話者自身の音声を利用したスピーチプライバシー技術の提案

川畑　尚也[†1]　藤枝　大[†1]

概要：近年，会話の内容や個人情報が周囲の第三者に聞こえ，情報が漏洩することが問題になっている．会話内容の漏洩を防ぐスピーチプライバシーを実現するために，様々な方法が提案されている．スピーチプライバシーを実現するために最も良い方法は，壁や衝立などの遮音壁を設置する方法であるが，設置時のコストや導入時に大規模な工事が必要という問題がある．遮音壁とは別の方法として，音のマスキング効果を利用した方法がある．音のマスキング効果を利用した方式は，低価格で設置も遮音壁を設置するよりも簡単に導入できる．本論文では，話者自身の音声を変形せずに長いフレーム長でデータベースに蓄積し，データベースから過去のフレーム信号を複数フレーム選択して，選択したフレームを重畳することでマスカーを生成する手法を提案する．提案法の有効性を確認するために評価実験を行い，提案法は従来法よりもマスキング効果は良く，聞き心地は従来法と同等になることを確認した．

1. はじめに

近年，不特定多数の人が存在する施設（例えば，病院，薬局，銀行等）にある受付カウンター，窓口，打合せスペース等で話者が会話の相手と会話を行うと，会話の内容や個人情報が周囲の第三者に聞こえ，情報が漏洩することが問題になっている．第三者に会話内容の漏洩を防ぐことをスピーチプライバシー技術[1][2]と言い，スピーチプライバシーを実現するために，様々な方法が提案されている．スピーチプライバシーを実現するために最も良い方法は，壁や衝立などの遮音壁を設置する方法である．遮音壁の中で話すことで，遮音壁の外に会話の音声が漏れることなく会話することができる．しかし，遮音壁は，設置時のコストや導入時に大規模な工事が必要という問題がある．

遮音壁とは別の方法として，音のマスキング効果を利用した方法が提案されている．音のマスキング効果とは，ある音（以下，ターゲット音）が聞こえている状態で，ターゲット音に近い音響特性（例えば，周波数特性，ピッチ，フォルマント等）を持つ別の音（以下，マスカー信号）が存在した場合，対象音が聞き取りにくくなる（マスクされる）現象である．音のマスキング効果を利用した方式は，遮音壁を設置するよりも低価格で簡単に設置可能なので，簡単に導入できる．最近では，音のマスキング効果を利用したスピーチプライバシー技術の研究[5]-[12]も行われており，スピーチプライバシー用のスピーカシステム[3][4]も販売されている．

従来の音のマスキング効果を利用した手法は，ピンクノイズなどの定常雑音をマスカー信号として使用する方法[5][6]や人の音声を使用した方法[7]-[12]がある．佐伯ら[5][6]は，ホワイトノイズ，ピンクノイズ，疑似音声雑音の 3 種類の雑音から帯域制限したピンクノイズが音声をマスクするために最も有効であるとし，スピーチプライバシー技術でマスキングノイズを設定する指標としてスペクトル距離が有効であると述べている．三戸ら[7][8]は，他人の音声が保存されているデータベースを用意し，入力された音声とデータベースに保存されている音声をフレーム毎に比較し，基本周波数が最も差が小さいデータベースの音声をそのフレームのマスカー信号とて使用している．さらに，発話者自身の音声を加工してマスカー信号として使用することで，他人の音声を使用して作成したマスカー信号よりマスキング効果が高くなるという報告[9][10]もされている．Ito ら[9]や Ueno ら[10]は，入力信号を長いフレームで切り出し，切り出したフレームを，短いフレームに分割し，時間軸方向に反転して，ランダムに置き換えた音をマスカー信号として使用している．また，太長根ら[11]や赤城ら[12]は，音声の言語意味理情報を担う音韻性を曖昧にする音をマスカー信号として使用するために，入力信号のスペクトル包絡のみを反転した音をマスカー信号として使用している．しかし，ピンクノイズをマスカー信号として使用する方法は，マスクキング効果は高いが，マスカー信号がピンクノイズなのでうるさく感じる．他人の音声や話者自身の音声を使用する場合は，データベースの他人の音声や話者の音声信号を短いフレームで変形してマスカー信号を生成しているので，マスカー信号が人工的な音になってしまい，不快な音になる可能性がある．

本論文では，話者自身の音声を変形せずに長いフレーム長でデータベースに蓄積し，テータベースに蓄積された過去のフレーム信号を複数フレーム選択して，選択したフレームを重畳してマスカー信号を生成することで，マスキング効果が高く，生成したマスカー信号の不快感を軽減する方式を提案する．

本論文は以下のように構成される．2 章で提案法を説明し，3 章で評価実験を示し，4 章でまとめる．

2. 提案法

話者自身の音声を変形させずに長いフレーム長でデータベースに蓄積し，テータベースに蓄積された過去のフレ

†1 沖電気工業株式会社　経営基盤本部　研究開発センター

ーム信号を複数フレーム選択して，選択したフレームを重畳してマスカー信号を生成する手法を提案する．長いフレーム長のフレーム信号を使用してマスカー信号を生成するため，人工的な音にならず，不快感を軽減することができる．長いフレーム信号を使用することでマスカー信号が言葉として聞こえるのを防止するために，データベースからマスカー信号を生成するために使用するフレーム信号を複数フレーム選択して,選択したフレームを重畳することで，ヒューマンスピーチライクノイズ（Human Speech-Like Noise：HSLN）[13][14]にし,マスカー信号として使用する．HSLN は音声信号を重畳して生成する信号なので，音声と類似した音響特性を有し，効果的に音声をマスクできると考えられる．また，話者自身の音声を使用してマスカー信号を生成する手法は，マスキング効果が高いという報告[9][10]がされているので，提案法もマスキング効果を高くすることができる．さらに，話者自身の音声を使用することで，ピッチの推定値などの推定結果が間違っていても，マスカー信号の音響特性が近くなるので，マスクできると考えられる．提案法のブロック図を図 1 に示す．

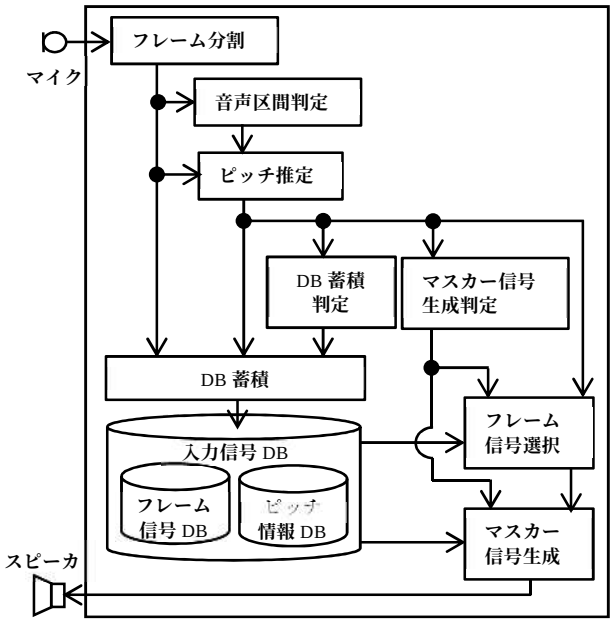

図 1 提案法のブロック図

2.1 入力信号のフレーム分割

提案法では，まずマイクロホンの入力信号$x(n)$を，式(1)に従いフレームに分割する．

$$x_{fram}(l;m) = x(l \cdot S + m) \tag{1}$$

ここで，$x_{fram}(l;m)$はフレーム分割したマイク入力信号，lはフレーム番号，mはフレーム内の時間$(m = 0,1,\cdots M-1)$，Mはフレーム長，Sはシフト幅である．

2.2 音声区間検出

フレーム分割したマイク入力信号$x_{fram}(l;m)$を用いて，音声区間か非音声区間かを判定する音声区間検出（Voice Activity Detection：VAD）を行う．本提案法では，初期数フレームを無音区間とし，無音区間の二乗平均平方根（Root Mean Square：RMS）の平均値を式(2)で求める．

$$RMS_N = \frac{1}{L_N}\sum_{l=0}^{L_N-1}\sqrt{\frac{1}{M}\sum_{m=0}^{M-1}\left(x_{fram}(l;m)\right)^2} \tag{2}$$

ここで，RMS_Nは無音区間の平均 RMS 値，L_Nは無音区間のフレーム数である．そして，現フレームの RMS 値を式(3)で求め，RMS_Nを閾値として VAD を式(4)で行う．

$$RMS(l) = \sqrt{\frac{1}{M}\sum_{m=0}^{M-1}\left(x_{fram}(l;m)\right)^2} \tag{3}$$

$$\mathrm{VAD}(l) = \begin{cases} 1 & (RMS(l) > 10^{\frac{R}{20}} \cdot RMS_N) \\ 0 & (\text{otherwise}) \end{cases} \tag{4}$$

ここで，$RMS(l)$は現フレームの RMS 値，VAD(l)は VAD の判定結果である．また，R[dB]は，$RMS(l)$に対する閾値を決定するための dB 尺度上のRMS_Nの係数である．

2.3 ピッチ推定

$x_{fram}(l;m)$に対してピッチ推定を行い，各フレームのピッチの推定値を求める．ピッチ推定は YIN 法[15]を使用する．ピッチを推定後に，ピッチの推定値の平均を式(5)で求め，ピッチの補間を式(6)で行う．

$$\bar{f}(l) = \frac{1}{N_{ave}}\sum_{q=0}^{N_{ave}-1} f(l;p-q) \tag{5}$$

$$f'(l) = \begin{cases} \bar{f}(l) & \left(\mathrm{VAD}(l) = 1,\ \bar{f}(l) \neq 0\right) \\ \alpha \cdot f'(l-1) & \left(\mathrm{VAD}(l) = 1,\ \bar{f}(l) = 0\right) \\ 0 & (\text{otherwise}) \end{cases} \tag{6}$$

ここで，$f(l;p)(p = 0,1,\cdots,P-1)$は YIN 法で求めたピッチの推定値，$P(P = M/M_{pitch})$はピッチ推定回数，$M_{pitch}$はピッチ推定のフレーム長，$N_{ave}(N_{ave} \leq P)$はピッチの平均フレーム数，$\bar{f}(l)$はピッチの推定値の平均値，$\alpha(0 < \alpha < 1)$はピッチの減衰経数，$f'(l)$はピッチの補間値である．式(6)は，VAD で音声区間と判定され，$\bar{f}(l)$が 0[Hz]以外の場合は音声区間でピッチが正しく推定されたとして$\bar{f}(l)$を，VAD で音声区間と判定され，$\bar{f}(l)$が 0[Hz]の場合は音声区間だがピッチ正しく推定されなかったとして，1 フレーム前のピッチの補間値$f'(l)$に減衰係数を乗算した値を，VAD で非音声区間と判定されたときは，0[Hz]をピッチの補間値とする．

2.4 データベース蓄積判定

データベースにフレーム信号とピッチの推定値を蓄積するか否かの判定は，式(7)に従い判定する．

$$flag_{DB}(l) = \begin{cases} 1 & (f'(l) > TH_{pitch}) \\ 0 & (\text{otherwise}) \end{cases} \tag{7}$$

ここで，$flag_{DB}(l)$は蓄積判定結果，TH_{pitch}はピッチに対する閾値である．$flag_{DB}(l)$が 1 の時のみ，データベースにフレーム信号とピッチ推定値を蓄積する．

2.5 データベース蓄積

データベースの作成は，$flag_{DB}(l)$が 1 の時のみ，式(8)～(10)に従い，クラスごとに$x_{fram}(l;m)$と$\hat{f}(l)$を蓄積する．

$$class(l) = \begin{cases} 0 & \left(f'(l) \le f_{TH}(0)\right) \\ c & \left(f_{TH}(c-1) < f'(l) \le f_{TH}(c)\right) \\ N_{class} & \left(f_{TH}(N_{class}-1) < f'(l)\right) \end{cases}, \quad c = 1, \cdots, N_{class}-1 \tag{8}$$

$$DB_{singal}(class(l);i;m) = x_{fram}(l;m) \tag{9}$$

$$DB_{pitch}(class(l);i) = f'(l) \tag{10}$$

ここで，cはクラス番号，N_{class}はクラス数，$class(l)$はクラス判定結果，$f_{TH}(c)$はクラスcの上限周波数，$DB_{singal}(f;i;m)$はフレーム信号のデータベース，$DB_{pitch}(f;i)$はピッチの推定値のデータベース，$i(i = 0,1,\cdots L_{DB}-1)$はデータベースにデータが蓄積されるとインクリメントされるインデックス（iはL_{DB}になると 0 になり，再びインクリメントされる），L_{DB}は各クラスのデータベース長である．

2.6 マスカー信号生成判定

マスカー信号を生成するか否かの判定は，式(11)に従い判定する．

$$flag_{mask}(l) = \begin{cases} 1 & (f'(l) > TH_{mask}) \\ 0 & (\text{otherwise}) \end{cases} \tag{11}$$

ここで，$flag_{mask}(l)$はマスカー信号生成判定結果，TH_{mask}はマスカー信号生成判定のピッチの閾値である．$flag_{mask}(l)$が 1 の時のみ，データベースに保存されている過去のフレーム信号から，マスカー信号を生成するために使用するフレームを複数フレーム選択し，マスカー信号を生成する．

2.7 データベースのフレーム信号の選択

フレーム信号の選択は，$flag_{mask}(l)$が 1 の時のみ,式(12)と(13)に従い，過去のフレームを複数フレーム選択する．

$$Sub(l;i) = \left| f'(l) - DB_{pitch}(class(l);i-1) \right| \tag{12}$$

$$T(l;u) = small(Sub(l;i), u+1) \tag{13}$$

ここで，$Sub(l;i)$は$f'(l)$と$DB_{pitch}(class(l);i)$のとの差の絶対値，$T(l;u)(u = 0,1,\cdots,U-1)$は選択したフレーム番号，$U$は選択するフレーム数である．また，式(13)の$small(Sub(l;i), u+1)$は，$Sub(l;i)$で$u+1$番目に小さい$Sub(l;i)$のインデックス$i$を出力する関数である．式(13)は，まず，$Sub(l;i)$が最小になるインデックス$i$が$T(l;0)$に代入され，$Sub(l;i)$が 2 番目に小さくなるインデックス$i$が$T(l;1)$に代入され，以降，$Sub(l;i)$が$U$番目に小さくなるインデックス$i$まで$T(l;U-1)$に代入される．つまり，ピッチが近い順にインデックスiが$T(l;u)$に代入し，マスカー信号作成時に使用することで，ピッチが近いU個のフレーム信号を選択できる．

2.8 マスカー信号の生成

マスカー信号の生成は，$flag_{mask}(l)$が 1 の時のみ,式(14)に従い，HSLN を生成する．

$$h(l,m) = \begin{cases} \frac{1}{U}\sum_{u=0}^{U-1} DB_{singal}(c(l);T(l;u);m) & (flag_{mask}(l) = 1) \\ 0 & (\text{otherwise}) \end{cases} \tag{14}$$

ここで，$h(l,m)$は生成した HSLN 信号である．(14)式は，選定したフレーム番号$T(l;u)$に対応するフレーム信号をデータベース$DB_{singal}(c(l);T(l;s);m)$から複数フレーム読出して，重畳することで$h(l,m)$を生成する．

HSLN 信号$h(l,m)$を生成後は，不連続点をなくすために窓掛けを行い，オーバーラップ加算を行ってマスカー信号として出力する．

3. 評価実験

提案法の有効性を確認するために，ターゲット音声とマスクするとマスカー信号をスピーカから再生し，被験者が聴取して，音を評価する主観評価実験を行った．

3.1 実験条件

従来法と提案法を計算機シミュレーションで実装し，ターゲット音声を入力してマスカー信号を生成した．従来法は，ピンクノイズをマスカー信号として使用する手法[5]と，提案法のデータベースを短いフレーム長で他者の音声から作成する手法とした（以下，ピンクノイズをマスカー信号として使用する手法を従来法 1，提案法のデータベースを短いフレーム長で他者の音声から作成する手法を従来法 2 とする．）．提案法は，動作開始時はデータベースに話者自身の音声信号が蓄積されていないので，データベースを蓄積するために，同じ話者の音声を入力してからターゲット音を入力してマスカー信号を生成した．

評価実験は，使用する音は，ターゲット音声と作成したマスカー信号の比（Target to Masker Ratio：TMR）が 0[dB]になるように，計算機シミュレーションで生成したマスカー信号の振幅を調整した．そして，ターゲット音声が Lch，振幅を調整したマスカー信号が Rch の音ファイルを 15 ファイル×3 方式の合計 45 個を作成し，2 つのスピーカから出力する．出力レベルは，スピーカから 3.0m の位置（被験者が座る位置），高さ 1.0m（被験者の耳の高さになる位置）

で約 55.0[dBA]になるように出力した．被験者は，スピーカから出力された音を聴取し，評価シートに記入する．被験者の人数は，聴力が正常な男性 5 人で行った．

評価尺度は，聴き取りにくさ，聞き心地評価の 2 つの評価尺度で評価する．「聴き取りにくさ」は，森本ら[16][17]が提案した音声を評価する指標である．「聴き取りにくくはない」，「やや聴き取りにくい」，「かなり聴き取りにくい」，「非常に聴き取りにくい」の 4 段階の評価尺で評価し，「聴き取りにくくはない.」以外の尺度は，程度は異なるものの「聴き取りにくい」という判断し，「聴き取りにくくはない.」以外の尺度の回答をひとまとめにし，これらが全回答に対して占める割合を聴き取りにくさとして算出する．聴き取りにくいと判断された割合を算出するので，値が高いほどマスキング性能が高い（聴き取りにくい）ことを示す．

表 1 に実験機器，表 2 に実験条件，表 3 に従来法と提案法のパラメータ値，表 4 と表 5 に各評価尺度，図 2 と図 3 に機器配置図，図 4 にターゲット音声信号（F101SF_B01.AD），図 5 図 5 から図 7 に各方式で生成した F101SF_B01.AD のマスカー信号を示す．

表 1　実験機器

機器	メーカ，型番
スピーカ	ELECOM MS-P08UWH
オーディオ IF	ROLAND OCTA-CAPTURE UA-1010

表 2　実験条件

実験場所	実験室
暗騒音レベル	38.5[dBA]
TMR	0[dB]
再生ファイル数	45 ファイル （各 15 ファイル×3 方式）
出力レベル	55.0[dBA]（被験者の耳の位置）
被験者	男性 5 名
サンプリング周波数	48,000[Hz]
量子化ビット数	16[bit]

表 3　提案法と従来法のパラメータ値

パラメータ	値
ターゲット音声	ATR デジタル音声データベース セット C 連続発声 B C3-F01 F101SF_B01.AD～F101SF_B15.AD
従来法 1 のピンクノイズの周波数帯域幅	176.75[Hz]～5656[Hz]
従来法 2 の他者音声	ATR デジタル音声データベース セット C 最重要単語 C1-M01（M102, M103, M104, M105） C1_F01（F102, F103, F104, F105）
従来法 2 のシフト幅S	4096[sample]
従来法のフレーム長M	8192[sample]
提案法のデータベースの蓄積に使用した音声	ATR デジタル音声データベース セット C 連続発声 B C3-F01 F101F101SF_B16.AD～ F101SF_B20.AD
提案法のシフト幅 S	10240[sample]
提案法のフレーム長M	20480[sample]
無音区間のフレーム数L_n	5[frame]
VAD 閾値R	5.0[dB]
ピッチの平均フレーム数N_{ave}	4[frame]
ピッチ推定回数P	10
ピッチ推定フレーム長M_{pitch}	1024[sample]
ピッチの減衰経緯数α	1.0(減衰なし)
ピッチの閾値 TH_{pitch}	0
クラス数N_{class}	10
各クラスの下限周波数$f_{TH}(c)$	$f_{TH}(0) = 56.1, f_{TH}(1) = 70.7,$ $f_{TH}(2) = 89.8, f_{TH}(3) = 112.3,$ $f_{TH}(4) = 140.3, f_{TH}(5) = 179.6,$ $f_{TH}(6) = 224.5, f_{TH}(7) = 280.6,$ $f_{TH}(8) = 353.6, f_{TH}(9) = 449.0$
各クラスのデータベース長L_{DB}	100
閾値TH_{mask}	56.1
選択するフレーム数 U	4

表 4　聴き取りにくさ

1	聴き取りにくくはない.
2	やや聴き取りにくい.
3	かなり聴き取りにくい.
4	非常に聴き取りにくい.

表 5　聞き心地

1	非常に悪い.
2	悪い.
3	良い.
4	非常に良い.

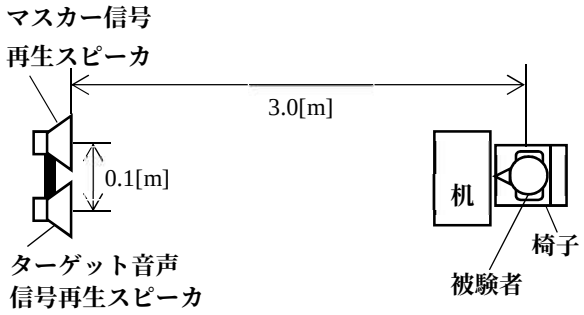

図 2　機器配置図（上面図）

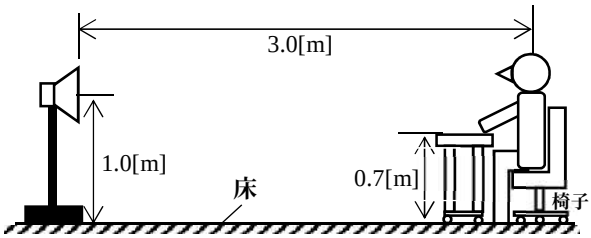

図 3 機器配置図（側面図）

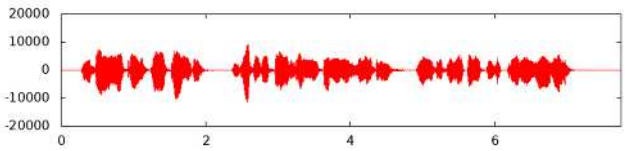

図 4 ターゲット音声（F101SF_B01.AD）

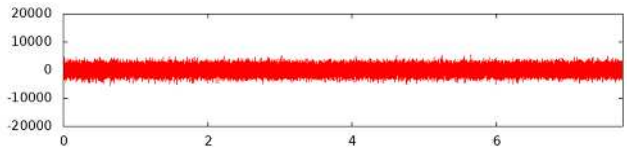

図 5 F101SF_B01.AD の従来法 1 マスキング音

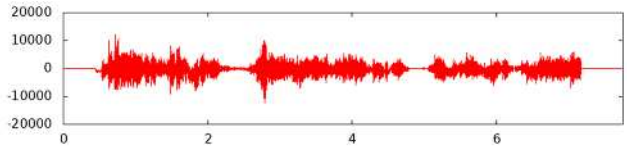

図 6 F101SF_B01.AD の従来法 2 マスキング音

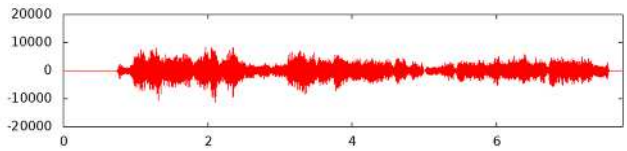

図 7 F101SF_B01.AD の提案法マスキング音

3.2 実験結果，考察

図 8 に聴き取りにくさの主観表結果，図 9 に聞き心地の主観評価結果を示す.

図 8 より，提案法は，従来法 1 や従来法 2 より聴き取りにくさの評価で良い結果になった．従来法 1 や従来法 2 はピンクノイズや他者の音声を使用してマスカー信号を生成しているため，ターゲット音とマスカー信号の音響特性が大きく異なり聴き取りにくさが低くなった（聴き取り易くなった）と考えられる．一方，提案法は話者の音声を使用してマスカー信号を生成しているため，ターゲット音とマスカー信号の音響特性が近くなり，聴き取りにくさが高くなった（聴き取りにくくなった）と考えられる．また，従来手 1 と提案法, 従来法 2 と提案法で T 検定を行った結果, 有意差がある結果になった.

また，図 9 より，提案手法は，聞き心地で従来法 1 と同等で，従来法 2 より良い結果になった．これは，従来法 1 はピンクノイズを使用しているので定常的な音になり，あまり聞き心地が悪くならなかったと考えられる．従来法 2 は短いフレーム長で他社の音声からマスカー信号を生成しているので，マスカー信号が耳障りな音になり聞き心地が悪くなったと考えられる．一方，提案法は，話者自身の音声を使用しているため，ターゲット音声とマスカー信号が違和感なくマスクされたため，従来法 1 と同等の聞き心地になったと考えられる.

以上の結果から，提案法は聴き取りにくさで従来法 1 と従来法 2 より性能が良く，聞き心地は従来法 1 と同等になり，提案法の有効性を確認できた.

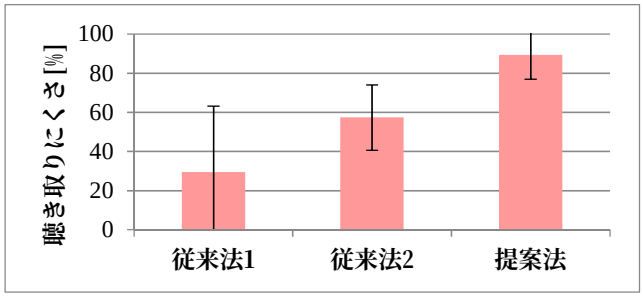

図 8 聴き取りにくさの主観評価結果

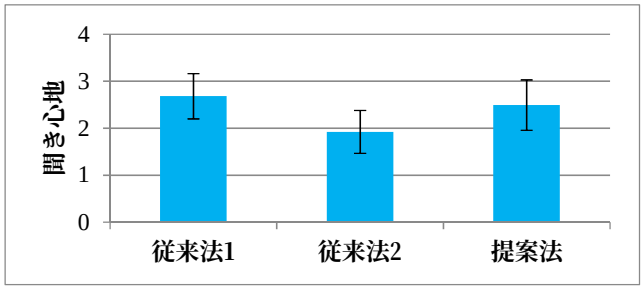

図 9　聞き心地の主観評価結果

4. まとめ

本論文では，話者の音声信号を変形させずに長いフレーム長でデータベースに蓄積し，テータベースに蓄積された過去のフレーム信号を複数フレーム選択して，選択したフレームを重畳してマスカー信号を生成する手法を提案した．評価実験の結果，提案法は，聴き取りにくさのは従来法より良く，聞き心地は従来法 2 より良く，従来法 1 と同等になることがわかった．

今後の課題として，システム開始直後はデータベースに話者の音声が蓄積されておらずマスキング性能が落ちるため，システム開始からデータベースに話者の音声が蓄積されるまでマスキング性能を良くする対策や，今回の評価ではオフライン処理でマスカー信号を作成したので，リアルタイムでマスカー信号を生成するリアルタイムシステムへの実装などがあげられる．

参考文献

[1] W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, and B. G. Watters, “Speech privacy in buildings”, The Journal of the Acoustical Society of America, vol.34, no.4, pp.475-492, 1962.
[2] 佐藤 洋, 清水 寧, “スピーチプライバシー研究の歴史と近年動向”, 日本音響学会誌, vol.64, no.8, pp.475-480, 2008.
[3] “スピーチプライバシーシステム VSP-1”. https://sound-solution.yamaha.com/products/speechprivacy/vsp-1/index, (参照 2019-12-20).
[4] “スピーチプライバシーシステム VSP-2“. https://sound-solution.yamaha.com/products/speechprivacy/vsp-2/index, (参照 2019-12-20).
[5] 佐伯 徹郎, 藤井 健生, 山口 静馬, 老松 建成, “音声をマスクするための無意味定常雑音の選定”, 電子情報通信学会論文誌, vol.J86-A, no.2, pp.187-191,2003.
[6] 佐伯 徹郎, 山口 静馬, 為末 浩二, “マスキングノイズによるスピーチプライバシー保護に関する一考察”, 日本音響学会誌, vol.61, no.10, pp.571-575, 2005.
[7] 三戸武大, 荒井隆行, 安啓一, “サウンドマスキングシステムにおけるデータベースを用いたマスカー作成法の提案”, 日本音響学会講演論文集秋季, pp.1243-1246, 2012.
[8] 三戸武大, 荒井隆行, 安啓一, “サウンドマスキングシステムにおけるデータベースを用いた音声マスカ作成法の提案”, 日本音響学会誌, vol.71, no.8, pp.382-389, 2015.
[9] A. Ito, A. Miki, Y. Shimizu, K. Ueno, H. J. Lee, and S. Sakamoto, “Oral information masking considering room environmental condition, Part 1 : Synthesis of Maskers and examination on their masking efficiency”, *Proc. of inter-noise 2007*, 2007.
[10] K. Ueno, H. J. Lee, S. Sakamoto, A. Ito, A. Miki, and Y. Shimizu, “Oral Information Masking Considering Room Environmental Condition Part 2 : Subjective Assessment for “Masking Efficiency and Annoyance” ” , *Proc. of inter-noise 2007*, 2007.
[11] 太長根 理会子, 赤木 正人, 入江 佳洋, “音源の近く的友邦にもとづいた会話のプライバシー保護の検討”, 日本音響学会講演論文集春季, pp.311-312, 2005.
[12] 赤木 正人, 入江 佳洋, “音情景解析の概念にもとづいた音声プライバシー保護”, 電子情報通信学会論文誌, vol.J97-A, no.4, pp.247-255, 2014.
[13] D. Kobayashi, S. Kajita, and K. Takeda, “Extracting speech features from human speech like noise”, *Proc. International Conference on Spoken Language 96*, pp.418-421, 1996.
[14] 梶田 将司, 小林 大祐, 武田 一哉, 板倉 文忠, “ヒューマンスピーチライク雑音に含まれる音声的特徴の分析” 日本音響学会誌, vol.53, no.5, pp.337-345, 1997.
[15] A. Cheveigne and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music”, The Journal of the Acoustical Society of America, vol.111, no.4, pp.1917-1930, 2002
[16] M. Morimoto, H. Sato, and M. Kobayashi, “Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces”, The Journal of the Acoustical Society of America, vol.116, no.3, pp.1607-1613, 2004.
[17] 佐藤逸人, 森本正之, 佐藤洋, “聴き取りにくさによる音声伝送性能の評価”, 日本音響学会誌, vol.63, no.5, pp.275-280, 2007.