母音部または子音部に対して選択的にエネルギーマスキングを施した音声の単語了解度*

☆澤井駿，荒井隆行，安啓一（上智大・理工）

1　はじめに

人が音声を認識する際には，様々な情報が重要とされている。例えば, 音声の子音部は，音響的特徴は様々であるが，エネルギーは小さい割に音声知覚に関する情報は多く含まれているとされている（例えば，[1]）。また，音声の母音部と子音部の振幅比によって，音声が聞きやすくなることや，逆に聞き取りにくくなることが指摘されている[2]。このように，音声の母音部・子音部にそれぞれ音声知覚に関する情報が含まれている。その結果，音声の母音部・子音部に対して選択的にエネルギーマスキングを施すことによって，人の音声認識に対して影響を及ぼすことが考えられる。そこで本研究では，スピーチプライバシーの保護を実現する手法であるサウンドマスキングにおいて，マスカーに関して母音・子音を考慮することを考えた。

サウンドマスキングに求められる重要な点として，マスカーが周りの人にうるさくなく不快に感じさせないことが挙げられる。マスカーとして雑音を使用する場合，レベルが低いほど，アノイアンスが低くなることが報告されている[3]。したがって，低いレベルの雑音をマスカーとして使用できれば，より良いシステムが実現できると考えられる。

そこで，本研究では従来の定常雑音マスカーに加え，定常雑音よりも低いエネルギー量の，音声の母音部を選択的にマスクするようなマスカーと，子音部を選択的にマスクするようなマスカーを使用した。そして，これらのマスカーを使用することによってターゲット音声の単語了解度を調査した。

2　本研究に用いたマスカー

2.1 ALL マスカー

ターゲット音の全体にわたってレベルが変化しない定常雑音として，ピンクノイズとホワイトノイズを用いた。ピンクノイズを用いたマスカーを PINK_ALL，ホワイトノイズを用いたマスカーを WHITE_ALL とした。

2.2 V マスカー

V マスカーは，ターゲット音の母音部を選択的にマスキングするマスカーである。V マスカーでは，ターゲットの母音部の振幅は前述の ALL マスカーの振幅と同じに保った(その振幅レベルを基準の 0 dB とする)。一方，子音部では振幅レベルを 3 dB，6 dB，10 dB の 3 段階で下げた（それぞれ V3，V6，V10 と呼ぶ)。実験では V マスカーを，ピンクノイズ，ホワイトノイズを使用して作成した。その結果，ピンクノイズを用いた V マスカーとして PINK_V3，PINK_V6，PINK_V10，ホワイトノイズを用いた WHITE_V3，WHITE_V6，WHITE_V10 の計 6 種類を実験に使用した。なお，子音と母音の境界である母音遷移部では，直線補間によってマスカーのレベルを変化させた。

2.3 C マスカー

C マスカーは，ターゲット音の子音部を選択的に隠すマスカーである。C マスカーでは V マスカーとは逆に，ターゲットの子音部の振幅を 0 dB に保つ一方, 母音部で振幅レベルを 3 dB，6 dB，10 dB の 3 段階で下げた（それぞれ C3，C6，C10 と呼ぶ)。同様に，ピンクノイズ，ホワイトノイズを使用した結果，ピンクノイズを用いた C マスカーとして PINK_C3，PINK_C6,

---

* Word intelligibility of speech with energetic masking on vowel or consonant, by SAWAI, Shun, ARAI, Takayuki and YASU, Keiichi (Graduate School of Science and Technology, Sophia University)。

マスカーの名称	説明	
ALL	従来の定常雑音マスカー	全体にわたってレベルの変化なし
V3	ALL と同じレベルでターゲットの母音部をマスクし，子音部は ALL より低いレベルでマスクするマスカー	子音部で 3 dB レベルを下げる
V6		子音部で 6 dB レベルを下げる
V10		子音部で 10 dB レベルを下げる
C3	ALL と同じレベルでターゲットの子音部をマスクし，母音部は ALL より低いレベルでマスクするマスカー	母音部で 3 dB レベルを下げる
C6		母音部で 6 dB レベルを下げる
C10		母音部で 10 dB レベルを下げる

Table 1 マスカーの種類(それぞれを PINK と WHITE で作成したため計 7×2=14 種類)

PINK_C10，ホワイトノイズを用いた WHITE_C3，WHITE_C6，WHITE_C10 の計 6 種類を実験に使用した。なお，母音遷移部では，直線補間によってマスカーのレベルを変化させた。本研究で用いたマスカーの種類をまとめたものを Table 1 に示す。

3 実験

3.1 刺激音

ターゲット音声は，辻ら[2]が作成した音声を使用した。その音声のもとになった原音声は，親密度別単語了解度試験用音声データベース(FW03) [4]から選ばれたもので，単語親密度 5.5 ~ 4.0 の日本語 4 モーラ単語の 42 単語である。これを音声合成ソフトウェア（アルカディア SPeeCAN SFT5）を用いて合成された女性話者による音声を使用した。そのターゲット音声は，辻らによって，時間波形やスペクトログラム上におけるフォルマントなどの周波数特性の特徴から，子音部・母音部・母音遷移部の 3 区間に分けられていた。なお，標本化周波数は 16 kHz であった。

刺激音は，ターゲットとマスカーを予め MATLAB を用いて加算したモノラル音を呈示するものとした。ターゲットの提示レベルを固定し，ALL マスカーの呈示レベルを 3 段階のターゲット対マスカー比(以後，TMR)になるように設定した。ターゲットの呈示レベルは，実験参加者の頭部中央で騒音レベル(A 特性)が 50 dB となるように騒音計（リオン・精密騒音計 NL-32）で調整した。

3.2 実験参加者

日本語を母語とする 18 ~ 24 歳（平均 22.1 歳）の男性 30 名，女性 12 名の健聴者（自己申告による）42 名が実験に参加した。

3.3 実験手順

実験は防音室で行われた。実験環境を Fig. 1 に示し，以下に詳細について述べる。実験参加者は，高さ 1.0 m の台に設置されたスピーカ (YAMAHA MSP-3) から 1.8 m 離れた位置に着席し，スピーカから提示される刺激音を聴取した。各刺激音の提示は 1 回のみとした。刺激音提示毎に，実験参加者が聴取したターゲットを PC ディスプレイ上のボックスにひらがなでタイピングしてもらった。42 条件（マスカー14 条件×ALL マスカーの TMR3 条件）に対して 1 条件につき 1 単語を使用したため，実験参加者は 42 刺激を聴取した。また，ターゲットと処理条件の組み合わせについては，ALL マスカーの TMR 毎に 14 単語（マスカー14 条件×1 単語）を割り当てることで，TMR 毎にカウンターバランスをとり，各単語がすべてのマスカー条件において出現するようにした。刺激の呈示順は ALL マスカーの TMR 毎にランダムに並べ替えた。また，ALL マスカーの TMR の小さい方から，−10 dB，−5 dB，0 dB の順番で実験を行った。ただし，V マスカーと，C マスカーは 42 単語ごとにマスカーのエネルギーが異なるため，それぞれ TMR は異なる。

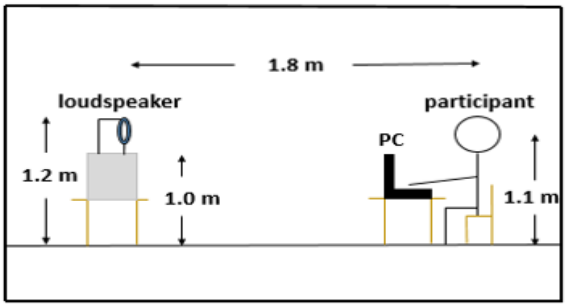

Fig. 1 実験環境

4 単語了解度試験の結果

Fig. 2 に，ピンクノイズを使用したマスカーの TMR に対する単語了解度を示し，Fig. 3 にホワイトノイズを使用したマスカーの TMR に対する単語了解度を示す。ここで単語了解度は，各刺激に対する正解を 100%，不正解を 0%として，全参加者の結果を平均したものである。なお，V マスカーと C マスカーは，単語毎にマスカーのレベルが異なるが，42 個のデータ点に対してロジスティク関数による回帰分析を施し，曲線を求めた。

また，Table 2 に単語了解度が 30%の際のピンクノイズ，ホワイトノイズを使用したときのマスカー呈示レベルを示す。

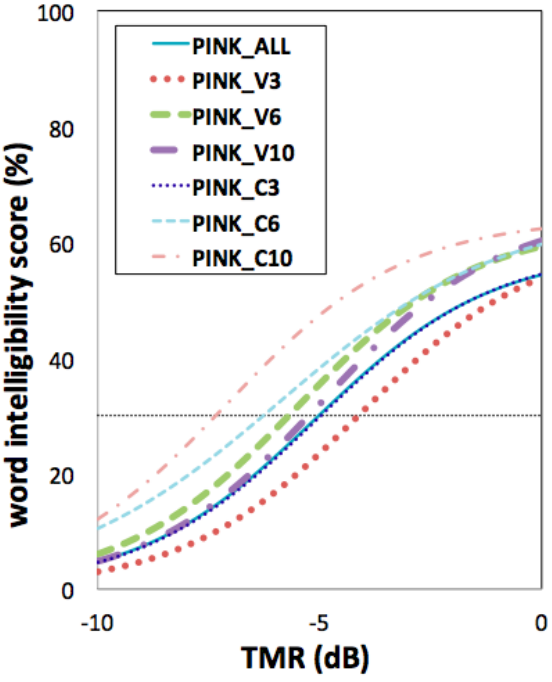

Fig. 2 ピンクノイズを使用したときの単語了解度試験結果（横軸に平行な点線は単語了解度 30%を示す）

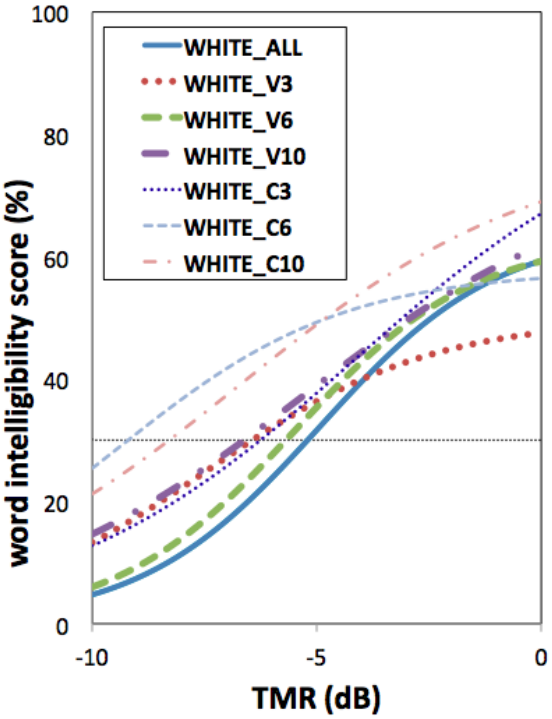

Fig. 3 ホワイトノイズを使用したときの単語了解度試験結果（横軸に平行な点線は単語了解度 30%を示す）

Table 2 単語了解度が 30%のマスカー呈示レベル

マスカーの種類		単語了解度 30%のマスカー呈示レベル[dB]
PINK	_ALL	55.0
	_V3	54.1
	_V6	55.7
	_V10	55.2
	_C3	55.0
	_C6	56.3
	_C10	57.3
WHITE	_ALL	55.2
	_V3	56.4
	_V6	55.7
	_V10	56.6
	_C3	56.3
	_C6	59.2
	_C10	58.3

5 考察

5.1 単語了解度試験結果の考察

Table 2 より， PINK_V3 は PINK_ALL よりも少ないマスカー呈示レベルで同じ単語了解度であるため，PINK_V3 は PINK_ALL よりもマスキングの効率が良い傾向がみられた。

この結果より，ピンクノイズではマスカーの振幅を子音部で 3 dB 低下させることは, ALL マスカーよりもマスキングの効率が上げられることが示唆された。

また，Fig. 2，Table 2 より，母音部または子音部で，レベルを 6 dB，10 dB 下げた場合(PINK_V6, V10, C6, C10)では，PINK_ALL, PINK_V3, PINK_C3 より単語了解度が高くなった。つまり，マスカーのレベルを 6 dB，10 dB 下げることは, 音声知覚に必要な情報が実験参加者に聴取されてしまうことが考えられる。

Table 2 より，WHITE_ALL が他のマスカーよりも少ないマスカー呈示レベルで同じ単語了解度であるため，WHITE_ALL が他のマスカーよりマスキングの効率が良い傾向がみられた。この結果より，ホワイトノイズではマスカーの振幅を，母音部または子音部で低下させることは，ALL マスカーよりもマスキングの効率が下がることが示唆された。

また，アノイアンスの試験を行った場合を考えると，母音部または子音部でマスカーの振幅を低下させる程，マスカーのエネルギーが少なくなるため，アノイアンスも低下することが考えられる。そのため，V マスカーと C マスカーは，ALL マスカーよりもアノイアンスが低下することが考えられる。しかし，振幅の変動がアノイアンスの上昇につながる可能性もあるため，実際にアノイアンスの試験を行う必要がある。

以上より，子音部でマスカーの振幅を低下させることは，そのレベルの下げる程度によっては ALL マスカーよりもマスキングの効率が上げられるということが分かった。

5.2 V マスカーと C マスカー比較

Table 2 より，PINK_V3 と PINK_C3，PINK_V6 と PINK_C6，PINK_V10 と PINK_C10 同士をそれぞれ比較すると，V マスカーはCマスカーよりも少ない呈示レベルで同じ単語了解度であるため，V マスカーは C マスカーよりもマスキングの効率が良い傾向がみられた。この結果は，人は音声の子音部よりも，母音部をマスキングすることで，音声知覚が難しくなることを示唆している。

音声知覚に関する情報は，子音部に多く含まれているとされているが，母音部にも音声知覚に関する情報は含まれている[1]。また，母音部は子音部よりも一般的にエネルギーが大きく持続時間も長い傾向にある。その結果，C マスカーで子音部を選択的にマスキングする際に，母音部に含まれる情報がマスクされずに知覚された結果，マスキングの効率がそれほど改善されなかった可能性がある。

6　おわりに

本研究では音声の母音部・子音部に対して選択的にエネルギーマスキングを施したときの単語了解度を調査した。その結果，子音部でマスカーの振幅を低下させることは，そのレベルの下げる程度によっては従来の定常雑音よりもマスキングの効率が上げられることが分かった。しかし，そのレベルを下げすぎることは，従来の定常雑音がよりもマスキングの効率は下がることが分かった。また，音声の子音部よりも，母音部をマスキングすることで，音声知覚が難しくなることが示唆された。

今後の課題として，マスキングを施す部分を様々なパターンで試すことが考えられる。例えば，ある特定の部分に注目せずに，マスカーの振幅をある一定の変調周波数で変化させるマスカーが考えられる。人が音声認識をする際に，音声のある特定の部分を重点的にマスキングすることが有効であるのか，あるいは，振幅を変化させるだけでターゲット音声を聴取しづらくさせることが可能であるならば，それだけでもマスキング効率を改善できるのかなどを調査する必要がある。

謝辞

本研究を進めるにあたって，実験参加者として協力してくださった方々に感謝申し上げます。

参考文献

[1] S. Furui, J. Acoust. Soc. Am., 80 (4), 1016-1025, 1986.
[2] 辻他，日音学誌，69（4），179-183，2013.
[3] 佐伯他, 日本音響学会誌, 59（4）, 209-214, 2003.
[4] 天野他，NⅡ音声資源コンソーシアム，2003.