ロングパスエコーによるオーバーラップマスキングと発話速度の関係*

○荒井隆行（上智大・理工），大田敦也（上智大院・理工研），安啓一（国リハ研究所）

1　はじめに

残響が顕著な室内において拡声音声の明瞭度は低下することが知られているが，その背景としていくつかの要因が指摘されている．その 1 つがオーバーラップマスキング（以下，OLM）であり，先行する音声に伴う残響の「尾」が後続の音声をマスキングすることによるものである[1]．この OLM を減らす試みとして音声信号に対する前処理なども検討されているが（例えば[2-4]など），それに合わせて OLM 量の定式化[5]や，発話速度と明瞭度の関係[6,7]などについても議論がされている．

一方，屋外拡声においてはロングパスエコーといった孤立反射音が特徴的であるが[8]，同様にエコーによっても OLM は起きる．そこで，本研究ではロングパスエコーを伴う屋外の環境を模擬し, OLM 量の変化を観察しながら，発話速度が明瞭度に与える影響を調べた．

2　実験

本研究では先行研究[9]にならい，発話速度の調整に 2 つのパラメータを用いた．1 つは音声そのものの速度である 1 秒あたりのモーラ数（以後，調音速度），もう 1 つは音声の文節間に挿入する無音区間の長さである．以後，前者は AR（articulation rate の略, 単位は mora/s），後者は SI（silent interval の略，単位は ms）という記号で表すこととする．このように 2 つのパラメータによって発話速度を調節した音声信号に対し，ロングパスエコーを模擬したインパルス応答をたたみ込み，日本語を対象とした単語了解度試験を実験室環境にて行った．

2.1 音声資料

本研究では，親密度別単語了解度試験用音声データセット 2007 (FW07) [10] から，高親密度（7.0-5.5）と中低親密度（2.5-4.0）の日本語 4 モーラ語，各 36 個，合計 72 個を用いた．

上智大学荒井研究室の防音室において，選出された単語を男性 1 名が読み上げ，その音声を録音した．録音には，コンデンサ型マイクロホン（SONY, ECM-23F5）および, ディジタルレコーダ（Marantz, PMD660）を使用し，ディジタル録音（非圧縮，サンプリング周波数 44.1 kHz，量子化 16 bit）を行った. なお, 上記ターゲット語とは別に，後に用いる「これから流す単語は○○○○です」というキャリア文についても合わせて録音した．

2.2 発話速度の調整

上記の録音音声に対し，キャリア文とターゲット語を組み合わせ，さらに発話速度の調整を行った．発話速度の調整には，前述の調音速度 AR と文節間の無音区間長 SI を制御した．

AR については，調音速度は Praat [12]を用いて各発話区間の持続時間を変更した．そして，変更後の調音速度が AR = 4, 6, 8 mora/s になるように音声信号を準備した．

SI については，対象となる文において「これから__流す__単語は__○○○○__です」の「 __ 」で示された箇所に無音を挿入した．挿入された無音区間長は，SI = 0, 500, 1000 ms であった．

2.3 劣化処理

屋外拡声において，再生される音声の周波数帯域やダイナミックレンジが狭いことなどが原因で音質が劣化することが指摘されている[11]．このことから，聴取実験に用いる刺激音声を作成する際，劣化処理を施す場合と，施さない場合の両方を条件に加えた．

劣化処理は，具体的に 2 つのステップで行った．第 1 ステップはクリッピングで，過大振幅（実際には 100 倍）に対し±1 を

* Relation between speaking rate and overlap-masking by long-path echo, by ARAI, Takayuki, OTA, Atsuya (Sophia Univ.), and YASU, Keiichi (Research Institute of NRCD).

超える場合に振幅値を±1 にする処理を施した．第 2 ステップは帯域制限で，32 次の FIR フィルタにより電話帯域である 300-3400 Hz の帯域通過フィルタを実現し処理を施した．

2.4 ロングパスエコー

ロングパスエコーについては，擬似的に作成したインパルス応答を音声信号にたたみ込むことで実現した．実際の聴取実験では遅れが短いものと長いものの 2 種類を用いたが，本稿では遅れが短いものを中心に報告するため，まずは短いものについて説明する．

インパルス応答の作成には，直接音に対するインパルス（振幅は 1.0）の他，エコーに対応する振幅の小さい遅れをもったインパルスを複数並べた．具体的には，500 ms の遅れに対して振幅を 1/2，1000 ms の遅れに対して振幅を 1/3，1500 ms の遅れに対して振幅を 1/4，2000 ms の遅れに対して振幅を 1/5 とした．

上記の短い遅れの条件と同様に，長い遅れのインパルス応答も準備した．長いものについては，インパルスを立てる間隔を 2 倍の 1000 ms おきに設定した．

2.5 刺激音声

72 語をキャリア文に埋め込む際，AR と SI のそれぞれに 3 種類あるため，この時点で 648 文となる．さらに，劣化処理のあり・なし，およびロングパスエコー2 種類によって，刺激音声の総数は 2592 となった．

なお，劣化処理を行う条件では，劣化処理後，インパルス応答とのたたみ込みを行った．劣化処理を行わない条件では，そのままインパルス応答とたたみ込んだ．

ところで，シミュレーションによってロングパスエコーを模擬すると，一般に音声が比較的クリアに聞こえる．そこで，すべての刺激において，最終段階にてピンク雑音を重畳した．その際，信号対雑音比は 0 dB とした．

一人の実験参加者が同じ単語を一度しか聴取しないよう考慮し，AR 3 条件×SI 3 条件×劣化処理 2 条件×インパルス応答 2 条件の 36 条件に対して，高親密度語と中低親密度語を割り当てた．その際，一人の参加者に対し，劣化処理ありの条件で高親密度語 18 語と中低親密度語 18 語を，劣化処理なしの条件で残りの高親密度語 18 語と残りの中低親密度語 18 語を提示した．そして最終的に，総刺激の半分である 1296 刺激を使用し，参加者間でカウンタバランスをとった．

2.6 実験方法

参加者は日本語母語話者 21～25 歳（平均 22.9 歳）の健聴者 18 名（男性 8 名・女性 10 名）であった．健聴か否かは参加者の自己申告とした．実験は，上智大学荒井研究室内防音室にて行った．刺激音声は，PC から USB オーディオインターフェース (Roland, UA-25EX)を介してヘッドホンから提示した（diotic 受聴）．参加者は各刺激が流れるごとに，聞こえたと思う 4 モーラ語を平仮名で PC に入力した．

3 結果および考察

Table 1 に，高親密度語を対象に短い遅れのロングパスエコーに対する実験結果を示

Table 1: Word intelligibility (%) for each experimental condition.

条件 No.	SI [ms]	AR [mora/s]	SR [mora/s]	Overlap [mora]	正答率［%］	
					処理あり	処理なし
1	0	4	4	4	55.6	88.9
2	0	6	6	4	66.7	83.3
3	0	8	8	4	44.4	77.8
4	500	4	2.7	2	77.8	83.3
5	500	6	3.5	1	77.8	88.9
6	500	8	4.1	0	77.8	94.4
7	1000	4	2.1	2	88.9	88.9
8	1000	6	2.5	1	83.3	94.4
9	1000	8	2.8	0	83.3	88.9

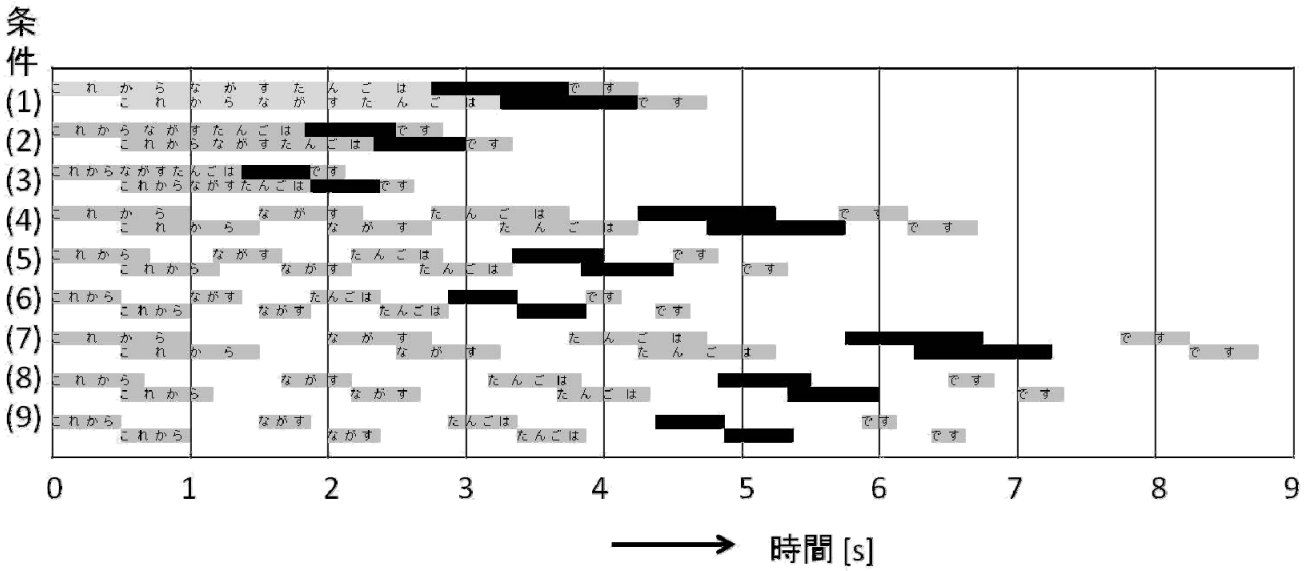

Fig. 1: 各条件における 17 モーラの時間配分.

す．この表において，一番右の 2 列が各語に対する正答率を平均したものである．ここで，処理あり・なしは劣化処理のあり・なしを意味している．この表を見ると，一般的な傾向として，文節間に挿入する無音区間の長さ(SI)が長くなるにしたがって，正答率が上昇する傾向があることが分かる．

一方, 同じ SI であっても, 調音速度（AR）が異なると正答率が変動する．その様子を考察するため, 2 つのパラメータを求めた．1 つめは, 1 文の全体の持続時間から計算した平均発話速度（speaking rate，略して SR）であり，単位は mora/s で表す．2 つめは，オーバーラップマスキング量に関わる指標（表中の Overlap [mora]）である．

以下では，2 つめの指標である Overlap について，その求め方を述べる．そのために，まず Fig. 1 を作図した．この図は，各条件ごとに 1 文 17 モーラがどのように時間配分されているかを図式化したものである（横軸が時間に対応）．さらに，各条件では同じものを 500 ms だけ遅らせたものを 2 段目に表示してある．そして，色が濃く表わされている部分がターゲット語である．これを見ると，直接音におけるターゲット語が，500 ms 遅れて到来する 1 つめのエコー音声と何モーラ分，オーバーラップしているかが分かる．その「オーバーラップしているモーラ数」をカウントしたのが，Table 1 に示した指標 Overlap [mora]である．なお，この場合，ターゲット語が 4 モーラ語であるため，Table 1 の Overlap の最大値は 4 である．

Table 1 における Overlap の値を見てみると，文節間に無音が挿入されていない場合は，常にオーバーラップされていることが分かる．一方，文節間に無音が挿入されると，オーバーラップが減少する様子が分かる．しかし，SI = 500 ms の場合と，1000 ms の場合を比較すると, 必ずしも SI が長いほうが Overlap 値が小さいわけではなく，調音速度やタイミングによって Overlap 値は変動する. Overlap 値と正答率を見比べると，大まかには Overlap 値が減少すると正答率が上昇する傾向が見られる．さらに詳しく見ると，文節間に適度な無音を挿入し，調音速度を速めて次のエコーが来る前に文節を言い切ってしまったほうが，結果としてオーバーラップが小さく抑えられるときがあり，その条件下での正答率が高い傾向があることも分かる．

SR と Overlap という 2 つのパラメータが，正答率とどのような相関関係にあるかを調べるため, 条件 4〜9 に対する散布図を Fig. 2(a)と(b)に示す．各図において，横軸はパラメータ（SR または Overlap），縦軸は劣化処理なしの結果に対する正答率である．これらの図を比べると, SR では同じ発話速度であっても異なる正答率を示す場合などがあるため，相関が弱い（Fig. 2a）．一方，Overlap ではターゲット語が実際にオーバーラップされている様子をより反映している結果，相関関係がよりはっきりと表れている様子がわかる（Fig. 2b）.

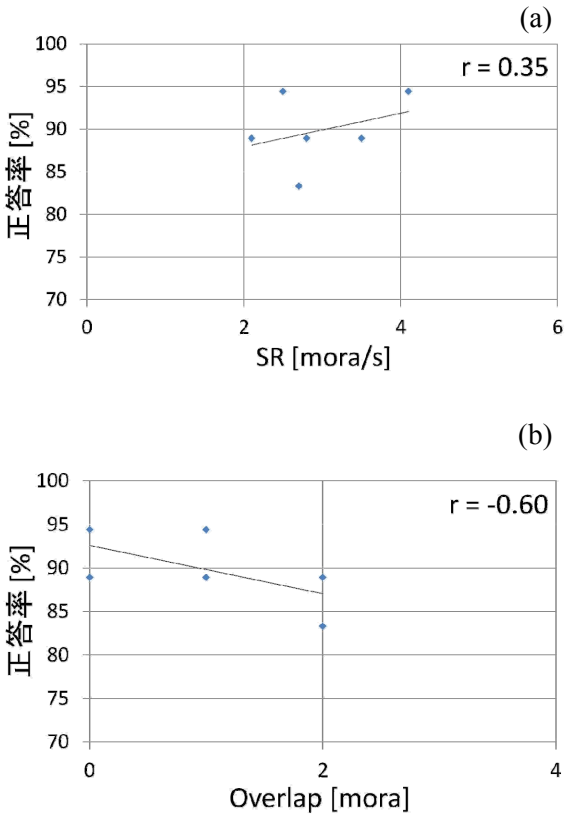

Fig. 2: 発話速度(a)および Overlap (b)と条件 4～9 に対する正答率（劣化処理なし）の関係

4　おわりに

本稿では，屋外拡声を想定しロングパスエコーを模擬したインパルス応答を用いて，発話速度の異なる音声に対する明瞭度を調べた．発話速度の変更には，調音速度，そして文節間に挿入する無音区間長の両方を変化させ，ターゲット語に対するオーバーラップマスキングを検討した．その結果，平均的な発話速度は明瞭度とは必ずしも相関が高くないことが分かった．これについて，文節間に適度な長さの無音を挿入した際，次のエコーが到来する前に直前の文節を言い切るために少し速めの調音速度で発話することの有効性が示唆された．

しかし，本研究での検討は，模擬的なロングパスエコーであること，1 つめのエコーのみのオーバーラップしか検討していないなど，限定的である．エコー音声自身に含まれるターゲット語のオーバーラップの状況の調査や，オーバーラップしたモーラ位置ごとの正答率の検討などは，今後の課題である．

謝辞

本研究にあたり，TOA（株）の栗栖清浩様に深く感謝申し上げます．

参考文献

[1] 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.

[2] 荒井隆行，木下慶介，程島奈緒，楠本亜希子，“音声の定常部抑圧の残響に対する効果,” 音講論（秋）, 1, 449-450, 2001.

[3] 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.*, 23(4), 229-232, 2002.

[4] T. Arai, “Padding zero into steady-state portions of speech as a preprocess for improving intelligibility in reverberant environments,” *Acoust. Sci. & Tech.*, 26(5), 459-461, 2005.

[5] T. Arai, Y. Murakami, N. Hayashi, N. Hodoshima and K. Kurisu, “Inverse correlation of intelligibility of speech in reverberation with the amount of overlap-masking,” *Acoust. Sci. & Tech.*, 28(6), 438-441, 2007.

[6] T. Arai, Y. Nakata, N. Hodoshima and K. Kurisu, “Decreasing speaking-rate with steady-state suppression to improve speech intelligibility in reverberant environments,” *Acoust. Sci. & Tech.*, 28(4), 282-285, 2007.

[7] 川島佑亮，荒井隆行，安啓一，“拡声音に対する零挿入による残響環境下での音声明瞭度改善の試み －挿入する位置と時間長の検討－,” 音講論（秋），831-835，2012.

[8] 戸井田義徳，“野外音場における明瞭度,” 日本音響学会誌, 43, 519-525, 1987.

[9] 大田敦也，荒井隆行，安啓一，“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,” 音講論（秋）, 183-186, 2014.

[10] 近藤公久，天野成昭，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 2007 (FW07) の作成,” 電子情報通信学会技術研究報告，思考と言語, 107, 43-48, 2007.

[11] 栗栖清浩，松本泰，山内昭弘，有賀成嘉，“屋外防災拡声システムの現状と課題,” 音講論（秋）, 1529-1532, 2013.