サウンドマスキングシステムにおける音声マスカーで使用する語彙の選定法による影響*

☆岩波達也，荒井隆行，安啓一（上智大・理工）

1　はじめに

現在スピーチプライバシー保護のための手段として，サウンドマスキングシステムが注目されている[1]。サウンドマスキングで使用されるマスカーは，そのマスキング効果から大きく分けて 2 つに大別される。1 つは，エネルギーマスキングによるものであり，その代表的なものにピンクノイズなどがある[2]。もう 1 つは情報マスキングによるものであり，その代表的なものに無意味音声によるマスカーがある[3]。この後者の場合，エネルギーマスキングだけでは説明できないそれ以上のマスキング効果をもたらす。情報マスキングで使用するマスカーは，主に人の音声を加工することで作られる。無意味音声を用いた場合，そのマスキング効率は場面によって差は生じず，常に同程度の効果をもたらすものと考えられる[3]。逆に，本来の音声が持つ意味を損なわずに作られたような有意味音声によるマスカーの場合，場面によってマスキング効率が上がるものと考えられる。

本報告では，有意味音声で使用するマスカーの語彙がマスキング対象となる会話(以後，ターゲット）の内容と類似するか否かによって，マスキング効率がどう変化するかを検討した。実験では，「特定の場面で使用される頻度が高い語彙から成るマスカー」(以後，scene dependent（略して SD）マスカー）と，「特定の場面で使用される頻度が必ずしも高くない語彙から成るマスカー」（以後，scene independent（略して SI）マスカー）の，2 種類のマスカーを作成し，比較を行った。加えて，従来法であるピンクノイズ，無意味音声のマスカーとの比較も行った。

2　実験方法

2.1 実験刺激

Ito *et al.*（2007）[3]は，ターゲットと同じ話者の音声を加工したマスカーの方が，異なる話者によるマスカーよりもマスキング効率が高くなることを示している。そこでターゲットとマスカーを足し合わせて刺激音を作成する際に，SD マスカー，SI マスカー共に，ターゲット音声と同じ話者の音声を加工して作成した。また Ito *et al.*（2007）を参考に，ターゲットを時間反転することによって無意味音声に加工したマスカー（以後，nonsense マスカー）も合わせて作成した。なお，刺激音の作成には MATLAB を使用した。

a)　ターゲット

サウンドマスキングシステムが応用される状況を考えると，実験で用いる刺激音声は実環境での対話である方が良いと考え，原音声として，RWCP 音声対話データベース[4]に含まれる音声サンプルをターゲットに選んだ。その中の旅行代理店において顧客と従業員が対話している様子を録音した音声（5～6 分）を対象としたが，そこには従業員 1 人に対して 12 人の顧客が順番に相談している際の音声が収められている。ターゲットは 12 人分の会話から約 16 秒の従業員の音声を 20 個抽出した。呈示レベルは騒音レベル（A 特性）で 50 dB とした。サンプリング周波数は 16 kHz，量子化ビット数は 16 bit であった。

b)　SD マスカー

ターゲットと同様に SD マスカーも従業員の音声から抽出した。音声コーパスの中に対話を文字に書き起こしたテキストファイルが収録されているので，そのファイルに対し，形態素解析ソフト「茶筅」[5]を使用して出現頻度の高い単語を頻度順に並び替えた。頻度 5 以上の単語を旅行代理店でよく使用される単語とし，従業員が発話した単語の音声を，

* Effects of vocabulary selection on speech masker in sound masking system, by IWANAMI, Tatsuya, ARAI, Takayuki and YASU, Keiichi (Sophia University).

全部で 53 語抽出した。なお，ターゲットとして使用されているものは除外した。

1 つの SD マスカーは次のように作成した。まず 53 語から 1 語ずつランダムに選択し，選ばれた単語に対する音声信号を連結していった。そして，ターゲットの長さに合わせて約 16 秒になるまで連結を続けた。1 つのマスカー内では同じ単語が 2 度と選ばれないようにした。以上の作業を 20 回繰り返し，ターゲットと同数の 20 個の SD マスカーを作成した。

なお，抽出した単語を連結するときに単語と単語の間で不連続が生じないよう，零交差点で波形を切り出した。

c) SI マスカー

ターゲットと同様に SI マスカーも従業員の音声から抽出した。SD マスカーと同じく，「茶筅」を使用し，頻度順に並び替えた単語リストから，使用頻度 1 の単語を「普段旅行代理店で使用される頻度が必ずしも高くない単語」として，全部で 50 語抽出した。作成方法は SD マスカーと同様とし，約 16 秒のマスカーを 20 個作成した。

d) nonsense マスカー

nonsense マスカーはターゲットを無意味音声に加工したマスカーである[3]。1 つの nonsense マスカーは次のように作成した。フレームの長さを 160 ms として，まずターゲットをフレームに分割した。次にフレーム毎に時間反転処理を行い，反転後のフレームをランダムに並び替えた。同じ処理を 2 回行い，異なる 2 つの信号を作成する。この時 2 つの信号はそれぞれランダムに並び替えるため，両者は異なるものとなっている。最後に，一方を 80 ms ずらして 2 つを加算した。以上の作業をターゲット 20 個の 1 つ 1 つに対して行い，各ターゲットに対応する nonsense マスカーを 20 個作成した。

e) ピンクノイズ

ピンクノイズとして，MATLAB で作成した 1 種類のものを使用した。

2.2 実験条件

a) 場面設定

ターゲットとして，旅行代理店の窓口で海外旅行計画について相談している場面の対話を用いたため，実験参加者には窓口のそばで順番を待っているという場面を想定してもらった。

b) 呈示方法

実験は防音室で行った。Fig. 1 に実験環境を示す。刺激音は USB オーディオインタフェース（Roland UA-25EX）の音声出力からアンプ内蔵型スピーカ（YAMAHA MSP3）を 1 つ用いて呈示した。ターゲットとマスカーは，事前に加算されたものをスピーカからモノラルで呈示した。実験参加者は，スピーカから 1.8 m の位置に着席してもらった。また，スピーカの位置は着席時の頭の高さに合わせ，1.2 m とした。

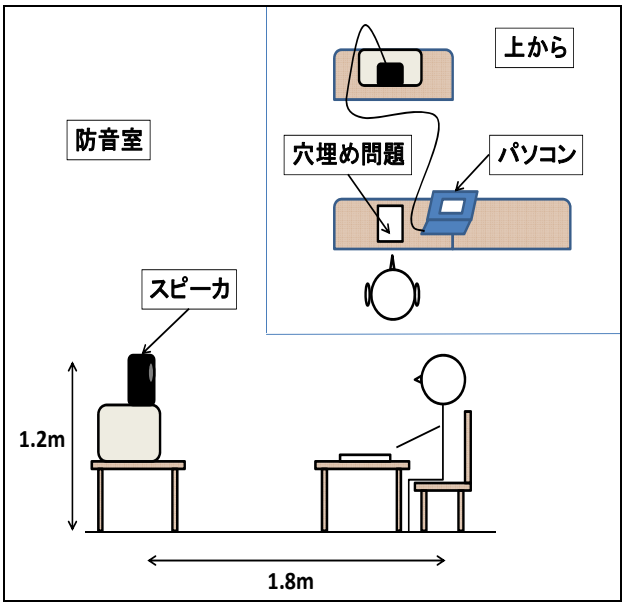

Fig. 1 実験環境

c) マスカー呈示レベル

今回実験で使用した刺激音は，異なるターゲットレベル対マスカーレベル比（以後，TMR）になるよう，予めターゲットとマスカーが加算されたものを用いた。ターゲットは常に 50 dB で呈示されるようにしたため，TMR 毎にマスカーの呈示レベルは変化した。TMR は－15，－10，－5，0，5 dB の 5 条件としたので，結局，マスカーの呈示レベルは騒音レベル（A 特性）でそれぞれ 65，60，55，50，45 dB の 5 段階となる。なお，実験は TMR の小さい順から順番に行った。

d)　実験参加者と回答方法

日本語を母語とする 22~24 歳（平均 22.7 歳）の健聴者 20 名が実験に参加した。実験参加者にはターゲット音声を書き出した文に対して穴埋め問題の書き取り試験を行ってもらい，回答用紙の穴の空いている箇所に聞こえた通りに記入させた。実験参加者が行う穴埋め問題は，1 人につき 20 問であり，1 問につき刺激音を 2 回再生した。なお，1 回目を再生した後，記入する時間を与え，2 回目は実験参加者のタイミングで再生できるようにした。実際に実験で使用した問題の一例を示す。

[問題例]

ノースウェストとコンチネンタルとジャル、が飛んでるんですけども、
[　　]がやはり[　　]に比べたら少ないのと，
また、[　　]で、[　　]のほうがまた[　　]できるかどうか

e)　評価方法

書き取らせた単語の正答率から Ueno *et al.* [7]を参考にして，シグモイド・ロジスティック回帰曲線で近似し，各マスカーのマスキング効率について比較した。

3　実験結果・考察

各実験条件（マスカー4 種類，TMR 5 種類）における単語了解度を全実験参加者の正答率の平均値として求めた。平均正答率に対し，シグモイド・ロジスティック回帰曲線で近似した結果を Fig. 2 に示す。

Fig. 2 から SD マスカー（赤の実線）が SI マスカー（緑の一点鎖線）よりも常に右に位置していることが分かる。これは同じ TMR で比較した時に常に SD マスカーの方が，SI マスカーよりも正答率が悪かったことを表している。つまり，SD マスカーの方が SI マスカーよりもマスキング効率が高いことが示された。特定の場面で使われる語彙がマスカーとして流れてくると，それがターゲットの一部であるか否かの判断がしにくくなることによって，ターゲットが聞き取りづらくなり，2 つのマスカーに差が生じたものと考えられる。以上のことから，マスカーで使用される語彙の選定法による影響は確かに存在し，語彙によってマスキング効率に差が生じることが確認できた。

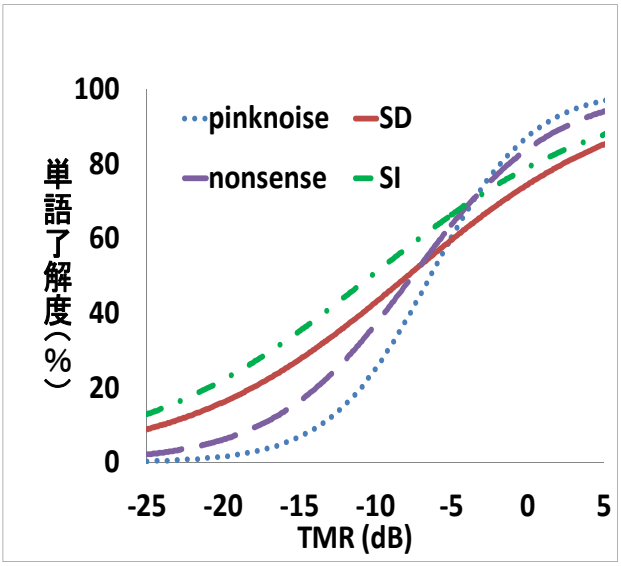

Fig. 2 各マスカーにおける TMR に対する単語了解度

また，従来法としてピンクノイズ（青の点線）と nonsense マスカー（紫の長破線）に対し，SD マスカーを比較した。Fig. 2 において単語了解度が低い時（40 %以下）について注目すると，従来法のピンクノイズ，nonsense マスカーに比べ，提案法の SD マスカーと SI マスカーでは TMR が下がっても正答率が高く，会話が聞こえてしまっていることを示している。つまり，従来法よりも提案法の方が，マスキング効率が低いということが分かる。この差については，マスカーの特性に原因があると考えられる。ピンクノイズは周波数特性もピーク性も低く，また時間包絡の変化も小さい。また，nonsense マスカーは 160 ms という短いフレームをランダムに並び替えたものを 2 つ用意し，一方を 80 ms ずらしてから加算しているので，やはり周波数スペクトルと時間包絡の変化が小さくなる。その結果，これらのマスカーではターゲットを全体的にマスクすることができる。しかし，SD マスカーと SI マスカーはもともと単語という長い単位の音声を用いているため，話者の発話した音声信号を時間・周波数平面上で全体を効率よくマスクできない。その結果，ターゲットの一部の情報が漏れてしまうことが考えられる。提案法のマスキング効率を高める方法として，定常雑音や環境音などを足し合わ

せる[7]ことで現在よりもマスカーの性能が上がることが期待できる。また，nonsense マスカーの生成方法のようにマスカー用の信号を 2 つ作り，それらを 1/2 フレームずらして足し合わせることで，マスキング効率が高くなることも期待できる。

藤原ら[8]は，実環境下における薬局窓口の等価騒音レベル（A 特性）が 50 dB～60 dB ほどで， また 56 dBA 程度のマスカー呈示レベルであれば，実際の薬局の環境下でマスキング効果が期待でき，かつアノイアンスを感じさせないことを確認したと報告している。また，李ら[9]の先行研究では，薬局利用者の 50 %が「プライバシーがだいぶ守られた」と判断するためには, 文章了解度は 60 %以下となる必要があると報告している。文章了解度は単語の説明を単語の前に配した短文を呈示し，書き取らせた単語の正答率を算出したものであり, 本実験の試験内容と類似している。そこで単語了解度が 60 %のとき，Fig. 2 を見るといずれのマスカーも，ほぼ同程度のマスキング効率を示しており，マスカーの呈示レベルも先行研究の 56 dBA 程度であった。したがって，今回提案したマスカーは実環境下を想定した単語了解度 60 %の場合, 従来法と同程度の効果が期待できるが，単語了解度を低く設定した場合には従来法よりもマスキング性能は下がるため，より効果的なマスカーの検討を行ってゆく必要がある。

Ueno *et al.* [7] はターゲットを加工した無意味音声によるマスカーの方が定常雑音よりもマスキング効率が高いということを示していたが，本実験の結果では，マスキング効率は必ずしも上回るという結果にはならなかった。理由として，従来のマスキング効率の評価方法は単語をターゲットとした明瞭度試験であるのに対し，本報告は会話文をターゲットとした単語了解度試験であるため，先行研究と異なる結果になった可能性も考えられる。今後，マスカーの見直しに合わせ，より実用的な環境に近い実験方法の検討が求められる。

4　まとめ

有意味音による音声マスカーの語彙が，マスキング対象となるターゲット音声の内容に類似するか否かによって，マスキング効率がどう変化するかを比較，検討した。単語了解度試験の結果，ターゲットと類似した場面で使用される頻度の高い語彙から成るマスカーの方が，マスキング効率が高くなることを確認した。実用化を考えると，低アノイアンスかつ高マスキング効率のマスカーが求められるため，アノイアンスの検討も行わなければならない。今後の課題として，アノイアンスの評価実験，異なる話者でも同様に語彙によるマスキング効率の差が生まれるかどうかの確認等も行う必要がある。

謝辞

本研究を進めるにあたり，実験参加者として協力してくださった方々に感謝いたします。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] Mapp, J. Acoust. Soc. Am., vol. 121, 3035-3036, 2007.
[2] Mellor, ICASSP, vol. 2, 87-90, 1993.
[3] Ito *et al.*, Proc. INTER-NOISE, 2007.
[4] RWCP 音声対話データデース，技術研究組合，1996.
[5] “専門用語（キーワード）自動抽出システム”，入手先 <http://gensen.dl.itc.u-tokyo.ac.jp/> (参照 2012-01-17).
[6] 藤原他，音講論（春），1075-1076，2009.
[7] Ueno *et al.*, Proc. INTER-NOISE, 2007.
[8] 藤原他，音講論（秋），1127-1130，2011.
[9] 李他，音講論（秋），1131-1134，2011.