社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
SP2009-69 (2009-11)

雑音・残響下における発話の音響的特徴の話者変動

程島 奈緒† 荒井 隆行† 栗栖 清浩‡

†上智大学理工学部 〒102-8554 東京都千代田区紀尾井町 7-1
‡TOA 株式会社 〒665-0043 宝塚市高松町 2-1
E-mail: †{n-hodosh, arai}@sophia.ac.jp, ‡kurisu_kiyohiro@toa.co.jp

あらまし 私達は周囲の音響環境に応じて発話が変化する。本報告では、静か（Q）、雑音（N）、2 種類の残響（R1, R2）環境下で 4 名が発話した文章の音響分析を行った。その結果、Q に対する N・R 下の音響的特徴には、発話の強調という面で同様の変化を示すもの（F0，F1，発話レベル，子音と母音のインテンシティ比）と，マスキングの違いにより異なる変化を示すもの（単語の長さ）が確認された。発話者間の変動は N 条件よりも R 条件で増加し、特に残響下では発話者の経験や意識によっても発話が変化する可能性が示された。また N・R 条件で母音が鼻音化し，その度合いは R 条件の方が強くなった。

キーワード 雑音，残響，音声生成，マスキング

Speaker variabilities of speech in noise and reverberation

Nao HODOSHIMA† Takayuki ARAI† and Kiyohiro KURISU‡

†Faculty of Science and Engineering, Sophia University 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan
‡TOA Corporation 2-1 Takamatsu-cho, Takarazuka-shi, Kobe, 665-0043 Japan
E-mail: †{n-hodosh, arai}@sophia.ac.jp, ‡kurisu_kiyohiro@toa.co.jp

Abstract Ambient noise and reverberation change our speech production. This paper is a report of the acoustical analyses of sentences produced by 4 speakers in quiet (Q), noise (N) and 2 reverberation (R1, R2) environments. Results showed that acoustical characteristics in N and R relative to Q changed similarly due to increase in the speaker's vocal effort (F0, F1, speech level, and consonant-vowel intensity ratio), as well as differently due to different masking patterns in N and R (word duration). Speaker variabilities were more dominant in R than in N, indicating that speech production also varies by the speaker's experience and intension. Vowels were nasalized in both N and R, and the degree of nasalization was more prominent in R than in N.

Keyword Noise, Reverberation, Speech production, Masking

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、フォルマント周波数の増加） [2,3]，信号対雑音比が 10～-5 dB 程度の雑音下で単語了解度が上昇する[2,3]。ロンバート効果は音声認識や話者認識においても応用されている[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り，雑音下での発話の特徴が残響下でも同様にみられるとは限らない。その一例として clear speech（聴覚障害者に話しかけることを想定し，静かな環境で明瞭に発話された音声）[5]があげられる。雑音下では通常発話の音声よりも clear speech の正解率が上昇したが[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]。

雑音と残響下で特性が異なる理由の一つとして，マスキングパターンの違いがあげられる。雑音下では同時マスキングが発生するのに対し，残響下ではマスキングが発話と同時に起こる self-masking に加え，反射音が遅延することでマスキングが発話よりも遅れる overlap-masking[8]が発生する。

また、雑音下では発話中の音声と発話者が聴取する雑音に相関がなく，提示雑音によって聴覚フィードバックが減少し，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 overlap-masking によって，聴覚遅延フィードバック（DAF）のように発話中の音声と発話者が聴取する残響音に相関がみられる。DAF 下の発話ではロンバ

ート音声で観測された特徴のようにインテンシティ・F0 の増加や話速が低下するが[9]，発話の間違いや脱落によって音声明瞭度は低下する[9]。

本研究の最終的な目的は，雑音や残響が存在する公共空間で利用者に明瞭な音声案内を提供することである。これには音声生成・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) 残響下での発話にはどのような音声的特徴が観測されるのか，(2) (1) は雑音下での観測[1-3]と同じか，(3) (1) は残響時間によって変化するかを調査した。本稿で用いる音響的特徴として，先行研究[1-3]で使いられたものを参考に，周波数のファクタ（F0，F1，F2），振幅のファクタ（発話レベル，子音と母音のインテシティ比（CVR)），時間のファクタ（単語の長さ）とした。

2. 録音

2.1. 発話者

東京方言話者 4 名（男女 2 名ずつ，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 10 単語、(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。なお原音声は、今後行う聴取実験で検討を行うため，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，2 : 予測度が高い）が異なるように選定した。

発話環境は，表 1 に示す静か（Q)，雑音（N)，2 種類の残響 (R1, R2) の 4 条件とした。N は白色雑音を，R は残響時間の異なる 2 種類のインパルス応答を使用した。本稿では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 における初期減衰時間の平均を残響時間とし、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

録音環境を図 1 に示す。録音は防音室で行い，音声はヘッドセットマイク（SHURE, Beta-53)，マイクアンプ（PreSonus, DIGIMAX FS)，オーディオインタフェイス（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オーディオインタフェイス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

表 1　発話環境

発話環境	提示	音圧レベル
静か（Q)	なし	—
雑音（N)	ホワイトノイズ	発話者の耳元で平均 80 dBA
残響（R1，R2）*	残響音	

* 残響時間は R1 で 3.5 s，R2 で 12.3 s

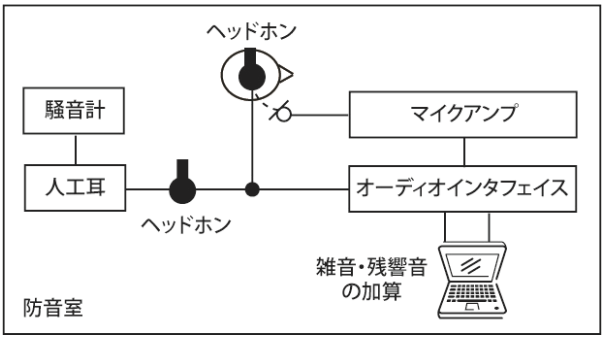

図 1 録音環境

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。

録音の手順は，まず練習として 8 試行（2 文×4 発話環境）行い，60 試行（15 文×4 発話環境）を 1 セットとして休憩を挟み計 2 セットを録音した。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，できるだけ明瞭に発話するよう教示を行った。録音は Q 条件の後に N・R 条件を行った。各発話条件内の 15 文の順番と、N・R 条件の順番は発話者ごとにカウンタバランスをとった。

3. 音響分析

図 2 に，ターゲットの F0，F1，F2，CVR，音圧レベル，長さを，発話者と発話環境毎に示す。図 2 のアスタリスクは，音響的特徴の平均値に対して，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 Q-N, Q-R1，Q-R2，R1-R2 間において 1%水準で有意なものを示す。図 3 に，母音ごとの F1-F2 の分布を発話者と発話環境毎に示す。

なお，原音声（1）（2）のターゲットの音響的特徴は似た傾向がみられたため，本稿では全ターゲットの平均値に対して分析を行なう。また，音響分析で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。

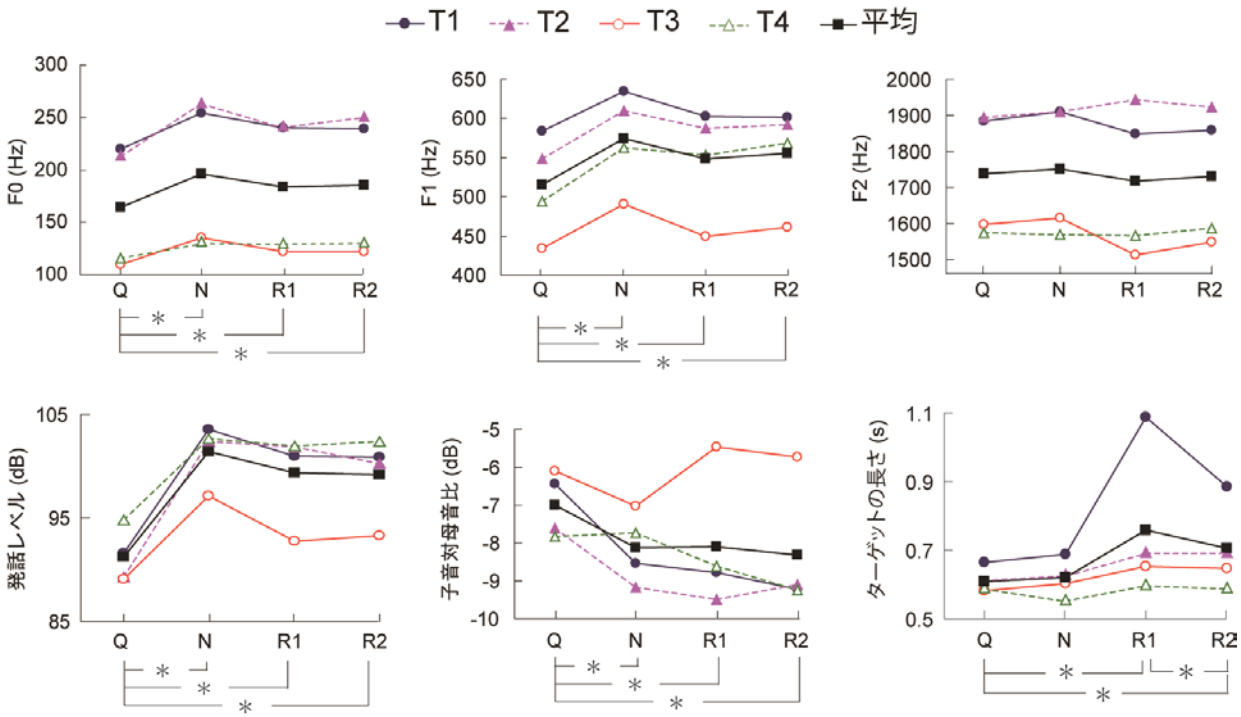

図 2 発話者（T1-T4 とその平均）と発話環境（Q: 静か, N: 雑音, R1, R2: 残響）における音響的特徴。
＊は 1%水準で有意差がみられた条件を示す。

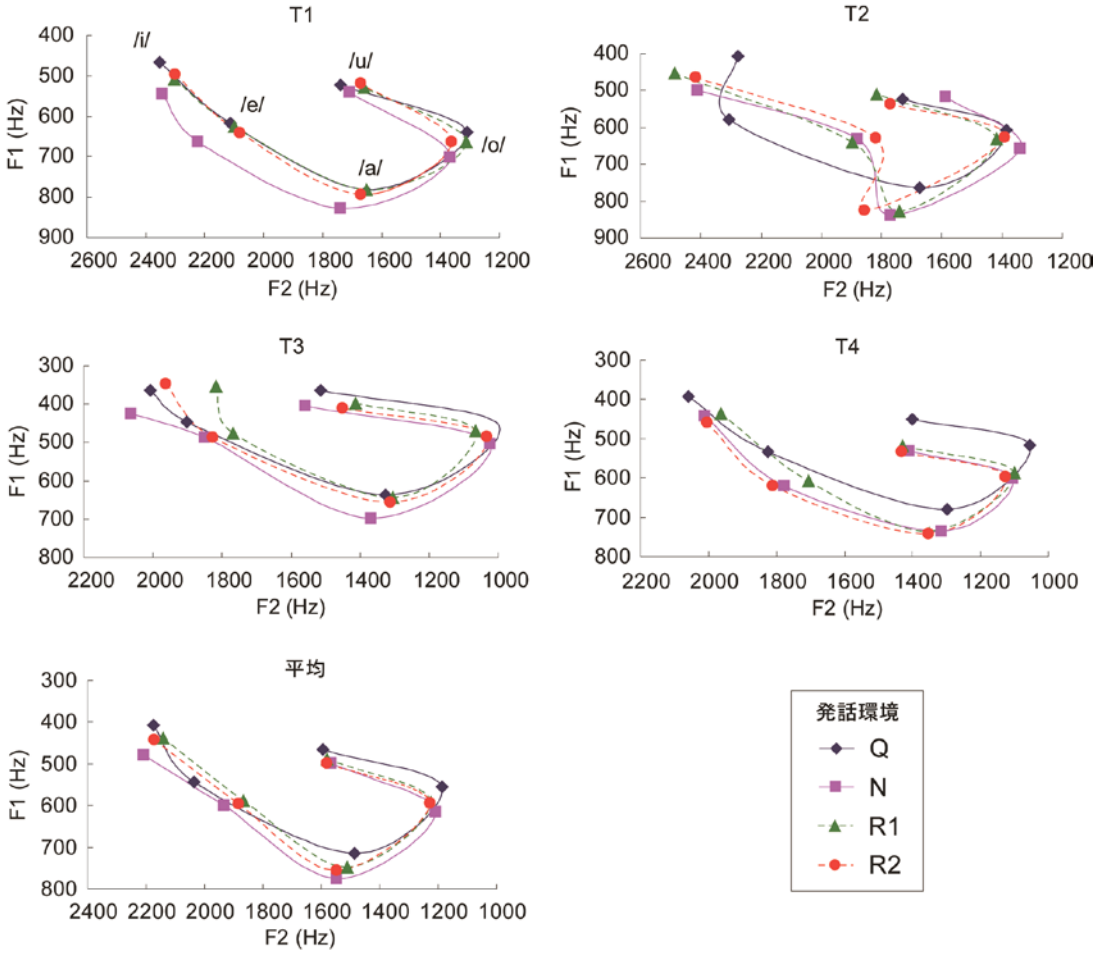

図 3　発話者と発話環境における母音空間

3.1. 周波数のファクタ

F0 と F1 は、Q 条件と比べて N・R 条件ともに増加した。これは，ロンバート音声や DAF の特徴[1-3,9]と同様に，発話が強調されたためだと考えられる。一方，F2 は Q 条件と比較して N・R 条件ともに有意な変化はみられず、先行研究[3]とは異なる結果となった。

F1-F2 の分布については、先行研究では Q 条件と比べて N 条件で F1 と F2 が増加（特に F1 の増加分が大きい）し、その変化の度合いは発話者によって異なる[2,3]。図 3 より、/i, a, o/の発話者の平均では、N・R 条件共に先行研究[2,3]と同様の傾向が得られている。また、N 条件よりも R 条件の方が増加分は少なくなった。一方/e, u/の発話者の平均では、N・R 条件共に F1 は上昇したが、F2 は減少した。

3.2. 振幅のファクタ

CVR は、Q 条件と比較して N・R 条件ともに減少した。これは，ロンバート音声の特徴[1-3]と同様に母音が強調されたためと考えられる。

発話レベルは、Q 条件と比較して N・R 条件ともに増加したが，Q 条件に対する増加量は N 条件よりも少なくなった。これは、まず発話の強調から，音圧レベルが Q 条件下よりも増加する。しかし，音圧レベルが増加しすぎると，overlap-masking も同様に増加する。それを防ぐために，発話者の音圧レベルの増加が N 条件に比べて抑制されたと考えられる。

3.3. 時間のファクタ

ターゲット長は、Q 条件と比較して R 条件のみで増加した。これは、話速を低下することで overlap-masking を減少させたためだと考えられる。

4. 考察

R 条件では Q 条件と比べて音響的特徴が変化し、N・R 条件下で同様に変化をした特徴（F0，F1，発話レベルの増加，CVR の減少）がみられた。このことから，N・R 条件下では発話の強調により声門下圧、声帯の緊張、口腔の開きが増加し[12]、それに対応した特徴が N・R 条件下で同様に変化したといえる。

R1・R2 間で音響的特徴に差はみられたのはターゲット長のみであった。従って、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[2,3]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。

発話者間では音響的特徴に変動がみられ，その変動は N 条件よりも R 条件の方が増加した。その理由として，発話と相関がない白色雑音が提示される N 条件と比べて，R 条件では発話と相関がある残響音が提示され、その環境で発話者が異なる方法で発話をしていると考えられる。R 条件に関する発話者の内観報告から，発話が困難であると回答した発話者と，聴取する残響音を減少させるために話速や発話レベルの低下や子音の強調を行ったと回答した発話者がみられた。

表 2　発話レベルと CVR の線形近似曲線

		Q	N	R1	R2
T1	発話レベル	y=0.5x +87.5	y=0.6x +98.4	y=0.5x +97.1	y=0.6x +96.3
	CVR	y=0.5x -10.1	y=0.6x -13.6	y=0.9x -15.9	y=0.8x -15.7
T2	発話レベル	y=0.4x +86.4	y=0.6x +97.9	y=0.5x +98.0	y=0.4x +97.1
	CVR	y=0.5x -11.9	y=0.6x -13.9	y=0.6x -14.1	y=0.5x -13.2
T3	発話レベル	y=0.6x +84.5	y=0.7x +91.5	y=0.7x +87.4	y=0.6x +88.3
	CVR	y=0.8x -12.2	y=0.7x -12.5	y=0.6x -10.4	y=0.6x -10.7
T4	発話レベル	y=0.6x +90	y=0.7x +96.8	y=0.7x +96	y=0.8x +96.3
	CVR	y=0.7x -13.3	y=1.0x -16.1	y=0.7x -14.3	y=0.7x -14.6

音響的特徴の発話者の変動に関して興味深い点として、R 条件下の CVR が T3 のみで増加している、つまり子音が強調されていることである。これは、T3 の歌唱経験（4 年間の混声歌唱）が関連していると考えられる。歌唱においては、歌唱者はホールに応じて子音対母音比などの歌唱方法を変化させ[13]、残響下では子音を強調することで歌詞の明瞭性が向上するという知識や経験を持っている。また楽器の演奏においても、演奏者はホールの残響などに加え、演奏への意識、経験的な知識、空間の認知などの後天的フィードバックも利用して演奏を行う[14]。以上から、歌唱や演奏と同様、発話に関しても発話者の知識や経験も発話方法に大きく関連する可能性が示唆された。

T3 の R 条件下での子音強調は、残響によるマスキングの観点からすると理想的である。それは、残響下では overlap-masking[8]により、母音に対してエネルギの小さい子音が特にマスクされるからである[15,16]。この子音へのマスキングを軽減するため、残響が付加される前に音声の定常部を抑圧する信号処理を用いることにより、残響下で若年者および高齢者の音声明瞭度を改善している[17-19]。その一方、本稿では R 条件下の CVR は T3 を除いて母音が強調されており、仮に知覚実験で残響下の音声明瞭度が減少する場合には信号処理を施すなどの措置が必要であるかもしれない。

ここで N・R 条件下における CVR の減少は、発話レベルの上昇により相対的にエネルギの大きい母音のエネルギが増加したためか、それとも発話環境によるものかを調べるため、発話レベルと CVR を発話環境と発話者毎に昇順に並べ替えて算出した線形近似曲線を表 2 に示す。発話レベルと CVR では、Q 条件に対する N・R 条件の傾きの変化が異なることから、N・R 条件では

発話環境によって母音が強調されることが示された。

著者の聴覚印象から 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。鼻音化の発生には、聴覚障害者で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による可能性があげられている[21]。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ずれたことが考えられる。その結果、健聴な若年者であっても発話が鼻音化した可能性が考えられる。

5. おわりに

静か（Q)、雑音（N)、残響（R1, R2）環境下で発話された単語の音響的特徴を調査した。その結果，

（1）R 条件では Q 条件と比べて異なる音響的特徴を示した（F0、F1、発話レベル、時間長の増加，CVR の減少)，

（2）N・R 条件下の音響的特徴は同様の変化を示すものがあった（上記のうち時間長以外の特徴)，

（3）ターゲット長のみ R1・R2 間で差がみられた，

（4）発話者間の音響的特徴の変動は，N 条件よりも R 条件で増加した，

（5）N・R 条件共に発話者全員に母音が鼻音化し，その度合いは R 条件の方が強くなった。

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，発話が強調されるという点では周波数や振幅のファクタのように同じ特徴で評価できる場合と，マスキングなどの違いによって時間のファクタのように異なる特徴を示す場合があった。特に残響環境下では、発話者が聞いている環境音だけではなく、発話者の意識によっても発話が変化する可能性が示唆された。

本稿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が雑音・残響下の音声明瞭度にどう寄与するかの音声知覚側の調査を行いたい。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の検討を行いたい。

6. 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方々に感謝いたします。

文　献

[1] H. Lane and B. Tranel, “The Lombard sign and the role of hearing in speech,” J. Speech Hear. Res., 14, pp. 677-709, 1971.

[2] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow and M. A. Stokes, “Effects of noise on speech production: Acoustics and perceptual analysis,” J. Acoust. Soc. Am., 84, 917-928, 1988.

[3] J.-C. Junqua. “The Lombard reflex and its role on human listeners and automatic speech recognizers,” J. Acoust. Soc. Am., 93, 510-524, 1993.

[4] J. H. L. Hansen, “Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition,” Speech Commun., 20, 151-173, 1996.

[5] M. A. Picheny, N. L. Durlach, and L. D. Briada, “Speaking clearly for the hard of hearing I,” J. Speech and Hear. Res., 28, 96-103, 1985.

[6] K. L. Payton, R. M. Uchanski, and L. D. Braida, “Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,” J. Acoust. Soc. Am., 95, 1581-1592, 1994.

[7] N. Hodoshima, T. Arai, and K. Kurisu, “Effects of training, style, and rate of speaking on s peech perception of young people in reverberation,” Proc. Acoustics 08, 2393-2397, 2008.

[8] A. K. Nabelek, T. R. Letowski and F. M. Tucker, “Reverberant overlap- and self-masking in consonant identification,” J. Acoust. Soc. Am., 86, 1259-1265, 1989.

[9] A. J. Yates, “Delayed auditory feedback,” Psychol. Bull., 60, 213-232, 1963.

[10]“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株式会社, 1999.

[12] H. Traunmuller and A. Eriksson, “Acoustic effects of variation in vocal effort by men, women, and children,” J. Acoust. Soc. Am., 107(6), 3438-3451, 2000.

[13] K. Kato, K. Fujii, T. Hirawa, K. Kawai, T. Yano and Y. Ando, “Investigation of the relation between minimum effective duration of running autocorrelation function and operatic singing with different interpretation styles,” Acta Acustica United with Acustica, 93, 421-434, 2007.

[14] 上野佳奈子, “演奏空間のリアリティに関する一考察,” 日本音響学会秋季研究会講演論文集, 1429-1430, 2009.

[15] V. O. Knudsen, “The hearing of speech in auditoriums,” J. Acoust. Soc. Am., 1(1), 56-82, 1929.

[16] 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.

[17] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, “Effects on suppressing steady-state portions of speech on i ntelligibility in reverberant environments,” Acoust. Sci. Tech., 23(4), 229-232, 2002.

[18] N. Hodoshima, T. Arai, A. Kusumoto and K. Kinoshita, “Improving syllable identification by a

preprocessing method reducing overlap-masking in reverberant environments," J. Acoust. Soc. Am., 119(6), 4055-4064, 2006.

[19] 小林敬, 安啓一, 程島奈緒, 荒井隆行, 進藤美津子, "母音のエネルギ定常部の抑圧による高齢者に対する音節強調の検討," 日本音響学会誌, 64(5), 278-289, 2008.

[20] M. Y. Chen, "Acoustic correlates of English and French nasalized vowels," J. Acoust. Soc. Am., 102(4), 2360-2370, 1997.

[21] S. G. Fletcher and D. A. Daly, "Nasalance in utterances of hearing-impaired speakers," J. Commun. Disorders, 9, 63-73, 1976.