屋外拡声を想定した音声の発話速度が単語了解度に及ぼす影響*

☆大田敦也，荒井隆行，安啓一（上智大・理工）

1　はじめに

防災行政無線などで使用される屋外拡声放送は屋外の建物や騒音，音声信号が拡声システムを通る間に受ける歪みなどによって音声明瞭性が低くなることが知られている[1]。東日本大震災においても政府が行ったアンケートによると「はっきりと聞き取ることが出来た」と回答した人は 6 割弱にとどまった[2]。

残響下での音声明瞭度の研究は広く進められており[3]，屋外放送の音声明瞭度についても研究は行われている。しかし，屋外放送の音声明瞭度についてはとても低いことが報告されている。これは屋外環境下には室内残響下では見られないロングパスエコーと呼ばれる孤立反射音が存在することが一つの大きな要因であると指摘されている[1]。また，他の要因として，音声信号が受ける歪みも挙げられる。先行研究で屋外のスピーカから流れる音声は，再生可能な周波数帯域やダイナミックレンジが狭いことなどによって音声伝送の質が低下する[4]ことが報告されている。

一般的に残響下で音声明瞭度を改善する方法として，発話速度を遅くすることが試みられている[5]。しかし，ロングパスエコーを想定し，後続音の時間間隔を変更しながら正答率を調査した研究は報告されているものの[6]，発話速度そのものを遅くすることで屋外環境下でも明瞭度が改善されるかどうかについてはまだ十分に検討されていない。

よって本研究では，屋外拡声において発話速度が音声明瞭度に及ぼす影響に着目した。特に発話速度の制御を，調音速度と文節間ポーズ長という 2 つのパラメータを介して行った。調音速度と文節間ポーズ長については次節で説明する。また，音声が受ける歪みを再現するために音声信号に対して帯域制限とクリップ処理を施した。

2　実験

日本語が母語である若年者を対象に，調音速度と文節間ポーズ長を変えた刺激による単語了解度試験を防音室環境下で行った。

2.1 参加者

参加者は日本語母語話者 19～27 歳（平均 21.7 歳）の健聴者 24 名（男性 16 名・女性 8 名）であった。ただし，健聴か否かは参加者の自己申告とした。

2.2 発話速度

発話速度を変えるにあたって，調音速度と文節間ポーズ長を変化させた。ここで調音速度とは，ポーズのない文節内での単位時間あたりのモーラ数(mora/s)とした。文節は単語と接語をセットとしたものを意味する。また文節間ポーズ長とは，文節と文節の間に挿入するポーズの長さとした。

調音速度は Praat を使用し PSOLA (Pitch Synchronous Overlap and Add)法で変換した。原音声の調音速度から対象の調音速度への倍率を求め duration point を追加して作成した。全部で 4 条件あり， 3 mora/s, 4 mora/s, 5 mora/s, 6 mora/s である（それぞれ w1, w2, w3, w4 とする）。

文節間ポーズ長は 400 ms, 700 ms, 1000 ms の 3 条件とした（それぞれ z1, z2, z3 とする）。

したがって，発話速度は調音速度の 4 条件と文節間ポーズ長の 3 条件の組み合わせによる計 12 条件である。

表 1　調音速度 w

表記	w1	w2	w3	w4
調音速度 (mora/s)	3.0	4.0	5.0	6.0

表 2　文節間ポーズ長 z

表記	z1	z2	z3
文節間ポーズ長 (ms)	400	700	1000

* Effect of speaking rate on word intelligibility of speech for outdoor mass notification, by OTA, Atsuya, ARAI, Takayuki and YASU, Keiichi (Sophia Univ.).

2.3 インパルス応答

インパルス応答は IR1 と IR2 の 2 種類とした。サンプリング周波数はどちらも 44.1 kHz である。

IR1 は，屋外環境下の特徴であるロングパスエコーを模擬して人工的に作成した。全部で 5 個のパルスを 500 ms ごとに設け, 振幅は順に 1, 1/2, 1/3, 1/4, 1/5 となるようにした。振幅の変化は，実環境の反射を考慮し，決定した。IR1 の時間波形とスペクトログラムを図 1 に示す。IR2 は屋外の実環境で測定したインパルス応答である。IR2 の時間波形とスペクトログラムを図 2 に示す。

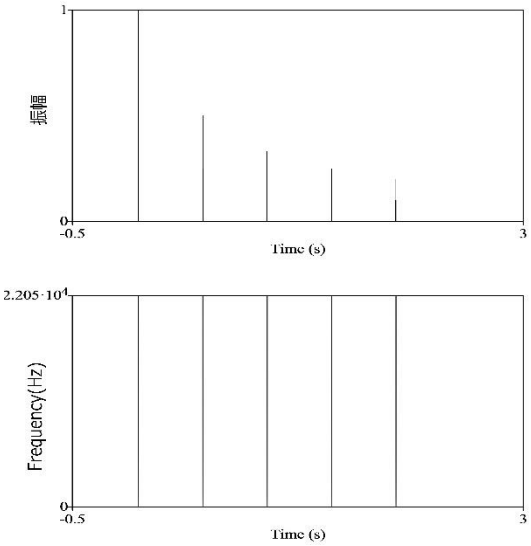

図 1　インパルス応答 IR1 の時間波形（上段）とスペクトログラム（下段）

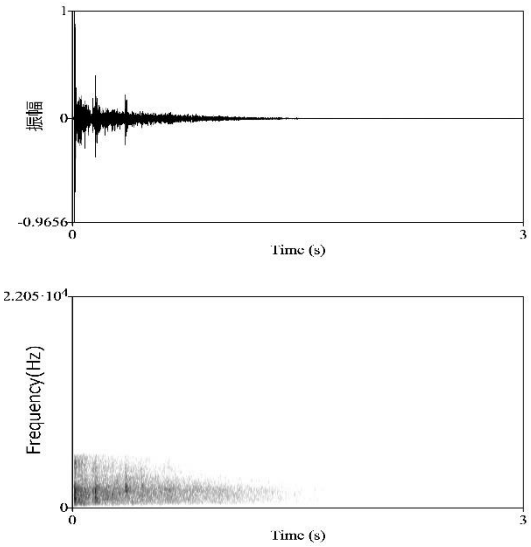

図 2　インパルス応答 IR2 の時間波形（上段）とスペクトログラム（下段）

2.4 音声処理

クリップ処理では，音声信号の振幅を 100 倍にした後，振幅の絶対値が 1 以上のときは 1 にクリッピングし, 1 以下のときはそのままの振幅で保存した。帯域制限では，音声信号の周波数特性を, 32 次の FIR フィルタにより電話の帯域である 300 - 3400 Hz に制限した。

ロングパスエコーを模擬した IR1 で畳み込んだ音声を予備実験として聴取したところ，非常にクリアな音声であり音声明瞭度の低下がほぼみられなかったため，IR1 を使用する刺激には，ピンクノイズを SN 比 0 dB になるように足し合わせた。

2.5 ターゲット語

本実験で使用した単語は，親密度別単語了解度試験用音声データセット 2007 (FW07) [7] より用いた。高親密度（7.0 - 5.5）と中低親密度（2.5 - 4.0）から日本語 4 モーラを 24 個ずつ計 48 個選出した。原音声は，男性 1 名の音声を使用した。

2.6 キャリア文

キャリア文は「これから__流す__単語は__〇〇〇〇__です」とした。「__」は文節間で挿入するポーズの位置を示し,「〇〇〇〇」は挿入するターゲット語の位置を示す。キャリア文の音声は，ターゲット語の発話者と同一男性の音声を使用した。

2.7 刺激

前節で述べたキャリア文の「__」の部分に文節間ポーズを挿入した。調音速度の変更は，ターゲット語だけでなくキャリア文にも行った。そして，キャリア文にターゲット語を挿入した音声にクリップ処理，帯域制限を施した後にインパルス応答を畳み込んだ。IR1 で畳み込んだ音声にはさらにピンクノイズを加えた。最後に全ての刺激について振幅の RMS 値（実効値）を全て同じ値に揃えた。

2.8 手順

聴取実験は上智大学の荒井研究室の防音室で行われた。刺激はスピーカ（ヤマハ MSP-3）から提示した。提示レベルは，騒音レベル(A 特性)で 50 dBA とした。各刺激は 1 回ずつ流れるものとし，参加者は各刺激が流れるごとに聞こえたと思う音声を平仮名で PC にタイプ入力した。

調音速度 4 条件×文節間ポーズ長 3 条件×インパルス応答 2 条件の 24 条件に対して, 高親密度語と中低親密度語に割り当てた。一人 48 刺激, 実験参加者 24 名で総刺激である 1152 刺激を各 1 回ずつ提示するようにカウンタバランスを施した。

2.9 仮説

実験に先立ち, 以下のように仮説を立てた。

仮説 1:調音速度(w)が低下するにつれ, 単語了解度が上昇する。

仮説 2:文節間ポーズ長が長くなるにつれ, ロングパスエコーの影響を受けにくくなり単語了解度が上昇する。

仮説 3:各条件において，ロングパスエコーによってターゲット語と反射音が重複する場合，単語了解度およびモーラごとの正答率が低下する。

3 実験結果

3.1 単語了解度試験

ターゲット語は有意味語であるため，回答した 4 モーラの平仮名と回答が完全に一致した時のみを正答とした。以降の図では横軸は条件を示し, 調音速度(w), 文節間ポーズ長(z)の順で表した。図 3 に実験結果一覧を示す。

IR1 と IR2 の調音速度ごとの正答率を図 4 に示す。IR1 の調音速度ごとの正答率は，3 mora/s から 5 mora/s にかけて少しずつ増加する結果になった。IR2 においてはほぼ変化が見られなかった。統計ソフトウェア SPSS を用いて分散分析を行ったところ，IR1 において，調音速度ごとの正答率では 5%以下で有意差が認められた。Tukey の多重比較検定では， 3 mora/s（平均 62.5%）と 5 mora/s（平均 70.8%）間に有意差 ($p < 0.05$) が見られた。

3.2 モーラごとの正答率

次に IR1 と IR2 のモーラごとの正答率を求めた。図 5 に示す。IR1，IR2 どちらにおいても 2 モーラ目の正答率が一番高かった。Tukey の多重比較を行った結果，IR1 においては，1 モーラ目と 2, 3, 4 モーラ目において有意差（$p < 0.05$）が見られた。IR2 においては，2 モーラ目と 1, 3, 4 モーラ目において有意差（$p < 0.05$）が見られた。

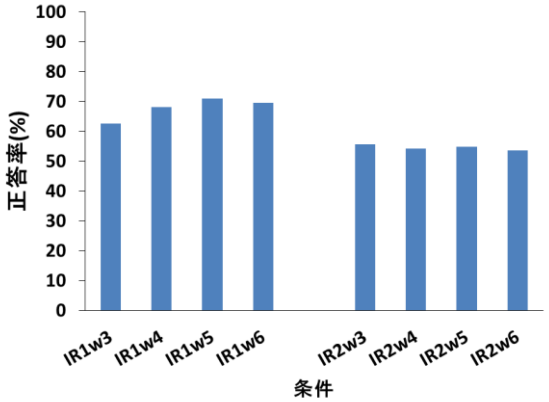

図 4 調音速度ごとの正答率

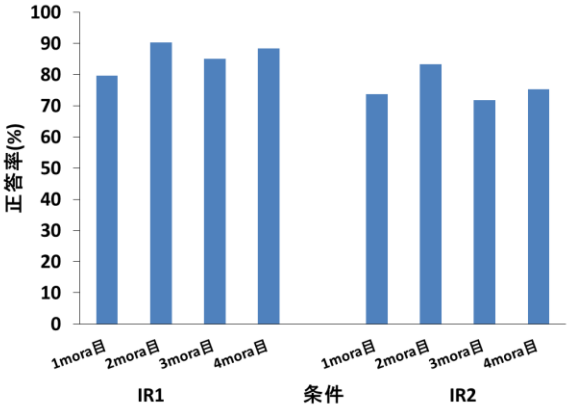

図 5 モーラごとの正答率

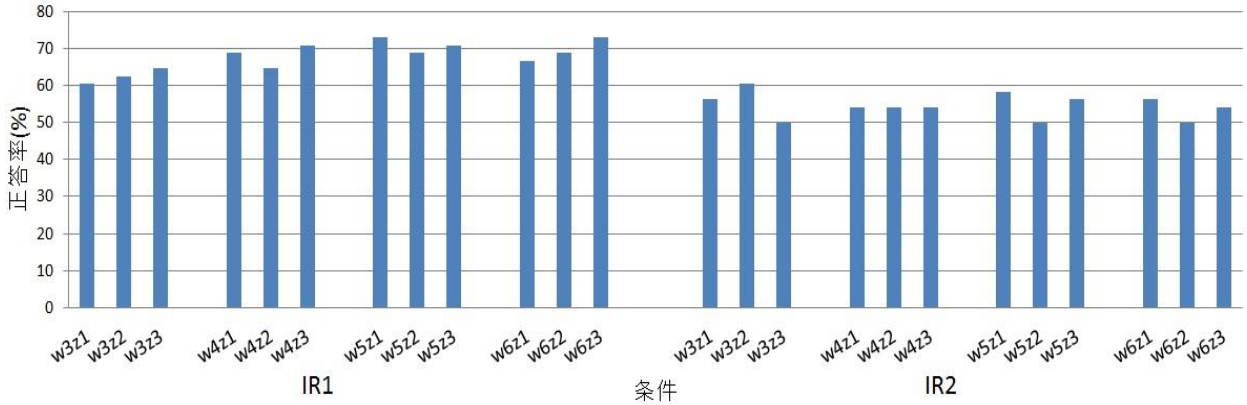

図 3 実験結果一覧

4　考察

4.1 調音速度について

図 4 より，IR1 の条件下において，調音速度ごとで 3 mora/s と 5 mora/s 間に有意差($p < 0.05$)が見られた。仮説 1 では調音速度が遅くなるにつれ単語了解度は上がるとしたが，今回は 5 mora/s の方が正答率は高くなり，仮説 1 は支持されなかった。これは調音速度が遅くなることによって，不自然さが増している可能性や母音部のエネルギーが大きくなりオーバーラップマスキングが増え明瞭性が低くなっていることによるものと考えられる。

4.2 文節間ポーズ長について

図 3 より IR1 において，1000 ms の条件下の正答率が一番高かったが，IR2 においては 400 ms の時が一番高くなった。よって，仮説 2 は IR1 では支持されたが, IR2 では支持されなかった。これは IR2 においては，文節間ポーズ長による影響よりも帯域制限やクリップ処理の音声処理の影響が大きく及ぼした可能性も考えられる。今後，音声処理を施さない条件でも実験を行い，音声処理の影響を調べる必要がある。

4.3 モーラごとの正答率について

図 5 のようにモーラごとの正答率は IR1 と IR2 のどちらの条件おいても 2 モーラ目が一番高かった。ロングパスエコーによるターゲット語の重複の度合は疑似インパルスである IR1 においては明確であり, 計算してみると 2 モーラ目より他のモーラの方が重複部分が多いことがわかる。 2 モーラ目と一つ目の反射音が重複しているのは，調音速度 4 条件×文節間ポーズ長 3 条件の 12 条件中 7 条件のみである。1 モーラ目はキャリア文の「たんごは」と重複する条件が多く，3, 4 モーラ目はターゲット語自身の反射音と重複する条件が多い。よって IR1 において，2 モーラ目の正答率が高かったのは，ロングパスエコーが 2 モーラ目にエコーによる重複が少ない条件が多かったことが考えられる。このことからロングパスエコーが重複している部分が少ないモーラは大きく重複しているモーラより正答率が高くなることがわかる。これより，仮説 3 は支持された。

5　おわりに

本研究では，屋外拡声放送における発話速度と音声明瞭度の関係を調査するため，単語了解度試験を行った。

その結果，ロングパスエコーを模擬したインパルス応答では調音速度が 3 mora/s より 5 mora/s の方が，有意に単語了解度が高くなることが確認された。調音速度が低下することでオーバーラップマスキングが増え，単語了解度が低下することによるものと考えられる。

モーラごとの正答率では，2 モーラ目の正答率と他のモーラの正答率との間に有意差が見られた。今回の実験条件下では 2 モーラ目がエコーによる重複をあまり受けなかったことによるものだと考えられ，エコー成分による重複が少ない部分は明瞭性が高くなると考えられる。

今回の実験では文節間ポーズ長に関する有意な改善が見られなかった。これは本実験で施した帯域制限やクリップ処理といった音声処理による影響の方が大きかった可能性が考えられる。よって今後は音声処理の有無を含めた多くの条件を対象に実験を行いたい。

謝辞

本研究を進めるにあたり，詳細に渡る指導をいただき，また各種データをご提供いただきました TOA 株式会社の栗栖清浩氏に大変感謝申し上げます。

参考文献

[1] 戸井田義徳, 日音学誌, 43, 519-525, 1987.
[2] 内閣府（防災担当），東北地方太平洋沖地震を教訓とした地震・津波対策に関する専門調査会（第 7 回）.
[3] 佐藤洋他，日本建築学会計画系論文集，484，1-8，1996.
[4] 栗栖他，音響論（秋），1529-1532，2013.
[5] 川島他，音響論（秋），831-835，2012.
[6] 崔他，音響論（春），953-954，2013.
[7] 近藤他, IEICE, 思考と言語, 107, 43-48, 2007.