

平成 25 年度
学士学位論文

ウェーブレット係数を特徴量とする音源 分離

Sound source separation of which the amount of the
feature is the wavelet coefficient

1140390 横田優佳

指導教員 福本昌弘

2014 年 3 月 6 日

高知工科大学 情報学群

要 旨

ウェーブレット係数を特徴量とする音源分離

横田優佳

近年、音声認識技術が向上しているが、複数話者の音声に対しては高精度な認識が困難である。複数の音声から 1 人ずつの音声を抽出できれば、抽出したそれぞれの音声を認識させることで会議の議事録の自動生成等ができるようになる。

複数の音声を分離する技術として音源分離技術がある。音源分離技術の従来手法には独立成分分析や DUET といった手法がある。これらの手法は音源数が未知の場合は適用できない、または、複数のマイクを離して設置できない等の問題点がある。しかし、これらの問題点は単一マイクを用いて録音することで解決できる。

そこで本研究では、単一マイクを用いて録音した複数話者音声からウェーブレット係数を特徴量として所望の音声を抽出するシステムの提案を行う。また、提案したシステムの有効性を確認するためにシミュレーションを行い、男性と女性の混合音声から所望の音声を抽出できることを確認している。しかし、男性 2 人や 3 人以上の混合音声に対しては所望の音声を抽出できていない。

キーワード 音源分離, ウェーブレット変換, 単一マイク

Abstract

Sound source separation of which the amount of the feature
is the wavelet coefficient

YOKOTA Yuka

In recent years, speech recognition technology has improved, but recognition with high accuracy is difficult for the speech of multiple speakers. Being able to extract the voice of one person from multiple voice, it becomes possible and be automatic generation of conference proceedings by to recognize sound each extracted.

There is sound source separation technology as a technique for separating the multiple voice. There is a technique such as independent component analysis or DUET in the conventional method of sound source separation technology. There is a problem that can not be applied when the number of sound sources is unknown, or can not be set the microphone of more than one of these approaches. However, these problem can be solved by making a recording using a single microphone.

Therefore, in this paper, we propose a system for extracting desired sound as a feature of the wavelet coefficient from multiple voice which recorded by single microphone. And, it has been confirmed that the simulation is performed in order to confirm the effectiveness of the system proposed, it is possible to extract desired voice from mixed voice of one man and one woman. However, it is not possible to extract the desired voice for the mixed sound of 3 or more and two men.

key words sound source separation, wavelet transform, signal microphone

目次

第 1 章	序論	1
1.1	本研究の背景と目的	1
1.2	本論文の構成	2
第 2 章	音源分離	3
2.1	音源分離技術	3
2.1.1	DUET ブラインド音源分離	4
2.1.2	問題点	5
2.2	ウェーブレット変換	6
2.2.1	連続ウェーブレット変換	7
2.2.2	離散ウェーブレット変換	8
2.3	多重解像度解析	8
第 3 章	抽出システム	10
3.1	音声抽出システム	10
3.1.1	音声抽出システムの流れ	10
3.1.2	音声の抽出方法	11
第 4 章	混合音声からの所望音声抽出処理	13
4.1	シミュレーション条件	13
4.2	シミュレーション結果	21
4.2.1	男女 2 人の混合音声 (条件 1)	21
4.2.2	男性 2 人の混合音声 (条件 2)	23
4.2.3	男女 3 人の混合音声 (条件 3)	25
4.3	考察	27

目次

第 5 章	結論	29
5.1	本研究のまとめ	29
5.2	今後の課題	29
	謝辞	31
	参考文献	32

目次

2.1	ブラインド音源分離システム	4
2.2	DUET の音源分離の過程	5
3.1	音声抽出システムの流れ	11
4.1	条件 1 における所望音声	15
4.2	条件 1 における非所望音声	15
4.3	条件 1 における簡易合成音声	16
4.4	条件 1 における混合音声	16
4.5	条件 2 における所望音声	17
4.6	条件 2 における非所望音声	17
4.7	条件 2 における簡易合成音声	18
4.8	条件 2 における混合音声	18
4.9	条件 3 における所望音声	19
4.10	条件 3 における非所望音声 I	19
4.11	条件 3 における非所望音声 II	20
4.12	条件 3 における簡易合成音声	20
4.13	条件 3 における混合音声	21
4.14	条件 1 における簡易合成音声から抽出した音声	22
4.15	条件 1 における所望音声と抽出音声の比較	22
4.16	条件 1 における混合音声から抽出した音声	23
4.17	条件 2 における簡易合成音声から抽出した音声	24
4.18	条件 2 における所望音声と抽出音声の比較	24
4.19	条件 2 における混合音声から抽出した音声	25

図目次

4.20 条件 3 における簡易合成音声音声から抽出した音声	26
4.21 条件 3 における所望音声と抽出音声の比較	26
4.22 条件 3 における混合音声から抽出した音声	27

表目次

4.1 録音環境	13
4.2 シミュレーション条件	13

第 1 章

序論

1.1 本研究の背景と目的

近年、音声認識技術が向上してきていることにより音声認識を利用したサービスが増加している。マイクの直近で 1 人の話者が発話した音声に対しては、高精度な音声認識が可能であるが、複数の話者が同時に発話した音声に対する認識率は低い [1]。複数音声の分離ができれば、会議の議事録の自動生成等が可能となる。

複数の音をそれぞれの音に分離測定する技術を音源分離技術と呼ぶ。現在、音源分離技術の中で独立成分分析に基づく手法が分離性能が高いためよく用いられる [2]。この手法は、音源数以上のマイクが必要となる。マイクの数が音源数より少ない場合は不良設定問題となり、独立成分分析による音源分離が不可能となる [3]。会議等の実環境において多くの場合は音源数が未知であるため、独立成分分析を用いた音源分離は難しい。そこで、音源数に関わらない音源分離の手法が必要となる。

一方、信号のスパース性を仮定することで音源数に関わらず 2 つのマイクで録音した音を分離できる DUET という手法がある [3–5]。DUET を用いた音源分離において、マイク間の距離とサンプリング周波数がトレードオフの関係にあるという問題点がある。マイク間の距離を数 cm 以内にする必要があり、実環境での利用や応用が難しい。

そこで本研究ではこれらの問題点を考慮し、単一マイクで録音した複数話者の音声からウェーブレット係数を特徴量として所望の音声を抽出する。

1.2 本論文の構成

本論文の構成について述べる．

第 2 章では，本研究に関連する音源分離技術について触れ，本研究内で分離の際に用いる技術であるウェーブレット変換について述べる．

第 3 章では，音源分離の従来手法の問題点を述べた後に，問題点を解決するための音声抽出システムの提案を行う．

第 4 章では，第 3 章で提案したシステムの有効性を確認するためのシミュレーションを行い，所望の音声抽出できたことを確認する．また，その結果から得られた考察を述べる．

第 5 章では，本論文の結論と今後の課題について述べる．

第 2 章

音源分離

2.1 音源分離技術

普段生活している環境には複数の音が混在している．人間の耳は様々な音の中から目的の音を聞き分けることができる．人間の複数音から目的の音を聞き分ける能力をカクテルパーティー効果と呼び，カクテルパーティー効果を工学的な問題として扱ったものが音源分離である [2]．抽出したい音源の到来方向が既知の場合は，音源の方向からの信号を強調し，それ以外の方向からの信号を無視することで所望の信号を得ることができる．このような分離手法をビームフォーミングと呼ぶ．

また，音源分離の中でも音源位置や混合信号になるまでの過程が未知の状態で，音源の最小限の知識と観測信号のみを用いて音源信号を復元する技術をブラインド音源分離と呼ぶ．

ブラインド音源分離のシステムを図 2.1 に示す [6, 7]．図 2.1 における S_1, S_2 は音源信号， X_1, X_2 は観測信号， Y_1, Y_2 は出力信号を示し， H は音源からマイクに入力されるまでの遅延や残響等の特性を示し， W は分離を行うためのフィルタを示す．音源信号 S_1 と S_2 は互いに独立であると仮定する．この仮定は実環境において成り立つ．混合音声を録音するマイクを 2 つ用いると，観測信号 X_1 と X_2 には相関がある．この相関がある観測信号を入力として，出力信号 Y_1 と Y_2 を互いに独立とする分離フィルタ w_{ij} を逐次的に学習し，出力 Y_1, Y_2 を分離・抽出する処理がブラインド音源分離システムである．

ブラインド音源分離の実現手法として観測信号のみから分離信号が互いに独立となるような線形的な分離行列と分離信号の両方を推定する手法である独立成分分析が，分離性能が高いため主流となっており，この手法は音源数以上のマイクが必要となっている．しかし，会

2.1 音源分離技術

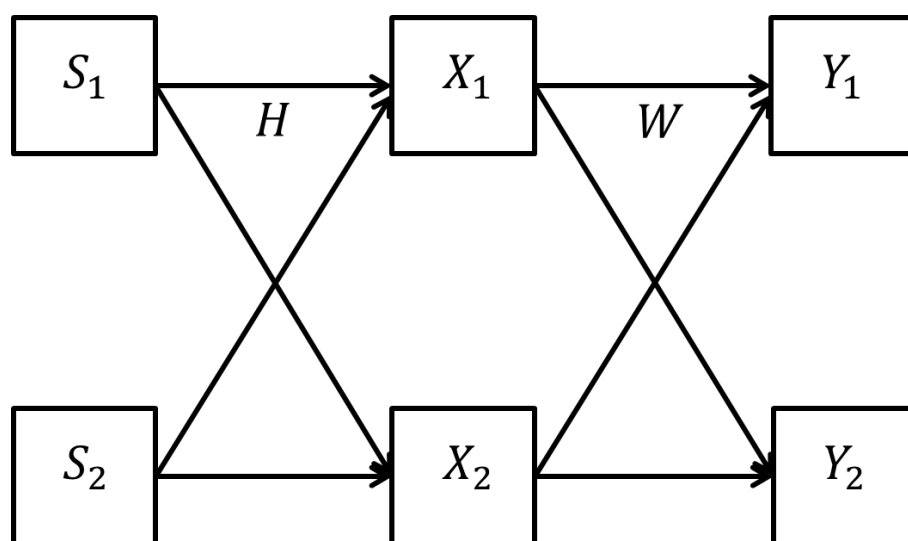


図 2.1 ブラインド音源分離システム

議等の実環境において多くの場合は音源数が未知であるため，独立成分分析を用いた音源分離は難しい．そのため，音源数に依存しない分離手法が必要となる．

2.1.1 DUET ブラインド音源分離

音声信号のスパース性を用いた手法は，マイクの数音源数よりも少ない場合でも音源分離が可能である．スパース性とは，信号のほとんどの時間周波数領域で 0 に近く大きな値を持つことが稀であるという性質である．スパース性に基づく音源分離では，各時間周波数において原信号のうち一つのみが支配的であることを仮定する．これは，音声信号において短時間フーリエ変換のフレーム長を適切に選ぶことで近似的に成り立つ [4]．音声信号のスパース性の仮定から，各時間に含まれる周波数は一つの音源信号の情報のみである．この各時間周波数の値を特徴量として音声の分離を行うことができる．

音声信号のスパース性を仮定することで音源数に関わらず信号を分離する手法の一つに DUET という手法がある [4]．DUET は音源数に関わらず 2 つのマイクで録音した音を分離することができ [3]，録音した音声の振幅-遅延ヒストグラムを特徴量として用いて分離を行う．

2.1 音源分離技術

DUET を用いた際の混合音声から個々の音源を分離するまでの過程を以下に示す。また、分離過程を図 2.2 に示す。

1. 2 つのマイクから混合音声 $x_1(t)$, $x_2(t)$ を取得する。
2. 混合音声 $x_1(t)$, $x_2(t)$ に短時間フーリエ変換を適用し、各時間周波数での振幅パラメータおよび遅延パラメータを算出する。
3. 算出したパラメータから振幅-遅延ヒストグラムを作成する。
4. 作成したヒストグラムを用いてピークサーチを行う。
5. ヒストグラムからクラスタに分類する。
6. 各クラスタに対応する時間周波数表現を短時間逆フーリエ変換する。

以上の過程によって混合音声から各音源を分離することができる。

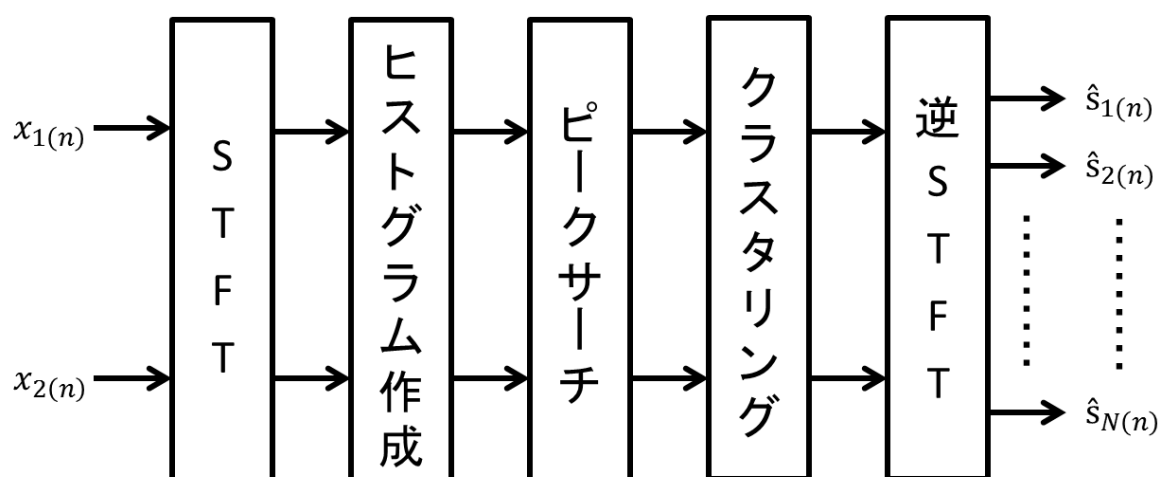


図 2.2 DUET の音源分離の過程

2.1.2 問題点

ブラインド音源分離の従来手法である独立成分分析に基づいた手法では、音源数以上のマイクを用いて録音する必要があるため、実環境での適用が困難である。一方、音声信号のスパース性を仮定することで音源数に関わらず信号を分離するのである DUET は 2 つのマイク

2.2 ウェーブレット変換

で録音した音声を用いて分離を行うため、上記の問題点を解決している。しかし DUET による音源分離において、マイク間の距離とサンプリング周波数がトレードオフの関係にあることが問題点として挙げられる。DUET では 2 つのマイク間の信号の遅延差が特徴量の一つとなっている。信号の遅延差が位相の制限範囲を越えた際、遅延差の導出に誤差が生じる。そのため、マイク間の距離を数 cm 以下に限定する必要がある。一方、マイク間の距離を大きくとる場合は最大周波数が低くなってしまいうため分解能が低くなる。この問題点は DUET の応用範囲を狭くするといえる。

音声信号の時間周波数変換を短時間フーリエ変換で行う場合、短時間フーリエ変換のフレーム長を短くすると時間分解能が上がるが周波数分解能が下がり、フレーム長を長くすると周波数分解能は上がるが時間分解能が下がるという関係がある。周波数分解能が下がると、特徴量とする周波数の精度が下がることから音源分離の特徴量として使用することができない。また、時間分解能が下がるとスパース性が成り立たなくなる。このような問題点があるため、本研究では時間分解能および周波数分解能の高いウェーブレット変換を時間周波数解析に用いる。

2.2 ウェーブレット変換

ウェーブレットは、“wave”(波)と“let”(小さい)の複合語であり、時間と周波数の小さなさざ波のことである。

観測された信号やパターン $h(t)$ に対して、ウェーブレットを用いてその信号 $h(t)$ の局在の様子を調べるためには、マザーウェーブレットと呼ばれる局所的な基本参照波 $\psi(t)$ およびそれを拡大縮小し、平行移動した参照波、

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

をいくつか用意して、これらとパターンとの関係を調べる。ここで、 a, b はそれぞれ伸張、移動パラメータである。ウェーブレット変換は、

$$\mathbf{W}_{\psi}(t)(a, b) \stackrel{\text{def}}{=} \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h(t) \psi\left(\frac{t-b}{a}\right) dt$$

2.2 ウェーブレット変換

と表せる.

マザーウェーブレットを拡大縮小し, 平行移動した様々な参照波のことを単にウェーブレットと呼ぶ.

観測パターンから特徴を抽出する際にフーリエ解析は, 観測パターンに含まれる全ての周波数成分, スペクトルを全ての時間軸上で調べる. 一方, ウェーブレット解析は, 観測パターンに最適なマザーウェーブレットおよびマザーウェーブレットを拡大縮小, あるいは平行移動したウェーブレットを用いて特徴のある時刻について詳細に調べることが出来る.

2.2.1 連続ウェーブレット変換

マザーウェーブレット $\psi(x)$ による関数 $h(t)$ の連続ウェーブレット変換は,

$$(\mathbf{W}_\psi h)(b, a) \stackrel{def}{=} \sqrt{\frac{1}{|a|}} \cdot \int_{-\infty}^{\infty} \overline{\psi\left(\frac{x-b}{a}\right)} \cdot h(x) dx$$

で与えられる. ここで, $\overline{\psi(\cdot)}$ は $\psi(\cdot)$ の複素共役である.

ウェーブレット $\psi((x-b)/a)$ は, マザーウェーブレット $\psi(x)$ を b 平行移動し, a 拡大縮小したものである. マザーウェーブレット ψ およびスケールパラメータ a , トランスレーションパラメータ b を適切に選ぶことにより, 信号 $h(x)$ の局在的様子を調べることが出来る.

逆に, ウェーブレット変換から元の信号 $h(x)$ を復元することも出来る. つまり, 逆ウェーブレット変換が存在し, 以下の様に定義される.

$$h(x) \stackrel{def}{=} \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mathbf{W}_\psi h)(b, a) \cdot \sqrt{\frac{1}{|a|}} \psi\left(\frac{x-b}{a}\right) \frac{1}{a^2} da db$$

ここで, 右辺が定義できるためには, 次のアドミッシブル条件を満たされなければならない.

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\widehat{\psi(\omega)}|^2}{|\omega|} d\omega < \infty$$

なお, $\widehat{\psi}$ は ψ のフーリエ変換である. しかし, 実際においては以下の条件式が代用されることがある.

$$\int_{-\infty}^{\infty} \psi(x) dx = 0$$

2.3 多重解像度解析

この式は $\psi(x)$ が振動的であることを意味する。

以上の式から基底により信号 $h(x)$ を表現することが出来る。しかし、ウェーブレット逆変換による信号の再構成のために a と b による無限積分が必要となるため、信号のウェーブレット変換および逆変換を行う際は離散ウェーブレット変換を用いる。

2.2.2 離散ウェーブレット変換

連続ウェーブレット変換では、パラメータ a , b が連続的な値を持つ。それに対して離散ウェーブレット変換では、パラメータが離散的な値をとる。

フーリエ級数に対応する離散ウェーブレットを考える。この基底関数列は、連続ウェーブレットのパラメータ a と b を 2 進ます目上に設定したものである。すなわち、

$$\psi_{m,n}(t) = \left(\frac{1}{2^{\frac{m}{2}}}\right)\psi(2^{-m}t - n)$$

とすればよい。ここで m , n はそれぞれ a , b に対応する正数である。 b に対応するのは $2^m \cdot n$ であり、移動量は基底関数の拡がりに依存する。この $\psi_{m,n}(t)$ の形式で 2 乗可積分関数空間 l_2 の直交基底を構成出来るので、直交基底のウェーブレットが実現できる。

マザーウェーブレット $\psi(x)$ による関数 $h(x)$ の離散ウェーブレット変換は、

$$d_k^{(j)} \stackrel{\text{def}}{=} 2^j \cdot \int_{-\infty}^{\infty} \overline{\psi(2^j x - k)} \cdot h(x) dx$$

で与えられ、その逆変換は、

$$h(x) = \sum_j \sum_k d_k^{(j)} \psi(2^j x - k)$$

で与えられる、離散ウェーブレット変換では、 ψ は基底関数となるものでなければならない。

2.3 多重解像度解析

音声や画像等のパターンを高い周波数成分と低い周波数成分に分けることが出来る。これを解像度を落とすといい、何度も行うことによりパターンの特徴を際立たせることが出来る。これを多重解像度解析と呼ぶ。

2.3 多重解像度解析

本研究では、観測信号に対して離散ウェーブレット変換による多重解像度解析を行うことで、音声の特徴量を解析し、所望する音声の抽出を行う。

第 3 章

抽出システム

3.1 音声抽出システム

2.1.2 の問題点の一つであるマイクに関する問題点は、単一マイクを用いて録音することで解決できる。しかし、従来の音源分離手法において複数のマイクを用いることで得られる遅延を特徴量の一つとしていたため、単一マイクを用いる場合は遅延以外の特徴量が必要となる。そのため、時間周波数解析に用いるウェーブレット変換を適用して得られるウェーブレット係数を特徴量として用いて所望の音声を抽出するシステムを提案する。

3.1.1 音声抽出システムの流れ

提案する音声抽出システムを用いて所望の音声を抽出するまでの過程を以下に示す。また、抽出システムの流れを図 3.1 に示す。

1. 単一マイクを用いて混合音声を取得する。
2. 複数の音声が入った観測信号において、単一話者区間音声と複数話者区間音声に信号を分割し、単一話者区間音声を音声抽出に用いるためのサンプルとする。
3. サンプルとする単一話者音声に対してフーリエ変換を適用し、ピークからなる信号を作成する。
4. 分割した複数話者音声とピーク信号に対してウェーブレット変換を適用し、得られたウェーブレット係数列を係数 w_1 , w_2 とする。
5. ウェーブレット係数列 w_1 とウェーブレット係数列 w_2 にパラメータをかけた値の差分

3.1 音声抽出システム

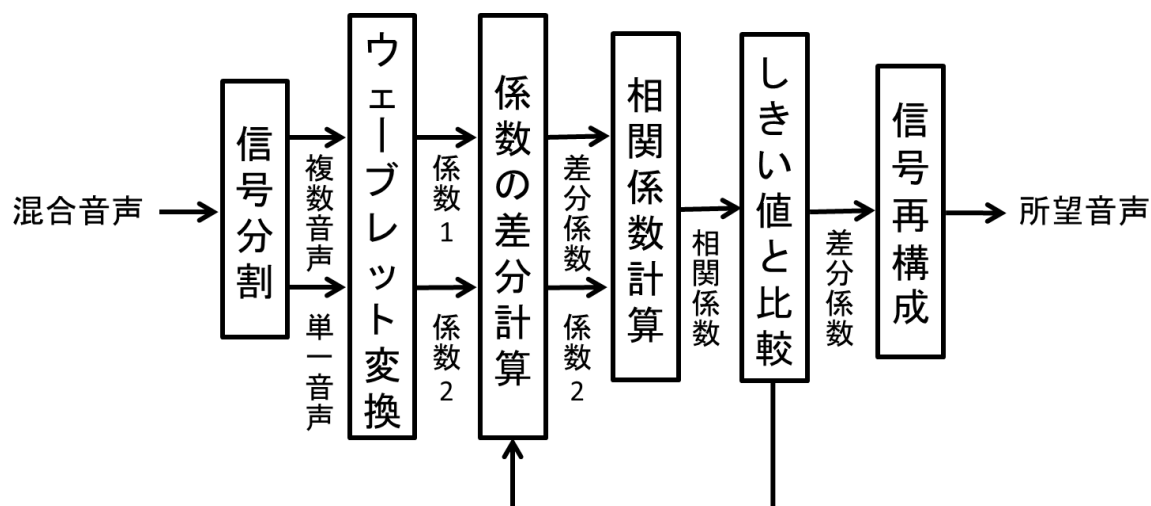


図 3.1 音声抽出システムの流れ

をとる。差分の式は、

$$w_{\text{diff}} = w_1 - p \times w_2$$

を用いる。ここで、 p はパラメータを表し、得られた差分を差分係数列とする。

6. 相関係数の値がしきい値に達するまで、図 3.1 内の係数の差分計算からの処理を繰り返す。
7. 相関係数がしきい値に達したとき、求めた差分係数列を用いて信号の再構成を行う。

以上の処理によって、混合音声から所望の音声を抽出する。

3.1.2 音声の抽出方法

提案する音声抽出システムは、単一マイクを用いて録音した音声から所望の音声を抽出する。そのため、従来手法の独立成分分析や DUET のマイクの数による問題点を解決できることが特徴である。

音声の抽出方法は、複数話者音声のウェーブレット係数列からサンプルとする単一話者音声のウェーブレット係数列にパラメータをかけた値を引くことで差分を計算する。この処理によって、複数話者音声からサンプルとする単一話者音声の周波数帯域を取り除く処理を行っている。この処理を相関係数がしきい値に達するまで繰り返すことで、複数話者音声か

3.1 音声抽出システム

ら 1 人分の音声を取り除くことができる．話者が 2 人以上の場合は，これらの処理を 1 度適用すると所望の音声を抽出できる．話者が 2 人以上の場合は，これらの処理を (混合音声に含まれている音声の人数 - 1) 回繰り返すことで所望の音声を抽出することができる．

所望音声の抽出に用いる単一話者音声区間は，音声信号のスパース性を用いて見つけることができる．しかし，スパース性が仮定される短時間の音声では周波数分解能が低いため，提案システムでは事前に単一話者区間がわかっていると仮定する．

第 4 章

混合音声からの所望音声抽出処理

4.1 シミュレーション条件

提案システムが有効であるかを確認するために 2 種類のシミュレーションを 3 つの条件に対して行った。1 種類目のシミュレーションは別々に録音した複数の音声进行簡易合成した音声に対して、2 種類目のシミュレーションは複数の人が同時に発話した音声に対して行った。

今回のシミュレーションでは、前提として録音した音声内において単一話者区間および複数話者区間がわかっているものとしてシミュレーションを行った。また、提案システム内の差分係数列とサンプルとする単一話者音声のウェーブレット係数列の相関係数との比較に用いるしきい値は 0 に固定した。今回のシミュレーションに用いた録音環境を表 4.1 に示す。

表 4.1 録音環境

録音機器	AUDIO TECHNICA/AT-VD3
録音場所	福本研究室

さらに、今回行ったシミュレーションについて表 4.2 に示す。

表 4.2 シミュレーション条件

	条件 1	条件 2	条件 3
混合音声	男性 A と女性	男性 A と男性 B	男性 A と男性 B と女性
所望音声	女性の音声	男性 B の音声	男性 B の音声

条件 1 のシミュレーションでは、男性 1 人と女性 1 人が別々に録音した音声を簡易合成し

4.1 シミュレーション条件

た音声と男性 1 人と女性 1 人が同時に発話した音声のそれぞれに対して提案システムを適用した。それぞれの音声の女性の音声を所望音声，男性の音声を非所望音声としてシミュレーションを行った。簡易合成した音声に対するシミュレーションにおける所望音声を図 4.1，非所望音声を図 4.2，2 つの音声を簡易合成して得られる音声を図 4.3 に示す。また，男性と女性が同時に発話した音声に対するシミュレーションにおける混合音声を図 4.4 に示す。それぞれの図の横軸はサンプル数，縦軸は振幅を示す。

条件 2 のシミュレーションでは，男性 2 人が別々に録音した音声を簡易合成した音声と男性 2 人が同時に発話した音声のそれぞれに対して提案システムを適用した。それぞれの音声の男性 B の音声を所望音声，男性 A の音声を非所望音声としてシミュレーションを行った。簡易合成した音声に対するシミュレーションにおける所望音声を図 4.5，非所望音声を図 4.6，2 つの音声を簡易合成して得られる音声を図 4.7 に示す。また，男性 2 人が同時に発話した音声における混合音声を図 4.8 に示す。それぞれの図の横軸はサンプル数，縦軸は振幅を示す。

条件 3 のシミュレーションでは，男性 2 人と女性 1 人が別々に録音した音声を簡易合成した音声と男性 2 人と女性 1 人が同時に発話した音声のそれぞれに対して提案システムを適用した。それぞれの音声の男性 B の音声を所望音声，男性 A の音声と女性の音声を非所望音声としてシミュレーションを行った。簡易合成した音声に対するシミュレーションにおける所望音声を図 4.9，非所望音声を図 4.10，図 4.11，3 つの音声を簡易合成して得られる音声を図 4.12 に示す。また，男性 2 人と女性 1 人が同時に発話した音声に対するシミュレーションにおける混合音声を図 4.13 に示す。それぞれの図の横軸はサンプル数，縦軸は振幅を示す。

4.1 シミュレーション条件

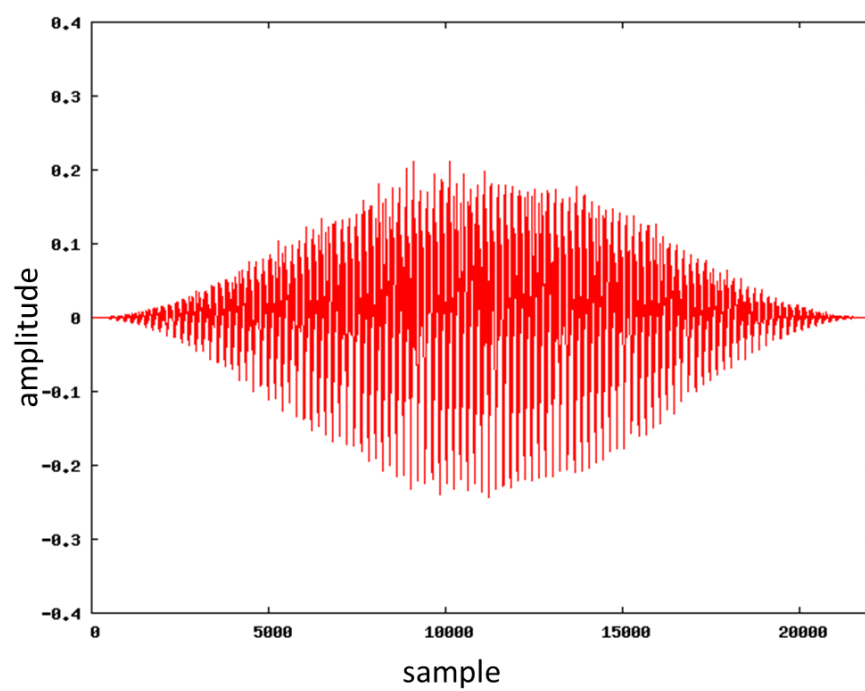


図 4.1 条件 1 における所望音声

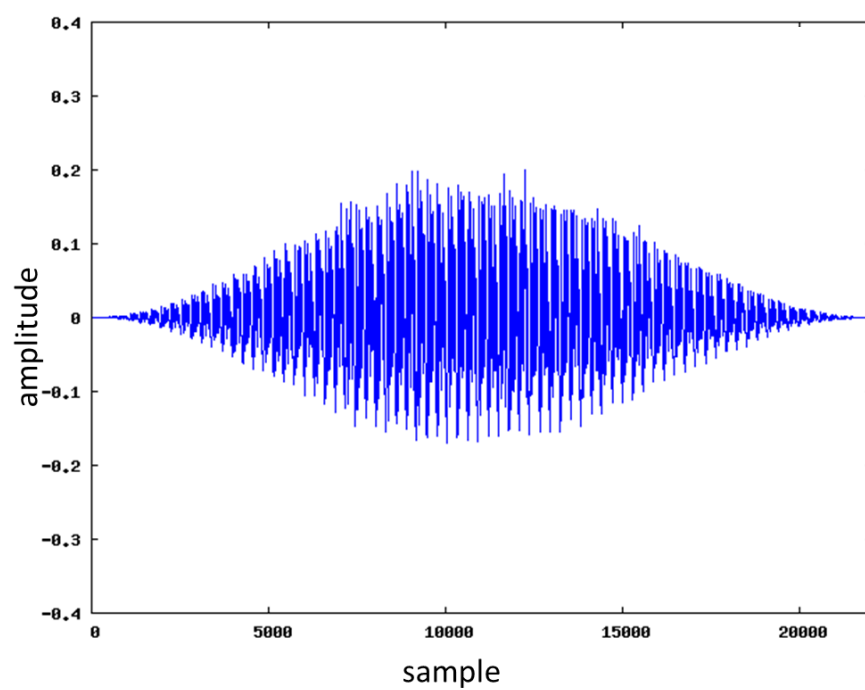


図 4.2 条件 1 における非所望音声

4.1 シミュレーション条件

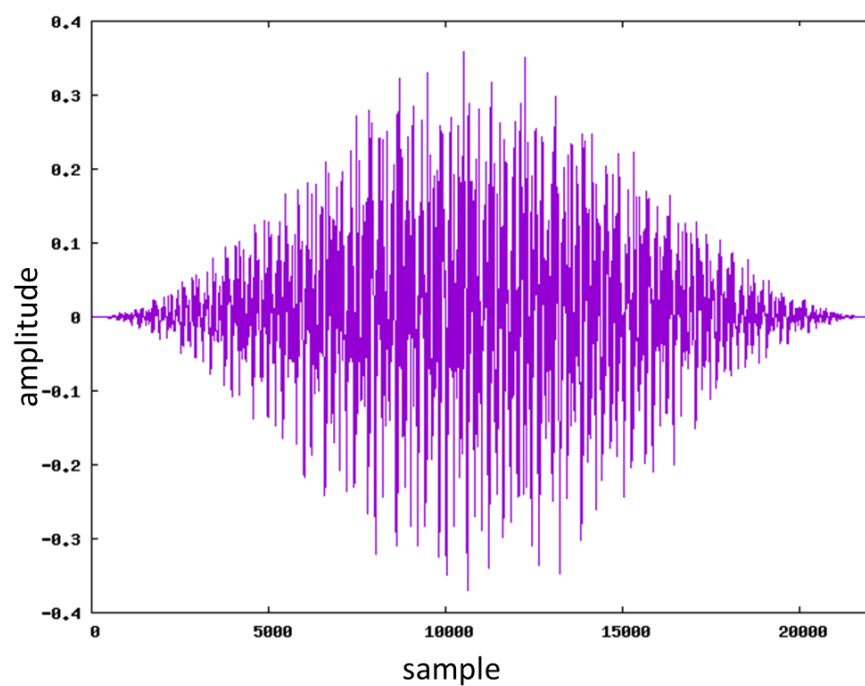


図 4.3 条件 1 における簡易合成音声

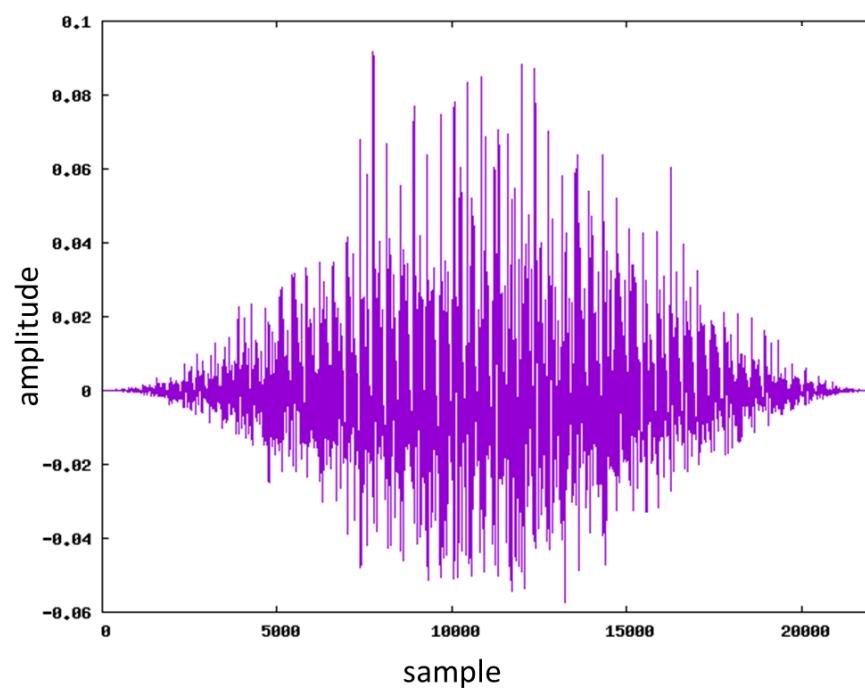


図 4.4 条件 1 における混合音声

4.1 シミュレーション条件

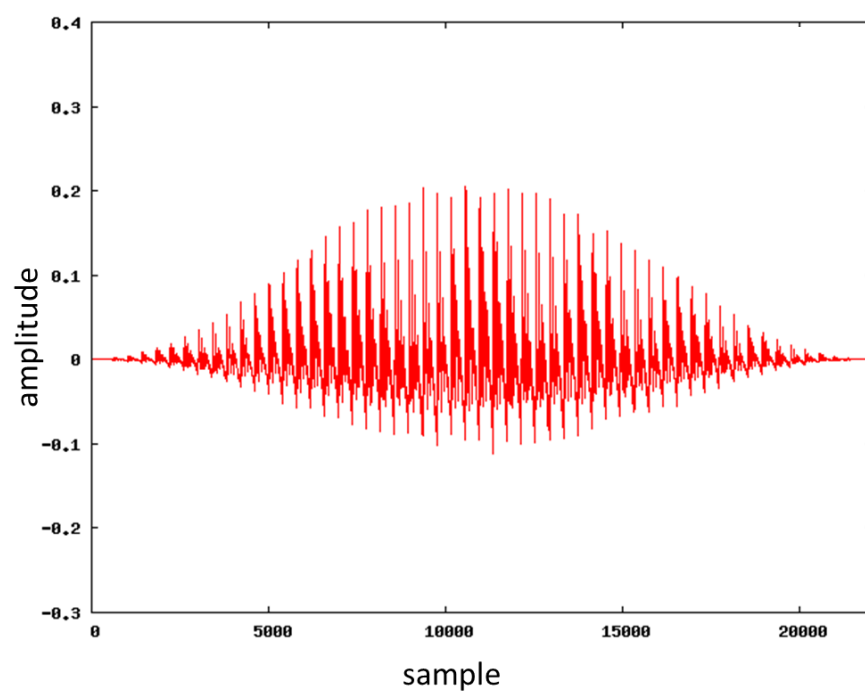


図 4.5 条件 2 における所望音声

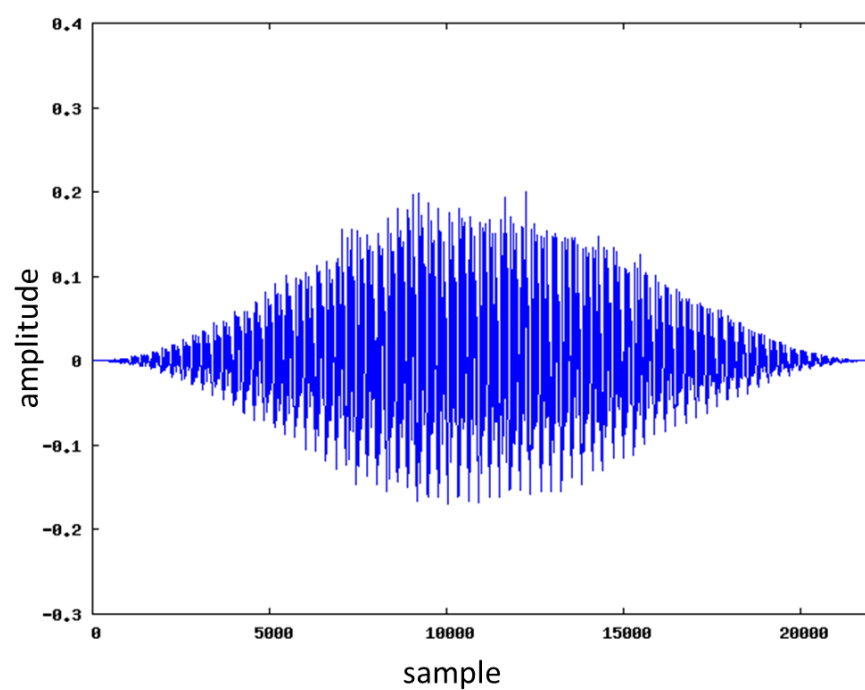


図 4.6 条件 2 における非所望音声

4.1 シミュレーション条件

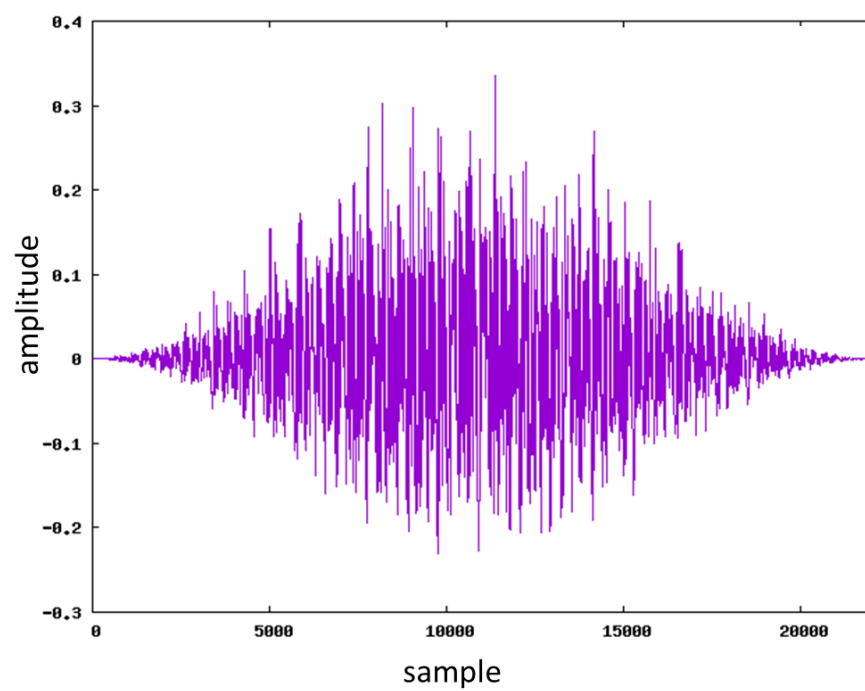


図 4.7 条件 2 における簡易合成音声

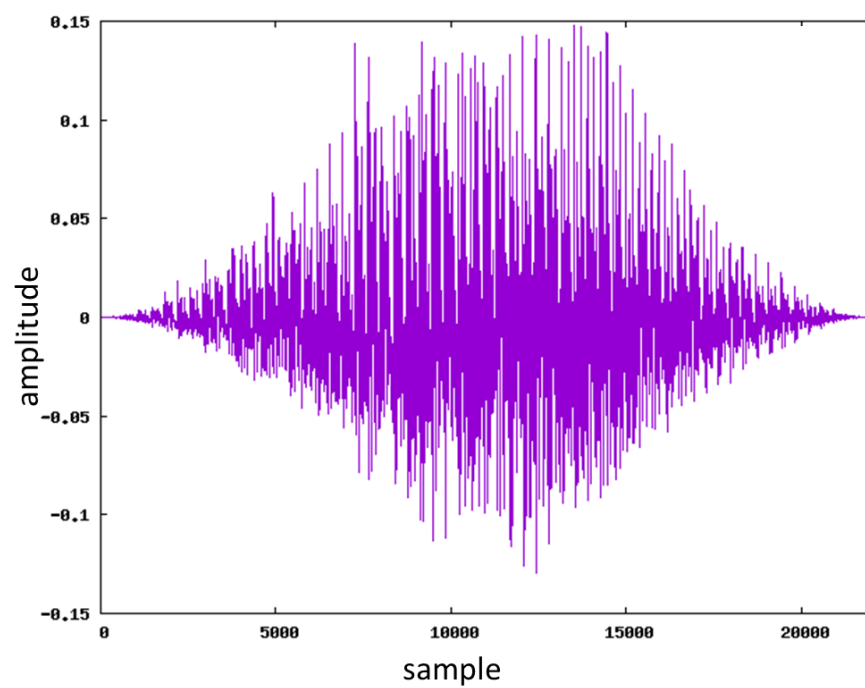


図 4.8 条件 2 における混合音声

4.1 シミュレーション条件

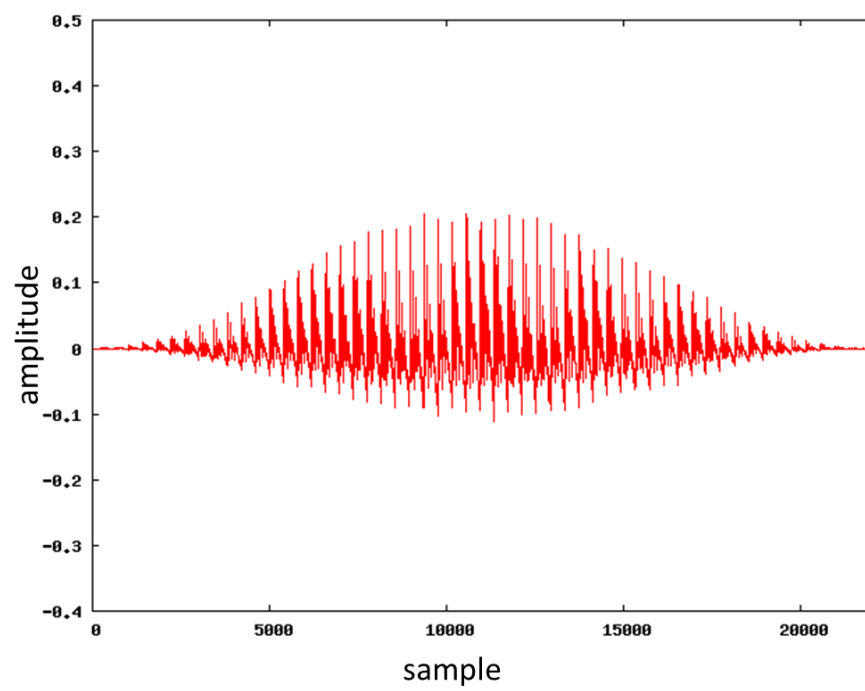


図 4.9 条件 3 における所望音声

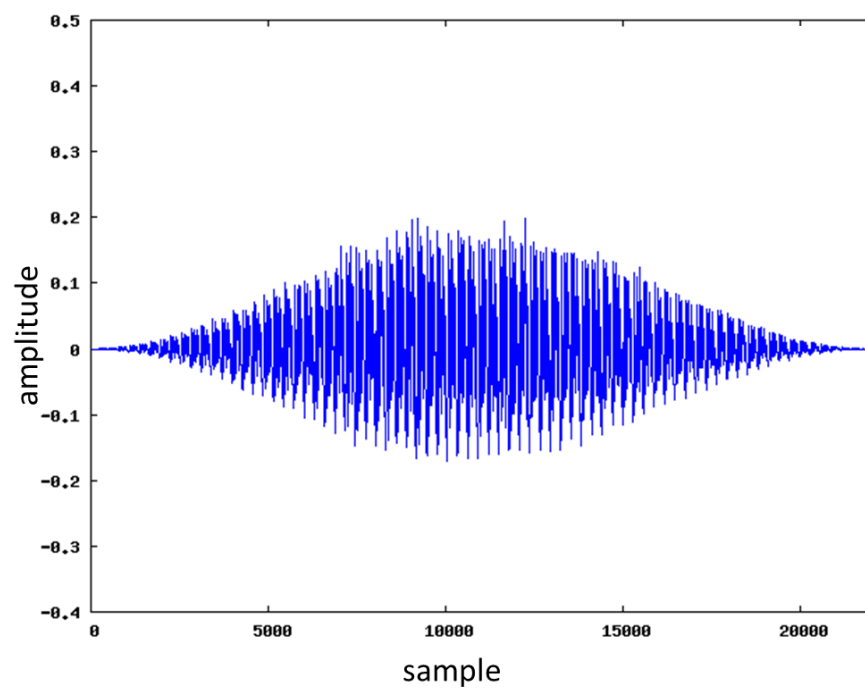


図 4.10 条件 3 における非所望音声 I

4.1 シミュレーション条件

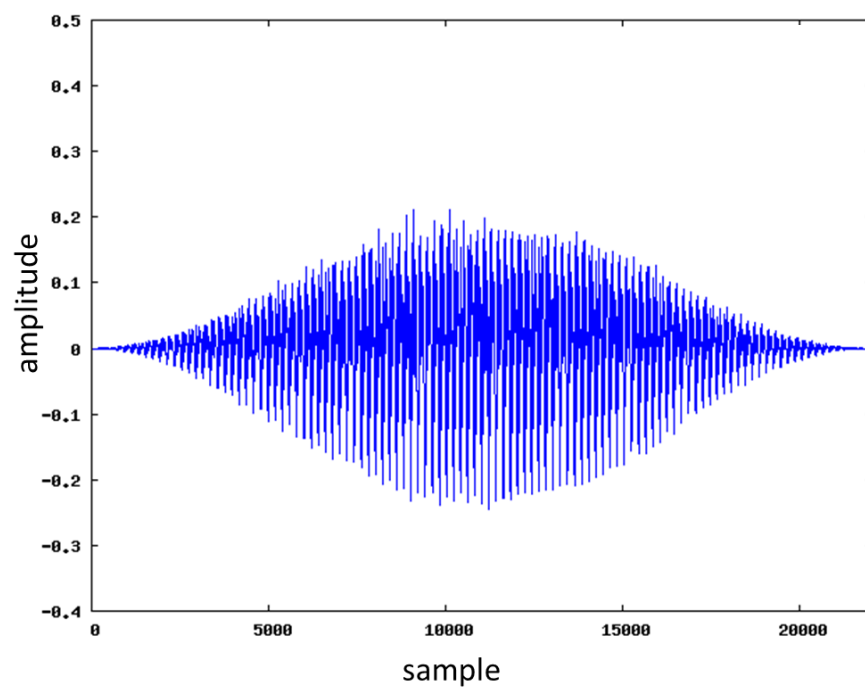


図 4.11 条件 3 における非所望音声 II

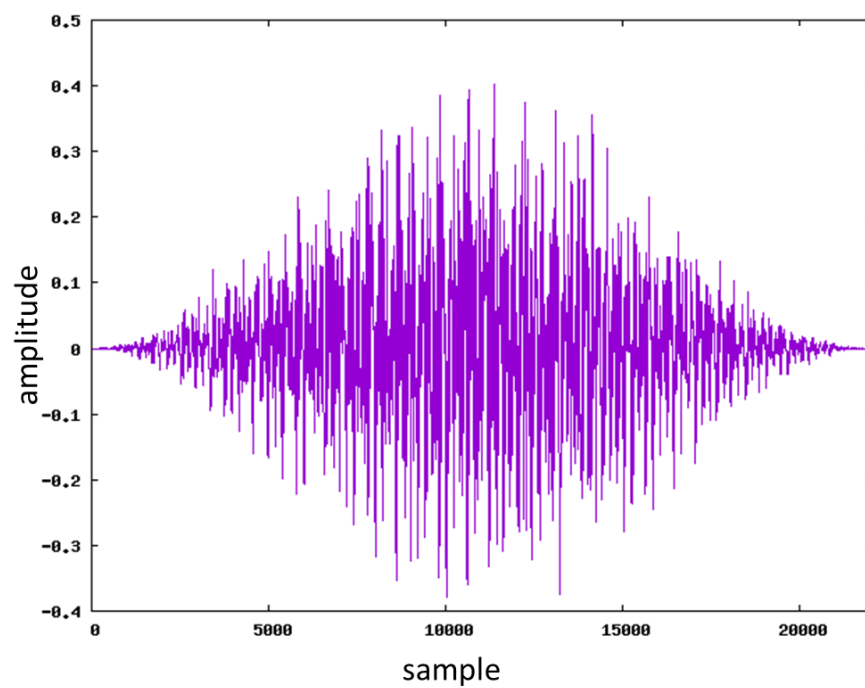


図 4.12 条件 3 における簡易合成音声

4.2 シミュレーション結果

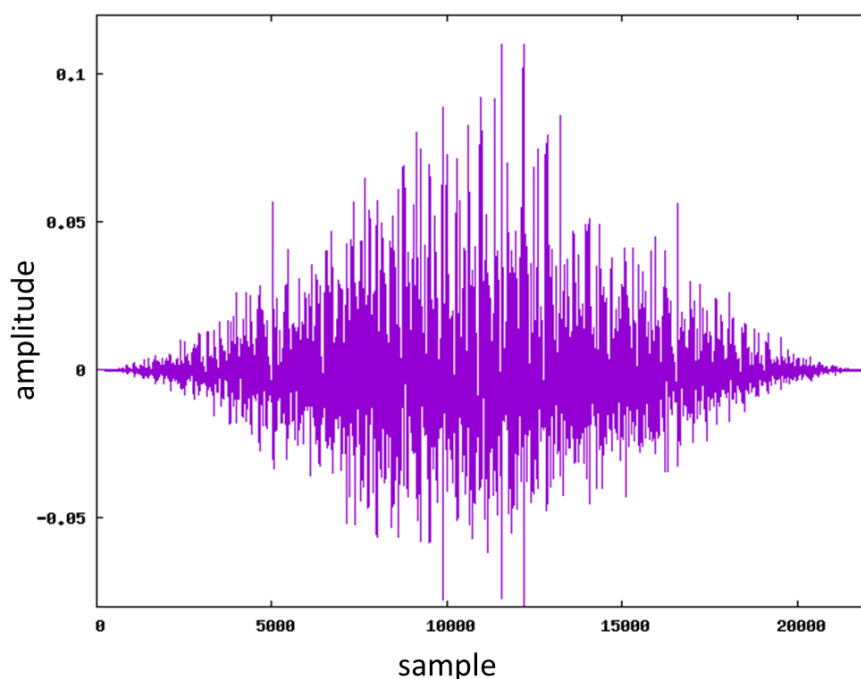


図 4.13 条件 3 における混合音声

4.2 シミュレーション結果

4.2.1 男女 2 人の混合音声 (条件 1)

男性と女性の音声の簡易合成音声に提案システムを適用した結果を図 4.14 に示す。図の横軸はサンプル数、縦軸は振幅を示す。抽出した音声を確認すると、所望の音声抽出されていることを確認できた。また、所望の音声と抽出した音声を比較した結果を図 4.15 に示す。図 4.15 の 2 種類の波形に大きな違いがないことから、所望の音声を抽出できたといえる。さらに、所望音声と提案システムを適用して得られた抽出音声の平均二乗誤差の値を以下の式を用いて求めた。

$$MSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{(x_i - y_i)^2}{y_i^2}}$$

ここで、 x_i を計算値、 y_i を真値とする。平均二乗誤差を求めた結果、 $-23[\text{dB}]$ であったことから十分に所望の音声抽出できたといえる。

混合音声に提案システムを適用した結果を図 4.16 に示す。抽出した音声を確認すると、

4.2 シミュレーション結果

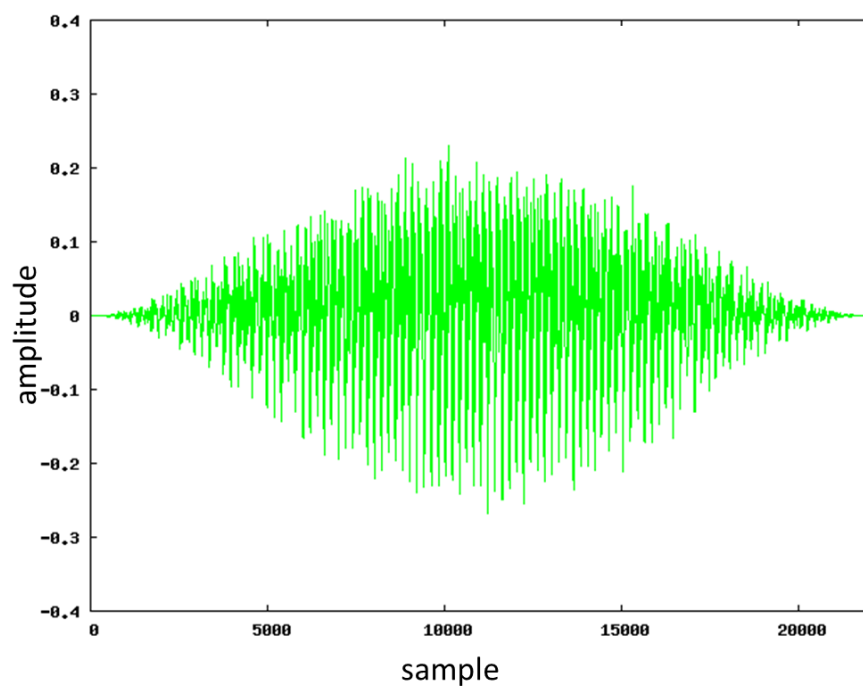


図 4.14 条件 1 における簡易合成音声から抽出した音声

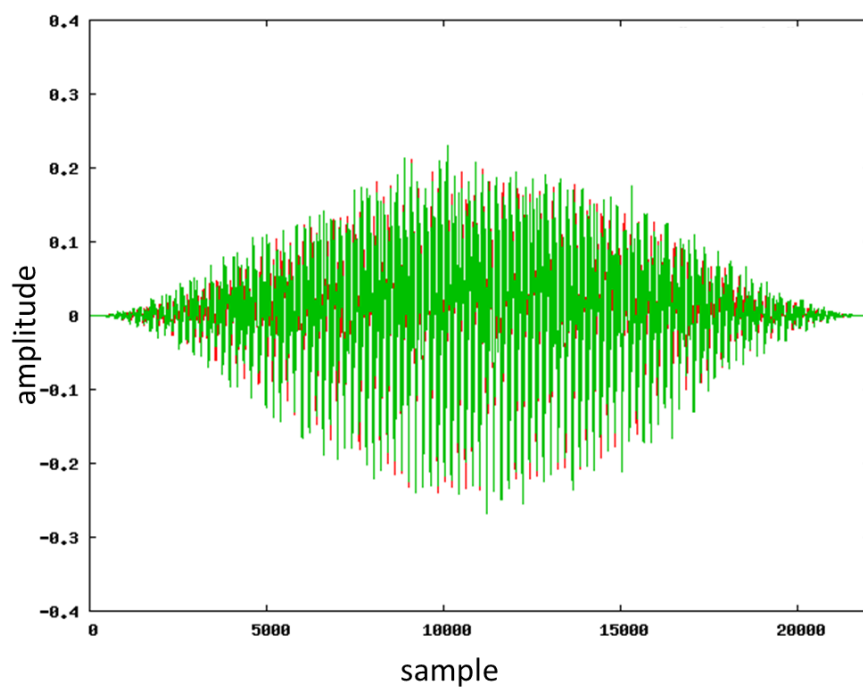


図 4.15 条件 1 における所望音声と抽出音声の比較

4.2 シミュレーション結果

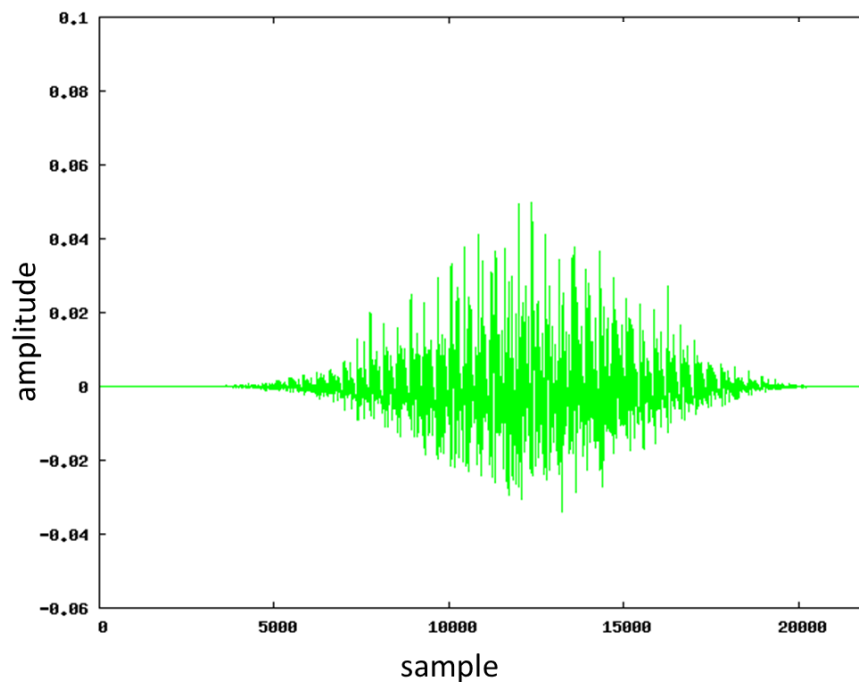


図 4.16 条件 1 における混合音声から抽出した音声

所望の音声抽出されていることを確認できた。しかし、非所望音声である男性の音声も聞こえることも確認できた。

4.2.2 男性 2 人の混合音声 (条件 2)

男性 2 人の音声の簡易合成音声に提案システムを適用した結果を図 4.17 に示す。図の横軸はサンプル数、縦軸は振幅を示す。抽出した音声を確認すると、条件 1 のシミュレーションと同様に所望の音声抽出できていることを確認できた。また、所望の音声と抽出した音声を比較した結果を図 4.18 に示す。2 種類の波形に大きな違いがないことから、所望の音声抽出できたといえる。さらに、所望音声と提案システムを適用して得られた抽出音声の平均二乗誤差の値が $-24[\text{dB}]$ であったことから十分に所望の音声抽出できたといえる。しかし、条件 1 と異なり抽出音声に少し非所望音声が含まれていることを確認できた。

混合音声に提案システムを適用した結果を図 4.19 に示す。抽出した音声を確認すると、所望の音声抽出できていなかった。抽出した音声と混合音声を聞き比べると、抽出した音

4.2 シミュレーション結果

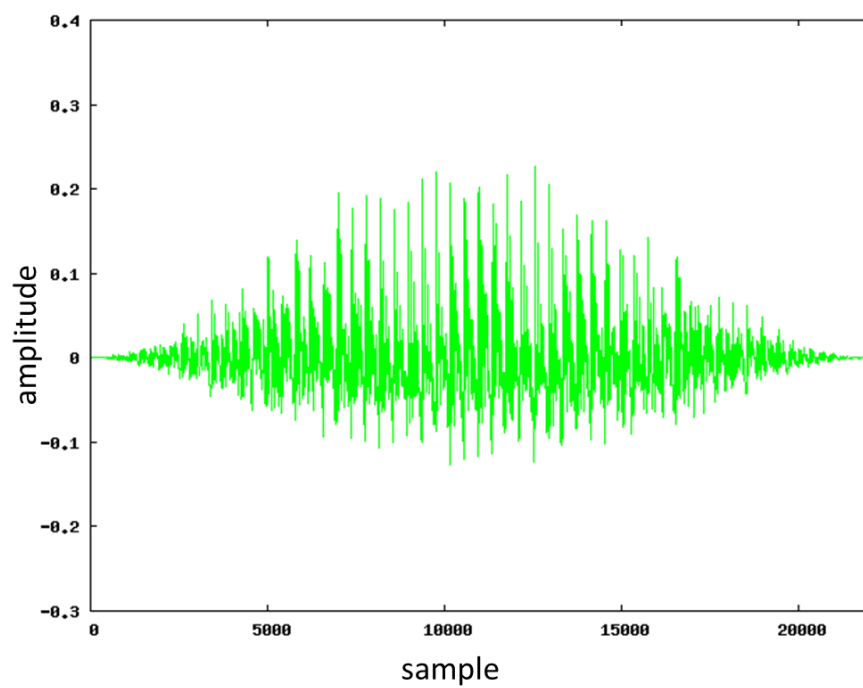


図 4.17 条件 2 における簡易合成音声から抽出した音声

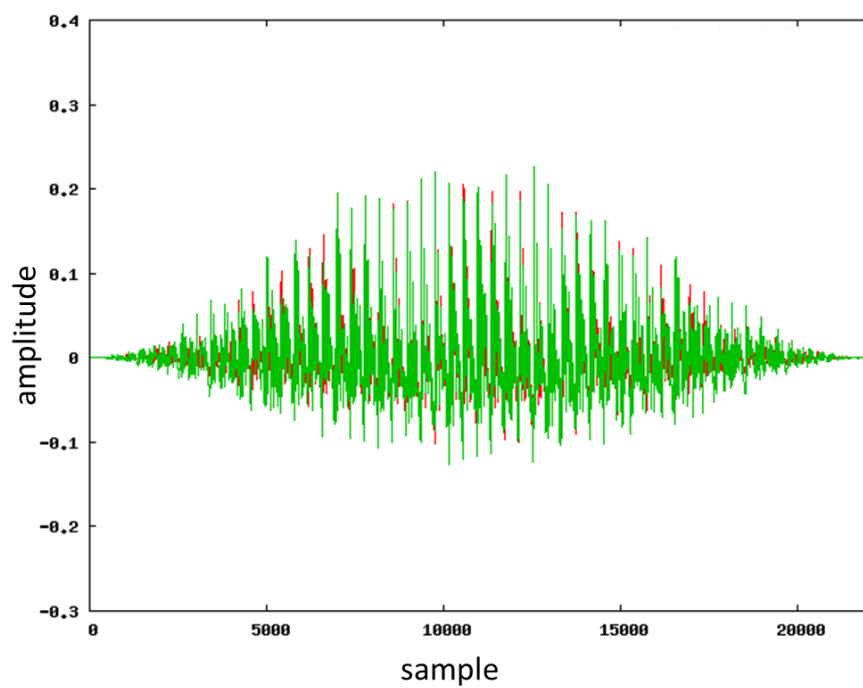


図 4.18 条件 2 における所望音声と抽出音声の比較

4.2 シミュレーション結果

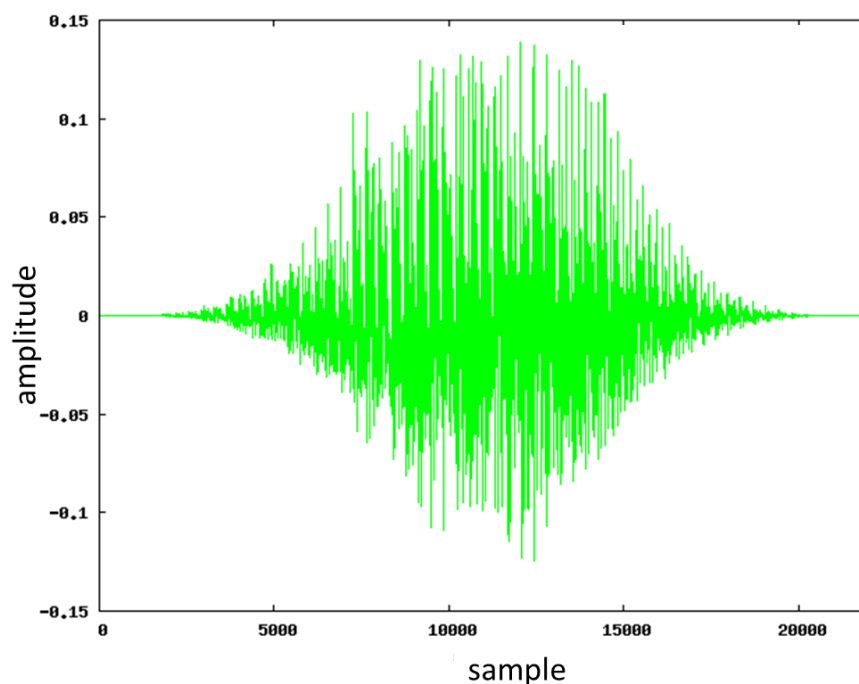


図 4.19 条件 2 における混合音声から抽出した音声

声内の非所望音声が少し小さくなっていることを確認できた。

4.2.3 男女 3 人の混合音声 (条件 3)

男性 2 人と女性 1 人の簡易合成音声に提案システムを適用した結果を図 4.20 に示す。図の横軸はサンプル数、縦軸は振幅を示す。抽出した音声を確認すると、所望の音声は抽出できていたが、非所望音声である男性 A の音声と女性の音声も確認できた。また、所望の音声と抽出した音声と比較した結果を図 4.21 に示す。2 種類の波形にあまり大きな違いがないことから、所望の音声は抽出できたといえる。さらに、所望音声と提案システムを適用して得られた抽出音声の平均二乗誤差の値が $-23[\text{dB}]$ であったことから十分に所望の音声は抽出できたといえる。しかし、抽出音声内に非所望音声が含まれていることも確認できた。

混合音声に提案システムを適用した結果を図 4.22 に示す。抽出した音声を確認すると、所望の音声は抽出できていなかった。抽出した音声と混合音声を聞き比べると、抽出した音声から所望の音声が消えているように聞こえた。

4.2 シミュレーション結果

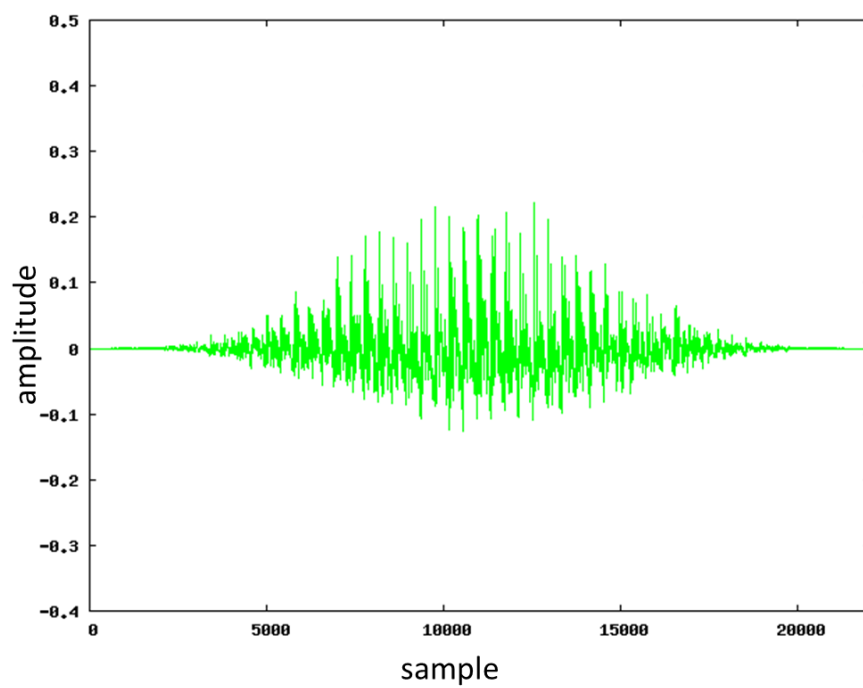


図 4.20 条件 3 における簡易合成音声音声から抽出した音声

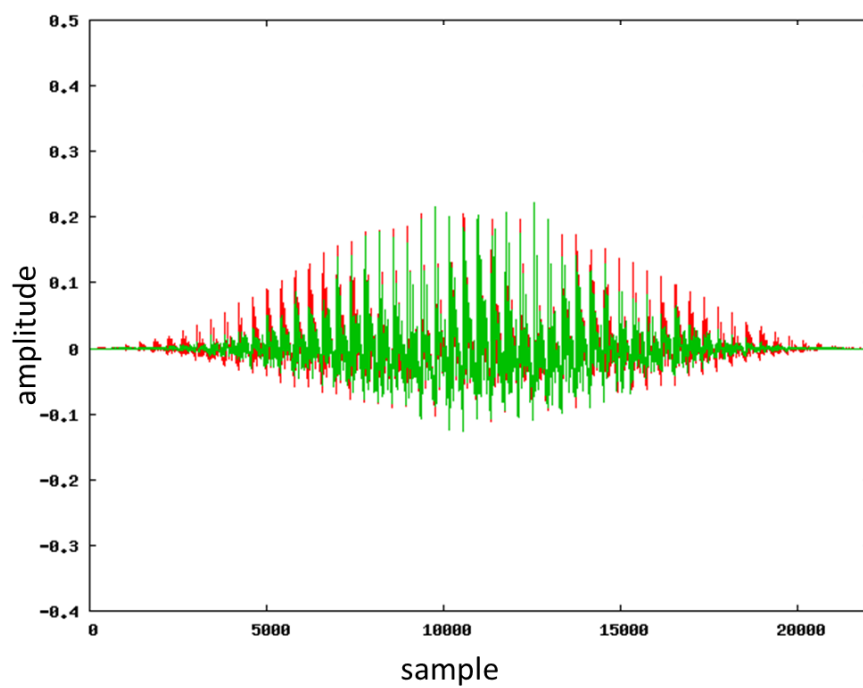


図 4.21 条件 3 における所望音声と抽出音声の比較

4.3 考察

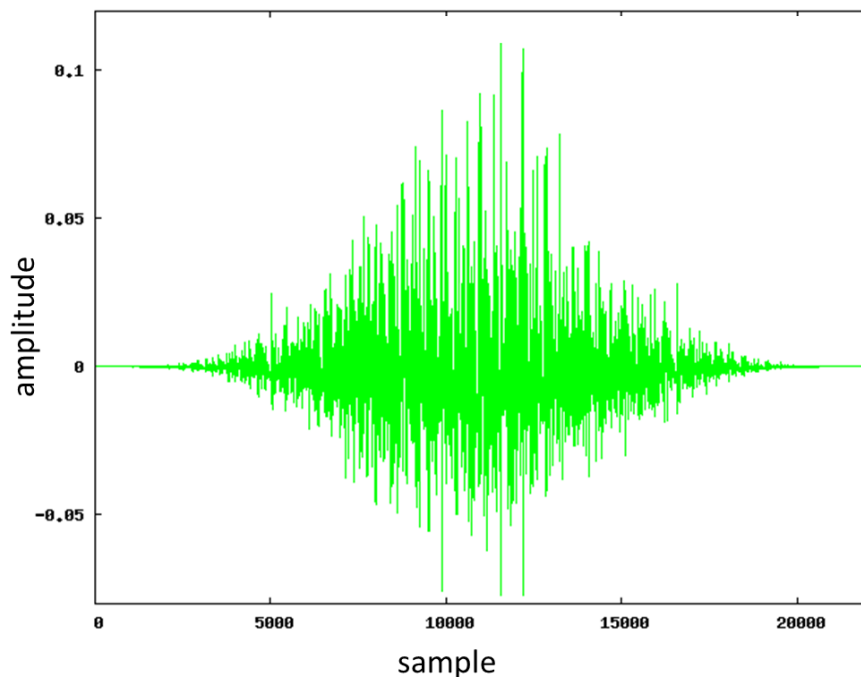


図 4.22 条件 3 における混合音声から抽出した音声

4.3 考察

2 種類のシミュレーションを 3 つの条件に対して行った結果、各条件の簡易合成音声に提案システムを適用した場合は所望の音声を抽出できることを確認した。しかし、条件 2 および条件 3 については非所望音声も確認できたため完全な分離はできなかった。一方、各条件の混合音声に提案システムを適用した場合は、条件 1 については所望の音声抽出できたが、条件 2 と条件 3 の場合は所望の音声抽出できなかった。また、条件 1 についても非所望音声を確認できたため、完全な分離はできなかったといえる。

条件 2 および条件 3 の簡易合成音声に対するシミュレーションにおいて非所望音声を除去できなかった原因としては、音声抽出処理のしきい値の値が適切でなかったことが考えられる。今回のシミュレーションでは、サンプルとする単一話者音声と複数話者音声の相関係数のしきい値を 0 に設定したが、男性同士の音声のように周波数帯域が重なる部分が多い場合は、しきい値の値を変更する必要がある。

条件 2 および条件 3 の混合音声に対するシミュレーションにおいて所望の音声を抽出でき

4.3 考察

なかった原因としては、サンプルとする単一話者音声の周波数帯域と混合音声に含まれるサンプルと同一話者の周波数帯域が異なっていることが挙げられる。そのため、条件 2 と条件 3 の混合音声から所望音声の抽出ができなかったと考えられる。このことから、条件 1 において所望の音声抽出できたのは、非所望音声とサンプルとする単一話者音声の周波数帯域に大きな差がなかったと推測できる。この問題は、混合音声に含まれる非所望音声の周波数帯域に合わせてサンプルを調整することができれば解決することができると予想される。また、条件 3 において所望の音声が消えているように聞こえたのは、所望の音声とサンプルの周波数帯域が近かった可能性が考えられる。

2 種類のシミュレーションにおいて、簡易合成音声に対して提案システムを適用した場合は所望音声既知であるため、所望の音声と抽出した音声の誤差を求めることで所望の音声抽出できたかどうかの評価を行った。しかし、混合音声に対して提案システムを適用した場合は抽出した音声を確認する以外に評価ができていない。これは、所望の音声未知であることから誤差の計算ができないためである。そのため今後は、混合音声に提案システムを適用した後に所望の音声抽出できたかどうかの評価を行うための指標が必要となる。

第 5 章

結論

5.1 本研究のまとめ

複数音声の音声認識率が低いことから、複数音声を分離する技術が必要である。音源分離技術の従来手法では、マイクの本数による問題点があった。そこで本研究では、単一マイクを用いることでこの問題点を解決した。また、音声の抽出方法としてウェーブレット係数を特徴量として用いることで所望の音声を抽出するシステムの提案を行った。提案した音声抽出システムの有効性を確認するためにシミュレーションを行った結果、男女 1 人ずつの混合音声から所望の音声を抽出できることを確認した。しかし、非所望音声も確認できたため、完全な分離はできなかった。一方、男性 2 人や 3 人以上の混合音声を用いたシミュレーションでは、所望の音声は抽出できなかった。これらは、抽出処理に用いたサンプルの周波数帯域が混合音声と重なっていなかったためと考えられる。このことから、混合音声に含まれる非所望音声の周波数帯域と抽出処理に用いるサンプルの周波数帯域が重なる場合は提案したシステムが有効であるといえる。

5.2 今後の課題

本研究では、複数の人の混合音声に対して提案システムを適用することで所望の音声は抽出できるかどうかの確認を行ったが、シミュレーションによって所望の音声を抽出できる音声と抽出できない音声があった。混合音声から所望の音声を抽出できた音声であっても非所望音声を確認できたため完全な分離はできていない。そのため、音声抽出処理における相関

5.2 今後の課題

係数と比較するしきい値の値を適切に設定する必要がある．また，音声の抽出方法としてウェーブレット係数同士の相関係数の値としきい値の比較を行ったが，別の分離方法の検討が課題である．また，所望の音声を抽出できなかった混合音声については，混合音声に含まれる非所望音声の周波数帯域に合わせて抽出処理に用いるサンプルを選択できるようになれば，提案システムが有効となる．

提案システムを用いたシミュレーションにおいて，簡易合成した音声については所望の音声は既知であったため，所望音声は抽出できたかどうかを平均二乗誤差を用いて評価することができた．しかし，混合音声については所望音声は未知であるため，所望音声は抽出できたかどうかの評価を音声を確認する以外の方法で評価することができなかった．今後は，抽出した音声は所望音声であるかどうかの評価方法の検討が課題である．

謝辞

本研究を行うにあたり，ご指導していただいた福本昌宏教授には感謝致します．本研究の副査をしていただいた浜村昌則准教授，吉田真一准教授にも感謝致します．また，福富英次氏にも感謝致します．

福本先生には，週次報告を全く出さない私に叱咤激励していただきここまで導いていただいたことに感謝しています．福本先生のお話をイヤホンをして聞いていないふりをしていたことは，今更ながらに申し訳ないと思いますが嫌なことは聞きたくないのでこのままやめることはないと思います．

卒業研究のみならず研究室生活の様々な面でお世話になった福富先生にも感謝いたします．忙しい中，論文添削や発表内容の確認などをしていただいて感謝してもしきれません．また，よく食事に連れて行ってくださって先輩方のおもしろいお話を聞かせていただいて楽しい時間を過ごすことができました．

一緒に研究に取り組んだ同学年の方々にも感謝致します．特に突っ走りがちな私を止めていただいた小田氏と田中氏は，3年生の頃から大変お世話になりました．4月の肉の日になんで始めた毎月恒例肉の日はとても楽しく，毎月の楽しみでした．最後の3ヶ月は忙しく肉の日を開催できなかったのは残念です．今後，4年生の皆さんはそれぞれの進路に進んでいくことになりますが，福本研究室で過ごしたことを糧にこれからもそれぞれ頑張っていくことができればいいと思います．

最後に，大学生活4年間支えてくださった皆様に感謝致します．

参考文献

- [1] 安藤厚志, 丹羽健太, 北岡教英, 武田一哉, “特徴量領域音源分離のためのクロススペクトル抑圧,” 電子情報通信学会, Vol.112, No.369, pp.107–112, Dec.2012.
- [2] 平井範之, 松本浩樹, 古川利博, 古屋清, “Wavelet 変換を用いたブラインド音源分離の提案,” 電子情報通信学会技術報告, Vol.104, No.456, pp.51–54, Nov.2004.
- [3] 関麻子, 渡部英二, “群遅延に基づく DUET ブラインド音源分離,” 電子情報通信学会技術報告, Vol.107, No.476, pp.55–60, Feb.2008.
- [4] 荒木章子, 澤田宏, 牧野昭二, “音声のスパース性を用いた Underdetermined 音源分離,” 信学会総合大会基礎・境界講演論文集, AS-4-5, 2008.
- [5] 足刈昇平, 川村新, 飯國洋二, “単一話者区間を利用したヒストグラム作成が不要な DUET ブラインド音源分離,” 電子情報通信学会, Vol.111, No.27, pp.113–118, May.2011.
- [6] 牧野昭二, 荒木章子, 向井良, 澤田宏, “独立成分分析に基づくブラインド音源分離,” 電子情報通信学会, Vol.103, No.129, pp.17–24, Jun.2003.
- [7] 猿渡洋, “音声・音響信号を対象としたブラインド音源分離,” 電子情報通信学会, Vol.101, No.667, pp.59–66, Mar.2002.
- [8] 中野宏毅, 山本鎮男, 吉田靖夫, “ウェーブレットによる信号処理と画像処理,” 共立出版株式会社, 2000.
- [9] 新井康平, “ウェーブレット解析の基礎理論,” 森北出版株式会社, 2000.