実時間処理を考慮に入れた サウンドマスキングシステムのためのマスカーの評価*

◎中嶋雄大，荒井隆行，安啓一（上智大・理工）

1　はじめに

サウンドマスキングは 1970 年代より多くの研究が行われており，従来の研究ではマスカーとして白色雑音やピンクノイズなどの定常雑音が用いられることが多かった[1]。近年の研究では，Ito *et al.* が音声を加工したマスカーを提案し，単語了解度試験の結果より，音声を加工したマスカーの方が定常雑音などの従来のマスカーよりもマスキング効率が高いことを報告している[2]。また，Ito *et al.* はターゲット音声と同じ話者の音声を加工したマスカーの方が異なった話者の音声を加工したマスカーよりもマスキング効率が高く，ターゲット話者の変更によりマスカーの性能が落ちてしまうということを指摘している[2]。このようなターゲット話者の変更によるマスキング効率減少の問題は，実時間処理を行うことにより対応できると考えられるが，音声を加工したマスカーの研究において，実時間処理に向けた検討の報告は多くはない。また，サウンドマスキングシステムの実用性を考慮すると，マスキング効率だけでなく，マスカーを呈示することによって生じる不快感などのアノイアンスや騒音環境としての側面を評価することも重要である[3]。

そこで，本研究では，サウンドマスキングシステムの実時間化を想定し，DSP（Digital Signal Processor）実装を考慮に入れた数種類のマスカーについて，単音節明瞭度試験によるマスキング効率の評価実験，および実験参加者によるアノイアンスの主観評価実験の 2 つの実験を行った。

2　本研究で用いたマスカー

本研究で用いたマスカー（Table 1）を以下に示す。

PINK : 従来の定常雑音のマスカーとしてピンクノイズ（以降，PINK ; NOISEX-92[4]）を用いた。

Table 1　本研究で用いたマスカー

マスカー	説明
PINK	・有色定常雑音
RAND [2]	・ターゲット音声を加工 ・フレーム（160 ms）毎に時間反転させて無作為に並び替えたものを 2 つ作成し，80 ms の時間差を持たせ足し合わせる
REV [5]	・ターゲット音声を加工 ・直前のフレーム（200 ms）を時間反転
MIX-RAND [6]	・RAND と PINK を足し合わせる
MIX-REV	・REV と PINK を足し合わせる

RAND : Ito *et al.* [2]により提案されたマスカーを参考にして RAND（Random Masker）を作成した。このマスカーは，複数のフレームを無作為に並べ替える処理があるため，DSP 実装を考慮した場合，ある程度の時間の音声をバッファに蓄える必要があり，最初にバッファの長さ分の遅延が生じる。本研究では，予め個々のターゲット音声の全区間に対してバッチ処理を行ったものを用い，最初の遅延はないものとした。

REV : Arai [5]により提案されたマスカーに従い REV（Reverse Masker）を作成した。このマスカー[5]は，DSP 実装を考慮した場合，最初に 1 フレーム分の時間（本研究では 200 ms）の遅延が生じる。また，前述の RAND の処理では複数のフレームをバッファリングする必要があったが，このマスカーの処理[5]ではバッファの長さは 1 フレーム分のみでよい。そのため，最初の遅延が 1 フレーム分の時間と短く，ターゲット話者の変更に瞬時に対応できる点が利点として挙げられる。

MIX-RAND : Ueno *et al.* [6]によって提案されたマスカーを参考にして，前述の RAND と PINK の音圧レベルが 1 : 1 になるように足し合わせたものを MIX-RAND（Mix Random Masker）として作成した。

MIX-REV : Ueno *et al.* [6]の処理を参考に，前

* Evaluation of maskers for sound masking system considering real-time processing by NAKAJIMA, Yudai, ARAI, Takayuki, and YASU, Keiichi (Sophia University).

述の REV と PINK の音圧レベルが 1:1 になるように足し合わせたものを提案し，MIX-REV （Mix Reverse Masker）として作成した。

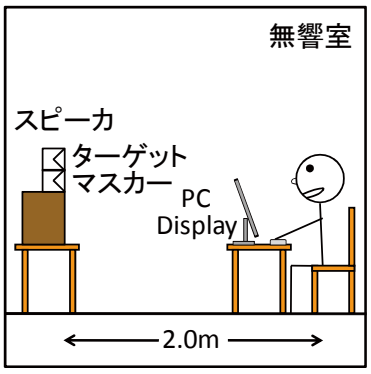


Fig. 1　実験環境

3　実験 1：単音節明瞭度試験による マスキング効率の評価実験

3.1 刺激

ターゲット：原音声に，日本語の単音節 CV（ATR 研究用日本音声データベース[7]）をキャリアセンテンス「題目としては__といいます」に挿入したものを用いた。V は/a/，C は/s/，/dz/，/tʃ/，/n/，/b/，/p/を用いた。呈示レベルは実験参加者の頭部中央で騒音レベル（A 特性）が 50 dB となるように設定した。

マスカー：第 2 節で述べた 5 種類のマスカーを用いた。呈示レベルは騒音レベル（A 特性）で 45，50，55，60，65 dB（T/M 比は-15，-10，-5，0，5 dB）の 5 条件とした。

3.2 実験参加者

日本語を母語とする 19～23 歳（平均 20.9 歳）の健聴者 24 名が実験に参加した（男 16 名，女 8 名）。健聴であるかどうかは自己申告とした。

3.3 実験手順

実験は無響室で行われた。Fig. 1 に実験環境を示す。刺激音は標本化周波数を 16 kHz とし，2 つのスピーカ（ONKYO D-312E）から，ターゲットとマスカーを各 1 チャンネルずつ呈示した。実験内容は，ターゲットとマスカーを同時に一度のみ呈示した後に，19 種類の選択肢（V を/a/とした単音節 18 種類と，その他）を PC ディスプレイ上に表示し，実験参加者にその中から一つを選択させた。ターゲット（6 種類）とマスカー（5 種類）の組み合わせである 30 刺激を 1 セットとして無作為に順番を入れ替えて呈示した。実験はマスカーの呈示レベルが高いセットから順番（65→60→55→50→45 dB の順）に行った。以上より，計 150 刺激に対して実験を行った。

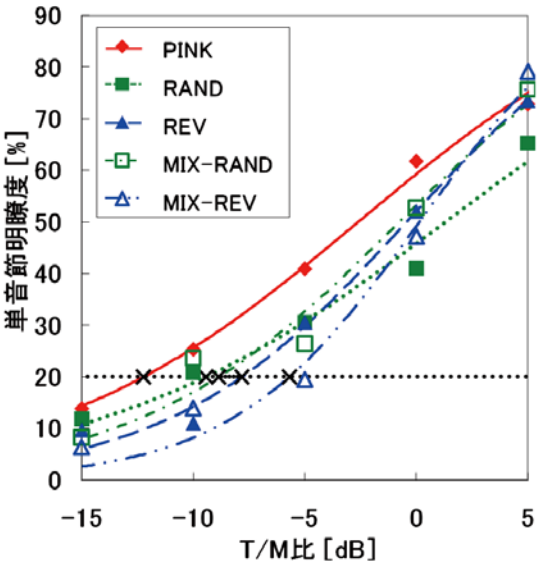


Fig. 2　マスキング効率評価実験の結果

3.4 結果・考察

Fig. 2 に，5 種類のマスカー毎の各 T/M 比における単音節明瞭度を示した。図中の曲線は，実験結果をシグモイド・ロジスティック回帰曲線によりフィッティングしたものである。

Fig. 2 から，音声を加工したマスカーである RAND や REV の方が，定常雑音である PINK よりも明瞭度が低くなっているため，マスキング効率が高いことが確認できる。多重比較検定を行った結果，PINK と RAND（$p < 0.01$），PINK と REV（$p < 0.05$）の間には各々有意差が見られた。RAND と REV の間には全条件で有意差が見られなかった。

また，Fig. 2 より T/M 比が-15 dB～0 dB の範囲で，MIX-REV の曲線は REV の曲線よりも下に位置しているため，MIX-REV は REV よりも明瞭度が低く，マスキング効率が高いと考えられる。このことから，音声を加工したマスカーに定常雑音を足し合わせるとマスキング効率が高くなるということが確認できた。しかし，MIX-RAND の曲線と RAND の曲線は T/M 比が-5 dB 付近で交差してしまっているため，この範囲では優劣を決め難い。このことから，音声を加工したマスカーに定常雑音を足し合わせるとマスキング効率が高くなるとは必ずしも限らないということも同様に確認された。多重比較検定を行った結果，PINK と MIX-RAND の間には有意差が見られなかったが，PINK と MIX-REV の間には有意差が見られた（$p < 0.01$）。以上の結果から，

Table 2　単音節明瞭度 20 %の際の
マスカー呈示レベル

マスカー	単音節明瞭度 20 %の際の マスカー呈示レベル[dB A]
PINK	62.22
RAND	59.44
REV	57.82
MIX-RAND	58.86
MIX-REV	55.69

Table 3　刺激音に対する質問内容

Q1. 計算問題を解くのに集中できましたか。(5 段階)
Q2. スピーカから聞こえてくる音はうるさかったですか。(4 段階)
Q3. 隣の部屋のプライバシーは保護されていましたか。(5 段階)
Q4. スピーカから聞こえてくる音はあなたが置かれている状況に適切でしたか。(5 段階)

定常雑音と足し合わせるマスカー（音声を加工する際の処理方法）によって定常雑音を足し合わせた際の効果（明瞭度の変化量）が変わってくるということが考えられる。

4　実験 2 : アノイアンスの主観評価実験

4.1 刺激

ターゲット：原音声に，男性 2 人の対話（内容：海外旅行計画について）の様子を録音した音声（RWCP 音声対話データベース[8]）を用い，30 秒の音声を 24 個切り出した。呈示レベルは騒音レベル（A 特性）で 50 dB とした。

マスカー:実験 1 と同じ 5 種類のマスカーに，マスカーなしの条件を加え,計 6 条件とした。各マスカーの呈示レベルは，Ueno *et al*. [6]にならい, 単音節明瞭度が 20 %となる際 (Fig. 2 の横線，Table 2）の音圧レベルとし，実験 1 の結果の曲線より各々求めた。

4.2 実験参加者

実験 1 の参加者と同様である。

4.3 実験手順

実験環境や，刺激音の標本化周波数は実験 1 と同様である。実験内容は，まず実験参加者に「学校の教室で自習をしていて，隣の部屋では進路相談のような他者に漏れてはならない会話をしている」という状況をイメージさせた後に，ターゲットとマスカーを同時に 30 秒間呈示し，実験参加者に簡単な計算問題（2 桁 + 2 桁）を解かせた。その際，実験参加者には刺激音が流れ始めたら問題を解き始め，音が止まったら問題を解くのを止めるように指示した。そして，問題を解くのを止めた後に，刺激音に対する質問内容（Table 3）に答えさせ，アノイアンスを主観的に評価させた。以上の試行を，カウンターバランスを考慮してマスカーとターゲットの組み合わせたものを,無作為に順番を入れ替えて 24 回繰り返した。

アノイアンスは，Ueno *et al*. [6]にならい，各々Q1～Q4 において，“計算問題への集中度”，“うるささ”，“プライバシー”，“適切さ”の 4 つの視点から評価した。Q3 の “プライバシー” は，隣の部屋の話声（ターゲット音声）の内容がどの程度聞こえるか, Q4 の“適切さ”は，自分が置かれている状況に対して，自分がその場にいたら刺激音を適切と思うかどうかという質問である。

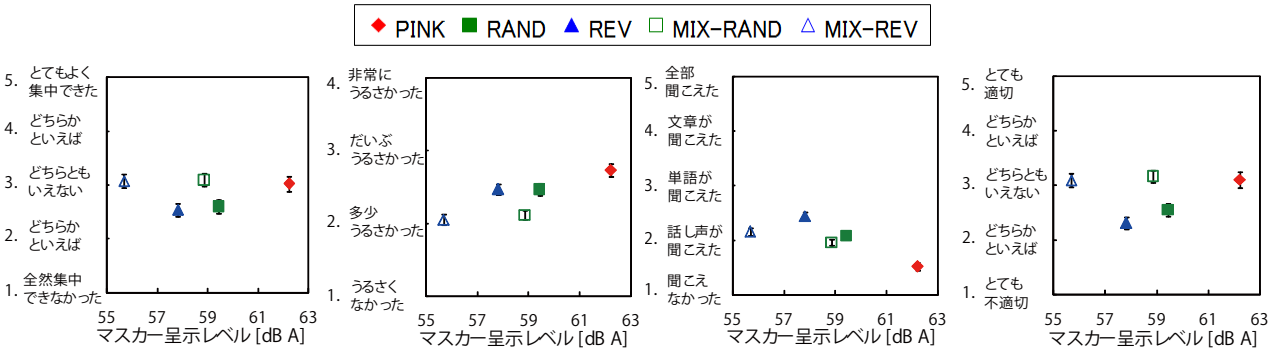
4.4 結果・考察

5 種類のマスカー毎の各質問における全実験参加者の平均値を求め，結果を Fig. 3 に示した。なお，各マスカーにおける結果の標準誤差を求め，誤差範囲を図中のエラーバーで示した。

Fig. 3 より，各質問において MIX-RAND と MIX- REV は, 各々RAND と REV よりも良い結果となっていることが確認できる。分散分析の結果，Q1 : 集中度，Q2 : うるささ，Q4 : 適切さについて，MIX-RAND と RAND の間と，MIX-REV と REV の間に各々有意差が見られた。

そして，Fig. 3(a)，(d)より，Q1 および Q4 について， PINK の結果は，MIX-RAND と MIX-REV の結果と各々ほぼ等しい結果となり，RAND や REV に比べて良い結果となったことが確認できる。このことより，マスカーの特性を PINK のような定常雑音に近付けることにより，集中度および適切さが高くなる傾向にあることが確認できた。

Fig. 3(c)より, Q3 : プライバシーについて，PINK の結果が他のマスカーに比べて良い結果となったことが確認できる。このことより，単音節明瞭度が等しい（20 %）際でも，文の内容が聞き取れるかの主観的判断は異なって



（a）Q1：集中度　（b）Q2：うるささ　（c）Q3：プライバシー　（d）Q4：適切さ

Fig. 3　主観評価実験の結果

くるということが確認できた。また，Q3 とマスカー呈示レベルの間の相関係数が r = -0.783（p = 0.12）であり，負の相関関係が見られた。このことより，文の内容が聞き取れるかの主観的判断は，マスカー呈示レベルの高さに影響されるということが考えられる。

さらに，マスカーの呈示レベルと各質問の結果の間の相関を求めた結果，Q2：うるささとマスカーの呈示レベルの相関係数が r = 0.811（p = 0.09）であり，正の相関関係が見られた。このことより，マスカーの呈示レベルを低くすることにより，うるささを減少させることができると考えられる。

各質問間の相関関係を求めた結果，Q3 と Q4 の相関係数が r = -0.925（p < 0.01）であり，強い負の相関関係が見られた。このことより，実験参加者は Q4：適切さについて Q3：プライバシーを基準に評価していたということが考えられる。一方，Q1：集中度と計算問題の得点の間に相関関係が見られた場合，計算問題の得点を集中度の客観的指標として利用できると想定したが，相関関係は見られなかった。

5　まとめ

本研究では，DSP 実装（実時間処理）を考慮に入れた数種類のマスカーを作成し，各マスカーについてマスキング効率の評価実験，およびアノイアンスの主観評価実験を行った。単音節明瞭度試験によるマスキング効率評価実験の結果，5 種類のマスカーの中で MIX-REV のマスキング効率が一番高いことが示された。また，実験参加者によるアノイアンスの主観評価実験の結果，各条件で MIX-RAND と MIX-REV がよい結果であった。これらの結果からマスキング効率とアノイアンスの両条件において，MIX-REV の優位性が示された。

今後の課題としては，本研究で用いたマスカーを DSP に実装し，DSP に実装したマスカーの評価をすることや，よりマスキング効率が高く，かつアノイアンスの程度も低いマスカーを提案することが挙げられる。また，アノイアンスの評価方法として，本研究では実験参加者に主観的に評価させていたため，実験参加者が何を基準に評価しているのかが明確ではなく，実験参加者毎に結果に影響が出てしまった可能性がある。よって，今後は主観評価の基準や客観的指標などを明確に定める必要があると考えられる。

謝辞

本研究の一部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「人間情報科学研究プロジェクト」の支援を受けて行われた。

参考文献

[1] A. C. C. Warnock *et al.*, J. Acoust. Soc. Am., 53 (6), 1535-1543, 1973.

[2] A. Ito *et al.*, Proc. INTER-NOISE 2007.

[3] K. Ueno *et al.*, Proc. INTER-NOISE 2007.

[4] A. Varga *et al.*, Speech Commun.,12(3), 247-251, 1993.

[5] T. Arai, Acoust. Sci. Tech., (to appear).

[6] K. Ueno *et al.*, *Proc. Acoustic'08 Paris.*

[7] 武田ら，日本音響学会誌, 44 (10), 747-754, 1988.

[8] RWCP 音声対話データベース，技術研究組合, 1996.