有意に向上したことから、先行研究[2, 6]より短い残響時間の残響下においても明瞭度の高い発話が生成されることが示唆された。しかし、先行研究[6]では単語の正解率が高くなった発話者は他の発話者と比較し、静音下の音声よりも基本周波数と単語の時間長が長くなるという結果が得られているが、FM2とM2のうち今回その結果が得られたのはM2のみであった。M2は条件Rの方が条件Qと比較し、ターゲット語の基本周波数は8 Hz、ターゲット語の時間長は0.04 s上昇した。しかし発話条件で有意差が得られたもう1名の話者FM2は、条件RAのターゲット語の時間長が条件Qと比較し、0.067 s長くなったが、ターゲット語の基本周波数は条件RAの方

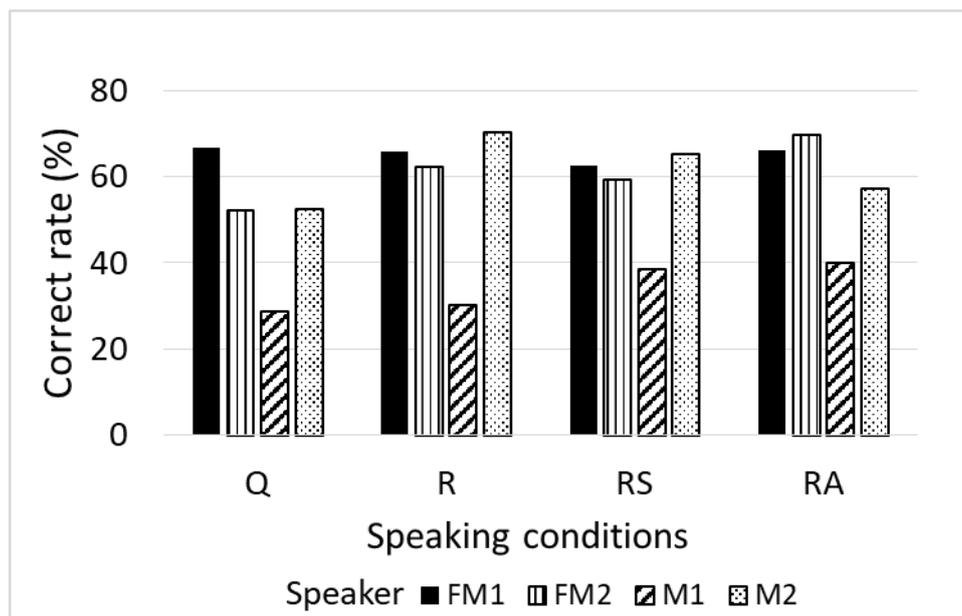


図 2. 発話条件ごとの話者別モーラ正解率

が 14 Hz 減少するという結果になった。

さらに今回の実験で条件 RA が条件 Q と比較し、モーラ正解率が向上したことから、残響下での発話において明瞭度が向上する要因の中に、使用したインパルス応答の D 値が影響していると考えられる。今回の実験で使用したインパルス応答の D 値は低い順に条件 RA, R, RA である。D 値の低い残響下での発話は、話者へのフィードバックが少ないため、話者が明瞭な発話をしようとするのが予測される。そして条件 RA は条件 R の直接音から 50 ms 以内をカットしたインパルス応答を使用したため、残響下 3 条件の中で最も D 値が低くなり、その結果条件 Q よりもモーラ正解率が有意に向上したと考えられる。

4. おわりに

本実験では Lombard 効果を応用した周波数特性の異なる残響下での発話が音声明瞭度に影響するかを聴取実験にて調査した。静音下で発話された音声(Q)と残響下[R: コンサートホールで録音されたインパルス応答, RS: R の 1k Hz~2k Hz の帯域の振幅を 1.5 倍にしたもの, RA: R の直接音から 50 ms 以内をカットしたもの, 残響時間は全て 2.2 s]の合計 4 条件で発話された 4 名の話者の音声を、若齢者に残響下(残響時間: 2.2 s)で聴取させた。

聴取実験の結果、発話条件による主効果が確認され発話条件 R・RA の発話音声は静音下での発話音声よりもモーラ正解率が有意に高くなった。ここから公共空間(駅や空港)に近く先行研究[2, 6]より短い残響時間の残響下においても明瞭度の高い発話が生成されることが確認された。そして今回の実験で発話条件 R・RA

が静音下での発話音声よりも明瞭度が向上したことから、残響下での発話において明瞭度が向上する要因のひとつに D 値が影響するのではないかと考えられる。D 値の低い残響下での発話は、話者への直接音のフィードバックが D 値の高い環境での発話よりも少ないため、話者が明瞭な発話をしようとするのが予測される。その結果、残響下 3 条件の中で D 値が最も低かった条件 RA(22%)と次に値の低かった条件 R(48%)での発話音声は静音下での発話と比較しモーラ正解率が有意に向上したと考えられる。本研究では駅や空港などの公共空間に近い環境下での発話が音声明瞭度にどのように影響するかを調査したが、今後はこのような環境で明瞭度の高くなった発話者の音声の特徴を調査するために音響分析を行いたい。

謝辞

本研究は JSPS 科研費 (18K04666) の助成を受けた。インパルス応答を提供して下さった橋秀樹先生、上野佳奈子先生、横山栄先生にも感謝いたします。

文献

- [1] Anna K Nabelek, Tomasz R. Letowski and Frances M. Tucker, "Reverberant overlap- and self-masking in consonant identification", J. Acoust. Soc. Am., 86, 1259-1265, 1989.
- [2] Nao Hodoshima, Takayuki Arai and Kiyohiro Kurisu, "Intelligibility of speech spoken in noise and reverberation", Proc. International Congress on Acoustics, 2010.
- [3] Harlan Lane and Bernard Tranel, "The Lombard sign and the role of hearing in speech", J. Speech Hear. Res., 14, 677-709, 1971.
- [4] W. Van Summers, David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow and Michael A. Stokes, "Effects of noise on speech production: Acoustic and

- perceptual analysis", *J. Acoust. Soc. Am.*, 84(3), 917-928, 1988.
- [5] 程島奈緒, 荒井隆行, 栗栖清浩, "雑音・残響下における発話の音響的特徴の話者変動", *電子情報通信学会技術研究報告*, 109(308), 43-48, 2009.
- [6] Nao Hodoshima, Takayuki Arai and Kiyohiro Kurisu, "Lombard-like effect against reverberation improves speech intelligibility in reverberation", *Proc. WESPAC.*, 93-97, 2015.
- [7] 伊積康彦, 藤井光治郎, 岩瀬昭雄, "駅コンコースの音環境に関する実態調査と主観評価実験", *日本建築学会環境系論文集*, 76(660), 115-124, 2011.
- [8] 横山栄, 向井ひかり, 橘秀樹, "公共空間の音環境に関する実測調査例", *日本騒音制御工学会*, 23(4), 228-231, 1999.
- [9] 天野成昭, 近藤公久, 坂本修一, 鈴木陽一, "親密度別単語了解度試験用音声データセット 2003 (FW03)", *音声試験コンソーシアム*, 2006.