

## 雑音・残響環境下の拡声音声に適した発話に関する検討

程島奈緒<sup>†</sup> 荒井隆行<sup>‡</sup> 栗栖清浩<sup>††</sup>

<sup>†</sup> 東海大学情報通信学部 〒108-8619 東京都港区高輪 2-3-23

<sup>‡</sup> 上智大学理工学部 〒102-8554 東京都千代田区紀尾井町 7-1

<sup>††</sup> TOA 株式会社 〒665-0043 兵庫県宝塚市高松町 2-1

E-mail: <sup>†</sup> hodoshima@tokai-u.jp, <sup>‡</sup> arai@sophia.ac.jp, <sup>††</sup> kurisu\_kiyohiro@toa.co.jp

**あらまし** 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と比べて一般的に明瞭であり、若年者に対しては複数の報告がある (Lombard effect)。著者らは、残響下で発話された音声も雑音下と同様に、若年者に明瞭であることを示した。本研究では、高齢者に対しても雑音下・残響下で発話された音声と同様に明瞭であるかを調査した。静かな環境/雑音/残響下で発話された音声を、32名の高齢者が雑音/残響下で聴取する実験を行った。その結果、雑音/残響下で発話された音声は、静かな環境で発話された音声と比べて、雑音/残響下での明瞭度が有意に上昇した。このことから、雑音もしくは残響を聞きながら発話をした音声を公共空間で拡声することで、音声案内がより明瞭であることが示唆された。本結果を受けて、公共空間で高い明瞭度を得ることを目的とした、発話環境の構築や発話者などへの発話訓練が考えられる。

**キーワード** 音声明瞭度, 音声生成, Lombard effect, 雑音, 残響

## Study on speech production that is suitable for speech announcements in noisy/reverberant environment

Nao HODOSHIMA<sup>†</sup> Takayuki ARAI<sup>‡</sup> and Kiyohiro KURISU<sup>††</sup>

<sup>†</sup> Faculty of Information and Telecommunication Engineering, Tokai University 2-3-23 Takanawa Minato-ku Tokyo 108-8619 Japan

<sup>‡</sup> Faculty of Science and Technology, Sophia University 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan

<sup>††</sup> TOA Corporation 2-1 Takamatsu-cho, Takarazuka-shi, Hyogo, 665-0043 Japan

E-mail: <sup>†</sup> hodoshima@tokai-u.jp, <sup>‡</sup> arai@sophia.ac.jp, <sup>††</sup> kurisu\_kiyohiro@toa.co.jp

**Abstract** Speech uttered in noise is generally more intelligible than speech produced in quiet when they are heard in noise, and several studies have been carried out for young people (i.e., Lombard effect). The authors showed that reverberation-induced speech as well as noise is more intelligible to young people than speech spoken in quiet. This paper examines whether noise/reverberation-induced speech is also intelligible for elderly people. We conducted listening tests wherein 32 elderly people listened to speech that was uttered in quiet/noise/reverberation with white noise or with reverberation. The results showed that noise/reverberation-induced speech were more intelligible than speech spoken in quiet for elderly listeners. The results suggest that using noise/reverberation-induced speech may improve the intelligibility of speech announcements in public spaces. The results lead to constructing speaking environments and conducting speech training for speakers at public spaces in order to yield intelligible speech.

**Keyword** speech intelligibility, speech production, Lombard effect, noise, reverberation

### 1. はじめに

駅などの公共空間の多くには雑音や残響が存在するため[1], 音声案内の聞き取りが難しいことがある。さらに、高齢者は健聴な若年者と比べて、雑音や残響下で音声明瞭度が低下する[2]。高齢者の人口は世界的に増加しており、我が国では高齢者の全人口に占める

割合が 2009 年度で 22.7%[3]と、世界で最も高い比率になっている。火災発生時など、音声による緊急放送のような高い明瞭性が求められる音声案内を拡声する場合には、特に考慮が必要である。高齢者にとっても音声案内が明瞭である公共空間の設計の必要性は今後増加するだろう。

公共空間の音声の明瞭度を改善する手法としては、これまで建築的・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[4-6]。これらは、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。

人は周囲の雑音に応じて発話を変化させることがある。これは Lombard 効果[7-9]として知られ、音声生成の面では、雑音下での発話の方が静かな場所での発話に比べて、時間長・インテンシティ・ $F_0 \cdot F_1 \cdot F_2$  などが増加する[7-9]。音声知覚においては、雑音下の発話の方が静かな場所での発話に比べて、雑音下で若年者の単語理解度が増加する[7,8]。Lombard 効果は音声認識や話者認識において応用されている[10,11]。

残響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[12]が発生するため、雑音による同時マスクングとは時間・周波数的に音声知覚への影響が異なる。我々は残響下においても、雑音下での Lombard 効果と同様の結果を得ている[13,14]。残響下での発話は、静かな場所での発話に比べて、時間長・インテンシティ・ $F_0 \cdot F_1 \cdot F_2$  の増加と子音対母音のインテンシティ比 (CVR) の減少がみられた[13]。そして、残響下の発話の方が静かな場所での発話に比べて、残響下で若年者が聞いた時に単語理解度が上昇した[14]。我々が知る範囲においては、雑音下・残響下の発話の明瞭度を高齢者に対して調べた研究はない。

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。本稿では雑音下・残響下で発話された音声で、雑音や残響下で高齢者が聞いた時に明瞭になるかを調査し、若年者の結果[14]と比較する。残響下では、発話と聴取の残響条件[残響時間 (RT)] が異なる場合において、明瞭性は向上するか調査した。また、ターゲット語の明瞭度は、ターゲット語だけではなくターゲットに先行する語句も残響下で発話をした方が明瞭になるのかも調査した。最後に得られた結果から、公共空間での発話訓練への応用について述べる。

## 2. 聴取実験

### 2.1. 参加者

日本語を母語とする高齢者 32 名 (男性 12 名, 女性 20 名, 年齢 65~85 才, 平均年齢 75 才) が実験に参加した。参加者の気導聴力レベルの平均を図 1 に示す。

### 2.2. 刺激

刺激は、信号対雑音比 (SNR) と RT 以外は先行研究[14]と基本的に同じである。

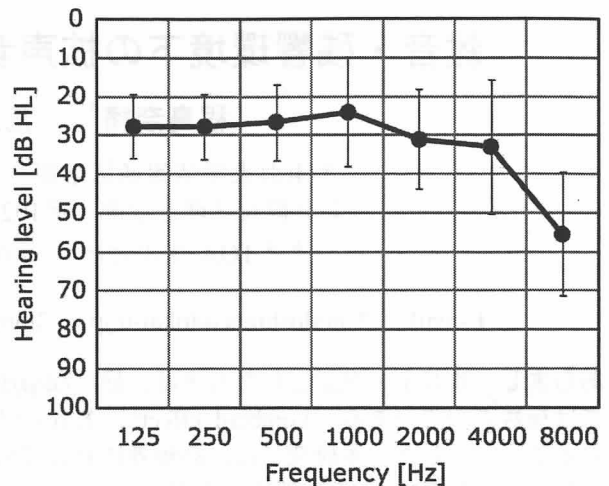


図 1 参加者の気導聴力レベルの平均

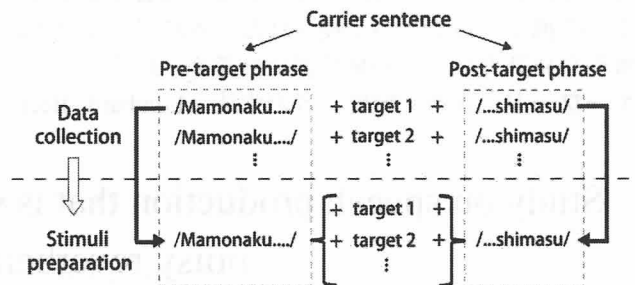


図 2 原音声の録音 (上) と刺激の作成 (下) の概略図。刺激ではキャリア文 36 文から 1 文を使用した。

#### 2.2.1. 録音

発話者は東京方言話者 2 名 (男女 1 名ずつ, 年齢 20, 22 才) である。原音声は図 2 上に示すような、キャリア文に挿入された 4 モーラのターゲット語である。ターゲット語として、FW03 [15]より親密度が 2.5~4.0 の 36 単語を使用した。

録音の環境を図 3 に示す。発話者はヘッドホン (SENNHEISER, HDA200; dynamic, closed circumaural type) を装着し、音声は防音室においてマイク (SHURE, KSM109), マイクアンプ (PreSonus, DIGIMAX FS), 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・残響音の付加は Adobe Audition 3.0 を使用した。ソフトウェアによる遅延時間は数 ms 以内である。

発話者は表 1 に示す 3 条件[静か (Q), 雑音 (N), 残響 (R)]下で発話を行った。N は白色雑音を, R は教会で測定されたインパルス応答を使用した。R の RT は 125-4000 Hz の平均で 3.6 s である。雑音・残響音の提示には直接音は含まれておらず、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。録音音声には、雑音・残響音のいずれも含まれていない。

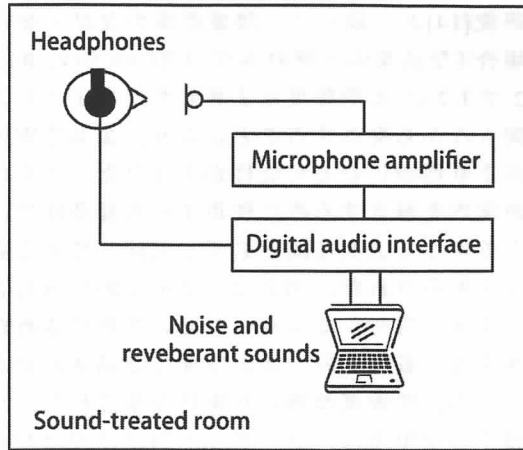


図3 録音の環境

表1 発話条件

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

各発話者は108文(36文×3発話条件)を録音した。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。

### 2.2.2. 刺激の作成

刺激の作成にあたって、ターゲットに先行する語句(図2の pre-target phrase)からターゲット語への overlap-masking 量の統制をとるため、発話条件内のターゲット語を同じターゲットに先行する語句とターゲットに後続する語句(図2の post-target phrase)の間に挿入した(図2下)。キャリア文とターゲット語の組み合わせは、Q(キャリア文とターゲット語ともにQ)、N(キャリア文とターゲット語ともにN)、R(キャリア文とターゲット語ともにR)の他、QR(キャリア文はQ、ターゲット語はR)の4種類(spliced conditions, 表2の列)とし、同一発話者の発話を使用した。キャリア文とターゲット語のインテンシティ比と無音区間の統制を発話条件と発話者毎にとった(無音区間の平均は0.07s)。

表2 聴取条件

		Spliced conditions			
		Q	N	R	QR
Listening conditions	N (SNR=0 dB)	Q_N	N_N	/	/
	R1 (RT=1.4 s)	Q_R1	/	R_R1	QR_R1
	R2 (RT=2.4 s)	Q_R2	/	R_R2	QR_R2

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。表2に8種類の聴取条件を示す。表2の記号の表記は、例えばQ\_Nは録音がQ条件、聴取実験がN条件であることを示す。聴取実験でN条件のSNRは0dB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Rの包絡を推定し、それを異なる時定数を持つ包絡に置換することでRTを1.4s(R1)と2.4s(R2)に短くしたものである。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。

### 2.2.3. 手順

聴取実験は参加者毎に防音室内で行った。刺激はコンピュータにオーディオインタフェース(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。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。

## 3. 結果

図4に各聴取条件における各発話者(S1, S2)と発話者の平均(average)のモーラ正解率を示す。発話者と聴取条件の主効果のいずれとも有意であった( $p < 0.01$ )。

N\_Nは話者の平均( $p=0.022$ )とS1( $p=0.001$ )でQ\_Nよりも正解率が有意に高くなった。S2については有意ではなかった。この結果より、雑音下での発話は静かな場所での発話よりも、雑音下で聞いた時に明瞭度が高くなることが示され、若年者に対するLombard効果の先行研究[7,8,14]と一致した。

R\_R1はS2( $p=0.032$ )でQ\_R1よりも正解率が有意に高くなり、R\_R2は話者の平均( $p=0.036$ )とS1( $p=0.022$ )でQ\_R2よりも正解率が有意に高くなった。R\_R1についてはS1や平均では有意ではなく、R\_R2

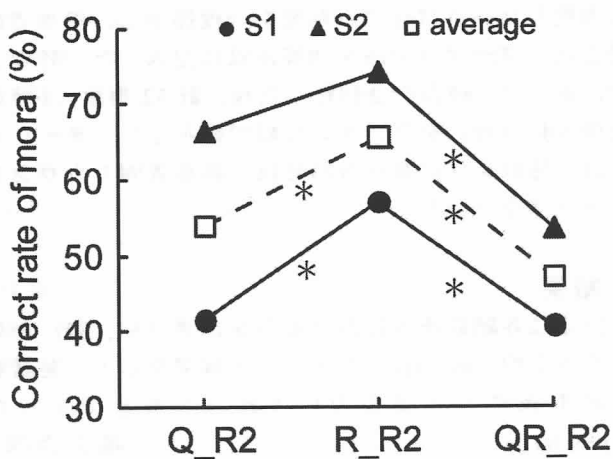
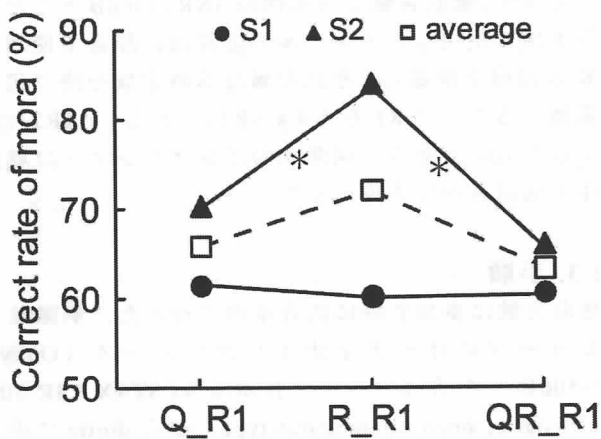
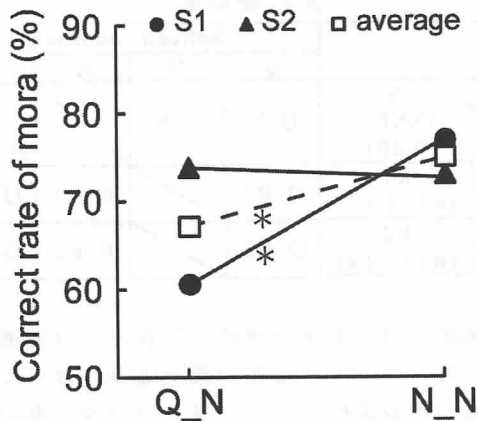


図4 各聴取条件（上：N，中央：R1，下：R2）における各発話者（S1：丸，S2：三角）と発話者の平均（average：四角）のモーラ正解率。アスタリスクは5%水準の有意差を示す。

についてはS2では有意ではなかった。この結果から、残響下の発話は静かな場所での発話よりも、残響下で聞いた時に明瞭になることが示され、若年者に対する

先行研究[14]と一致した。残響環境が発話と聴取で異なる場合（発話条件と聴取条件のRTの差は、R1で2.2s，R2で1.2s）で明瞭度は上昇した。以上より、残響音を聞きながら発話することにより、公共空間の音声案内がより明瞭になる可能性が示された。さらに、その音声案内を録音する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。例えば、他の空間で録音されたインパルス応答やシミュレーションで作成されたインパルス応答の畳み込み、他の空間で発話された音声で代用しても、明瞭度改善の可能性が示された。

残響下の明瞭度は、ターゲット語だけではなくターゲットに先行する語句も残響下で発話をした方が明瞭になるのかも調査した。まず、Q\_R1とQR\_R1及びQ\_R2とQR\_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。次に、R\_R1はS2( $p=0.003$ )でQR\_R1よりも、R\_R2は話者の平均( $p<0.01$ )、S1( $p=0.015$ )、S2( $p=0.001$ )でQR\_R2よりも正解率が有意に高くなった。よって、ターゲットに先行する語句が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。残響下の発話は静かな環境での発話と比べて無音区間・基本周波数・第一フォルマントの増加、CVRの減少がみられる[13]。そのため、キャリア文から残響下で発話をした方が、ターゲットのみを残響下で発話をするよりもoverlap-masking量が減少すると考えられる。つまり、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、ターゲットに先行する語句も考慮が必要があることが示された。

R1およびR2ともQ\_R1およびQ\_R2よりも正解率が有意に高くなった結果は若年者の結果[14]と一致したことから、残響下の発話は若年者だけではなく高齢者に対しても明瞭であることが示された。明瞭度の改善が得られたRTは、高齢者(1.4sと2.4s)の方が若年者(2.6sと3.6s)よりも短くなった。その理由の一つに、参加者の年齢と聴力の違いが挙げられる。高齢者の平均年齢は75才で、高周波数帯域で平均聴力が低下している。加齢と聴力は時間的に歪んだ信号の聴取に影響を与えることから、高齢者は若年者と比べて残響が付加された音声の聴取がより困難になったと考えられる[16,17]。

雑音/残響下で発話された音声の明瞭度は、先行研究[14]と同様N,R1,R2ともに発話者によって変化した。これは、発話者によって雑音や残響下での発話の仕方が異なる（例えば、F0や時間長の増加のさせ方やCVRの減少のさせ方[13]）ためだと考えられる。

公共空間で音声案内を拡声する際には、雑音や残響を聞きながら発話をした方が明瞭になる可能性が示された。よって、本実験結果を公共空間に反映し、高い明瞭度を得られるような発話環境の構築を行いたい。また、雑音/残響下での発話のどのような音響的特徴量が明瞭度に関連しているかの体系的な調査が必要である。発話訓練を行うことで主観評価による聞き取りやすさが向上する[18]ことから、明瞭度に関連した音響的特徴量に基づく発話者(例:公共空間の職員)への発話訓練も検討したい。さらに、音声案内を合成するシステムへの応用についても検討したい。

#### 4. 結論

高齢者に対する聴取実験の結果より、雑音・残響下での発話は、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。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残響環境と拡声する残響環境が異なる場合でも得られた。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的特徴と明瞭度改善の対応を調査したい。さらに、公共空間の利用者に適切な情報を伝えるような発話者の訓練への応用を検討したい。

#### 5. 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。

#### 文 献

- [1] R. Shimokura and Y. Soeta, Evaluation of speech intelligibility of sound filed in underground stations, *Acoust. Sci. Tech.* vol.32, no.2, pp.73-75, 2011.
- [2] A. K. Nabelek and P. K. Robinson, Monaural and binaural speech perception in reverberation for listeners of various ages, *J. Acoust. Soc. Am.*, vol.71, no.4, pp.1242-1248, 1982.
- [3] The Cabinet Office, Annual Report on the Aging Society, Japan, 2010.
- [4] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments, *Acoust. Sci. Tech.*, vol.23, no.4, pp.229-232, 2002.
- [5] 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.*, vol.119, no.6, pp.4055-4064, 2006.
- [6] A. Kusumoto, T. Arai, K. Kinoshita, N. Hodoshima and N. Vaughan, Modulation enhancement of speech by a pre-processing algorithm for improving intelligibility in reverberant environments. *Speech Communication*, vol.45, no.2, pp.101-113, 2005.
- [7] H. Lane and B. Tranel, The Lombard sign and the role of hearing in speech, *J. Speech Hear. Res.*, vol.14, pp.677-709, 1971.
- [8] W. V. 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.*, vol.84, pp.917-928, 1988.
- [9] J.-C. Junqua, The Lombard reflex and its role on human listeners and automatic speech recognizers, *J. Acoust. Soc. Am.*, vol.93, pp.510-524, 1993.
- [10] J. H. L. Hansen, Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition, *Speech Communication*, vol.20, pp.151-173, 1996.
- [11] J. H. L. Hansen and V. Varadarajan, Analysis and compensation of Lombard speech across noise type and levels with application to in-set/out-of-set speaker recognition, *IEEE Trans. Audio, Speech, Lang. Process.*, vol.17, pp.366-378, 2009.
- [12] A. K. Nabelek, T. R. Letowski and F. M. Tucker, Reverberant overlap- and self-masking in consonant identification, *J. Acoust. Soc. Am.*, vol.86, pp.1259-1265, 1989.
- [13] 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報告, vol.SP2009-69, pp.43-48, 2009.
- [14] N. Hodoshima, T. Arai and K. Kurisu, Intelligibility of speech spoken in noise and reverberation, *Proc. ICA*, 2010.
- [15] 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), 音声資源コンソーシアム, 2006.
- [16] S. Gordon-Salant and P. J. Fitzgibbons, Profile of auditory temporal processing in older listeners, *J. Speech Lang. Hear. Res.*, vol.42, pp.300-311, 1999.
- [17] P. J. Fitzgibbons and S. Gordon-Salant, Age effects on discrimination of timing in auditory sequences, *J. Acoust. Soc. Am.*, vol.116, no.2, pp.1126-1134, 2004.
- [18] 鹿島田千帆, 長谷川光司, 石川智治, 阿山みよし, よりよい聴取印象を与える音声の研究~現役のアナウンサーによるボイストレーニング効果~, 日本音響学会春季講演論文集, pp.1605-1606, 2011.