

# 音声強調を用いたホームビデオの音質改善

## Improvement of the Home video Sound Quality Using Speech emphasis

栗原 佑季

Yuki Kurihara

法政大学情報科学部デジタルメディア学科

E-mail:yuki.kurihara.va@cis.hosei.ac.jp

### Abstract

The noise by video camera is often included in the sound of home video, and its quality is controlled by environment of filming. Noise of regularity and others mixes with sound by environment of filming, sometimes the sound of necessity actually is not caught easily. It was done of improvement of the quality of home video using speech emphasis.

First, various problems of the recorded speech were described, and the Spectral Subtraction by Boll.S was introduced. Moreover, window function was described used for Fourier transform of Spectral Subtraction.

It was done of speech emphasis using some windows as window function of Fourier transform of Spectral Subtraction. Two methods for reducing Musical Noise were described. That is the noise reduction using spectrum of time frames, and noise reduction using spectrum of frequency frames. It was decided that using blackman window is better as window function of Fourier transform of Spectral Subtraction. As for the reduction of the musical noise, the noise reduction using spectrum of time frames was a great result. Finally, result was almost the same though the video camera was different.

### 1 まえがき

現在、ホームビデオの撮影では、主として家庭用デジタルビデオカメラが用いられる。8ミリビデオカメラなどのアナログのデジタルカメラから、デジタルビデオカメラへの転換と発展により、画質や音質が大きく改善された。しかし、録画された動画の音声には、ビデオカメラによるノイズが含まれていたり、撮影環境や音場によって品質が大きく左右される。環境によっては、風や空調、機械のモーター音のような雑音、残響音によるノイズなどが混ざり、実際に録音したかった音が聞き取りにくい場合がある。本研究では、このような実際に必要な音や聞きたい音が聞こえにくいというホームビデオにおける音質の問題点を改善するために、これらの音を低減または取り除き、必要な音のみを抽出することで、動画の音声の音質改善の実現を目指す。

雑音を除去するためにノイズキャンセラが必要である。録音の際に、単一マイクのみを用いる方式と、複数のマイクを用いて、音声と雑音を集音するマイクのほかに雑音のみを集音するマイクを用いる方式がある [4]。利便性と小型化において考えた場合、単一マイクのみを用いる方式のほうが有用であると考えられる。また、過去に録画された動画は、このような聞き取りにくい音が含まれていたとしても、再び同じものを撮影するのは困難である。すなわち、過去に録画された動画においては、マイク口ホーンに風防を取り付けたり、撮影環境を変えたり、音場を制御するといったことができないため、デジタル信号処理により音質改善を行う必要がある。

なお、あくまでホームビデオの音質改善を行うことが目的であるので、低コストで実現できることが望ましいが、デジタル音声

処理を行うことで実現できると考えられる。

今回処理を加えるのは動画の音声部分のみであるために、まず動画の映像と音声を分離し、音声のみを抽出する。抽出した音声より問題点の検証を行い、音質を改善するプログラムを作成する。音声に処理を行い、処理後の音声と元の映像を合成して、動画の音声の音質改善を目指す。

### 2 録音された音声の様々な問題点と音質改善の手法

#### 2.1 録音された音声に見受けられる問題点

ホームビデオの音声における問題点として、雑音挙げられるが、1. ビデオカメラによるノイズ、2. 風や空調、機械のモーター音、残響音によるノイズなどの撮影環境によるほとんど定常的な雑音、3. 突発的な音に分類される。定常的な雑音がほとんど感じられない環境で撮影された動画においても、サーというような定常的な小さい音が確認された。これは、ビデオカメラで撮影することにより発生したノイズであると考えられる。撮影環境における定常的な雑音について、この音は実際に定常なわけではないが、環境により多くの場合、ほとんど同じような音に聞こえ、近似して一定と考えることができる。突発的な音とは、ノイズとは異なるが、前者以外の音で聞きたい音とは異なる。録音の際に入り込んでしまった音のことである。環境により様々な音が考えられるが、車の通る音、人の声、足音などが挙げられる。

#### 2.2 スペクトル減算法

発話におけるノイズ(雑音)は、デジタル音声プログラムにより軽減することができる。それを行うものとして、スペクトル減算法が挙げられる [2]。

ノイズ抑制のスペクトルの推定量は、ノイズを含む発話で、発話のない状態におけるノイズスペクトルの推定量の減算により得られる。

#### 2.3 スペクトル減算法による音質改善を行うにあたっての課題

本研究では、ビデオカメラによるノイズ、風や空調や機械のモーター音、残響音によるノイズなどの定常的な雑音をスペクトル減算法を用いて、除去または低減することでその音質改善を目指す。しかし、定常的な雑音を含む音声にスペクトル減算法を用いた雑音除去の処理を行うと、機械音声のような音(ミュージカルノイズ)が確認されることがある。

また、突発的な音(実際に聞きたい音以外の音)に関しては、多くの場合で実際に聞きたい音よりも音量が小さく、それを仮定すると、比較的音量の小さな音を軽減することで音質改善が可能であると考えられる。しかし、実際に聞きたい音よりも突発的な音のほうが音量が大きい場合や、音量にあまり明確な差が感じられない場合は、実際に聞きたい音と雑音の区別が判断しにくく、聞きたい音と雑音をとらえ違えるなど、間違った音声処理を行ってしまう原因となる。突発的な音の音質改善においては、このように、実際に聞きたい音と雑音の区別を正しく判断することが課題となる。

#### 2.4 スペクトル減算法を用いた音質改善手法

前節に述べたように、スペクトル減算法とは、発話におけるノイズ(雑音)を音声信号処理により軽減する手法である。ある時

間枠の発話信号  $s(k)$  に付加されたノイズの信号を  $n(k)$ , またその合計を  $x(k)$  として,

$$x(k) = n(k) + s(k) \quad (1)$$

となり, ノイズを含む発話からノイズスペクトルの推定量を減算したものが, 発話スペクトルの推定量であるため,

$$\hat{S}(k) = X(k) - \hat{N}(k) \quad (2)$$

となる.

#### 2.4.1 スペクトル減算法における窓関数の使用について

窓関数は, ある有限区間以外が 0 となる関数である. ある無限区間数列  $h_d(n)$  が有限長のインパルス応答を得るためには,  $h(n)$  を簡単に打ち切ることが必要である. そのため, 無限区間数列  $h_d(n)$  に, ある有限区間以外が 0 となる窓関数を掛け合わせることで,  $h(n)$  を得ることができる. すなわち, 窓関数を  $w(n)$  とすると,

$$h(n) = h_d(n)w(n) \quad (3)$$

となる. 従って, フーリエ変換後の値は元の値の近似値となり, 窓関数によって定まる. そのため, 異なる窓関数を用いることで, スペクトル減算法を行った後の音声は変わることになる.

#### 2.4.2 代表的な窓関数

FIR フィルタ設計に用いられる代表的な窓関数として, 方形窓, ハミング窓, ハニング窓, パートレット窓, ブラックマン窓などが挙げられる. これらの窓は, 以下の方程式で与えられる.

方形窓:

$$\omega(n) = 1, \quad 0 \leq n \leq N-1 \quad (4)$$

パートレット窓:

$$\omega(n) = \begin{cases} \frac{2n}{N-1}, & 0 \leq n \leq \frac{N-1}{2} \\ 2 - \frac{2n}{N-1}, & \frac{N-1}{2} \leq n \leq N-1 \end{cases} \quad (5)$$

ハニング窓:

$$\omega(n) = \frac{1}{2} [1 - \cos(\frac{N-1}{2\pi n})], \quad 0 \leq n \leq N-1 \quad (6)$$

ハミング窓:

$$\omega(n) = 0.54 - 0.46 \cos(\frac{2\pi n}{N-1}), \quad 0 \leq n \leq N-1 \quad (7)$$

ブラックマン窓:

$$\omega(n) = 0.42 - 0.5 \cos(\frac{2\pi n}{N-1}) + 0.08 \cos(\frac{4\pi n}{N-1}), \quad 0 \leq n \leq N-1 \quad (8)$$

各窓関数の形状は図 1 のようになる. これらの窓関数は, 図 2 のような振幅特性をもつ. これらのグラフは, それぞれの窓関数  $w(n)$  に対し,  $20 \log_{10} |W(e^{j\omega})|$  とする. 角周波数 0 近くの一番大きな凸を主極, それ以外の凸を副極と呼ぶ. 振幅特性を見たときに, 主極の幅が広いと, 振幅が等しいが周波数が隣接する複数の信号成分を分離するには適さない. また, 角周波数が 0 から離れても副極のエネルギーが大きい (副極の減衰が小さい) と, 周波数が離れた微弱な信号を検出するには適さない. ここで, 五つの窓関数の振幅特性を比べてみると, 主極の幅については, ブラックマン窓 > ハミング窓 > ハニング窓 > パートレット窓 > 方形窓 となり, 副極のエネルギーの大きさについては, 方形窓 > ハミング窓 > パートレット窓 > ハニング窓 > ブラックマン窓 となる. ここで本研究で行う雑音除去を考えると, 周波数が離れた微弱な信号を検出する必要があるため, 副極のエネルギーが小さい (副極の減衰が大きい) 窓関数を用いるのが好ましいと考えられる. そのため, スペクトル減算法を用いた音質改善で使用する窓関数として, ブラックマン窓が好ましいと考えられる.

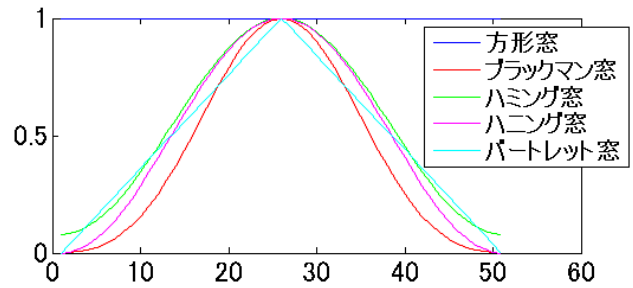


図 1. 代表的な窓関数の形状

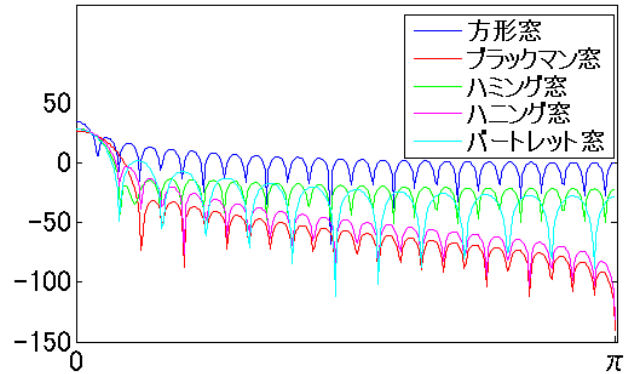


図 2. 代表的な窓関数の振幅特性

#### 2.5 ミュージカルノイズの軽減手法

ミュージカルノイズは, スペクトル減算法のノイズスペクトルの推定誤差による雑音の引き残り成分により発生する機械音のような音である. 雑音除去処理を行った後に残るミュージカルノイズは, 推定誤差であるため, 実際に必要な音に比べて微弱であると考えられる. よって, ある時間の相対的に小さな周波数成分がミュージカルノイズである可能性がある. また, 音声の倍音の必要な部分よりも微弱である可能性がある. そのため, ミュージカルノイズの軽減手法として, 時間フレームのスペクトルを用いた雑音除去法と, 周波数フレームのスペクトルを用いた雑音除去法を提案する.

##### 2.5.1 時間フレームのスペクトルを用いた雑音除去法

ある時間における相対的に小さな周波数成分除去を行う手法 (処理 1 とする.) を提案する. まず, 各時間フレームのスペクトルを求める. 次に, 音量に閾値を設ける. 閾値は, 各スペクトルの音量の最大値  $max$  に, 同じレート  $rate$  をかけたものとする. 各スペクトルの閾値以下をカットするために, スペクトルから閾値を減算する. 次に, 処理前と処理後の最大値を等しくように乗算を行う. その全フレーム分を足し合わせることで, 時間フレームのスペクトルを用いた雑音除去を行う.

$f$  を周波数軸のフレーム,  $t$  を時間軸のフレームとして,

$$X(f, t) = (X(f, t) - (rate \times max)) / (1 - rate) \quad (9) \\ 0 \leq rate \leq 1 \\ \text{if } X(f, t) \geq 0 \text{ then } X(f, t) = 0$$

##### 2.5.2 周波数フレームのスペクトルを用いた雑音除去法

信号から, ある周波数のスペクトルを求める. スペクトルの音量に閾値を設け, 閾値以下の部分の除去を行う. 閾値設定と処理方法においては, 時間フレームスペクトルを用いた雑音除去法と同じとする. 各周波数フレームに処理を行い, 全フレーム分を足し合わせることで, 周波数フレームのスペクトルを用いた雑音除去 (処理 2 とする.) を行う.

### 3 実験方法

#### 3.1 実験の流れ

まず、撮影した映像と音声とを分離し、音声のみを抽出する。これをサンプル音源として、雑音除去の処理を行った。処理後の音声と元の映像とを合成することで、ホームビデオの音質改善を目指す。

雑音除去の処理として、サンプル音源に異なる窓関数を用いてフーリエ変換を行う、スペクトル減算法を用いた音質改善手法を行う。その後、ミュージカルノイズの軽減手法として提案した2つの手法を行う。

スペクトル減算法を用いたプログラムのフーリエ変換では、窓関数として前節に述べた方形窓、パートレット窓、ハミング窓、ハニング窓、ブラックマン窓を用いた。また、使用するビデオカメラが違くと、音声も異なることが考えられる。そのため、二つのビデオカメラを並べて設置して撮影を行い同じ音声区間を抽出したものに、スペクトル減算法とミュージカルノイズの軽減手法を行った。

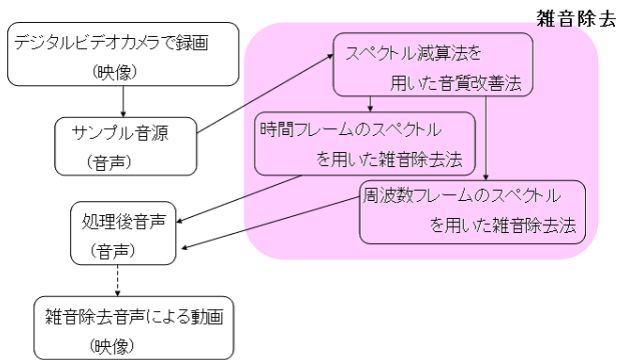


図 3. 実験の流れ

#### 3.2 実験に用いたサンプル

今回実験に用いたサンプルについて記述する。デジタルビデオカメラで撮影した動画の、一部区間を抽出した音声をサンプルとした。屋内で撮影した動画7つに加えて、屋外で撮影した動画を1つ用いた。また、これらのサンプルは、発話(人の声)をメイン音声とし、また、人の声やビデオカメラによるノイズなどを雑音とする。

#### 3.3 閾値設定について

今回提案したミュージカルノイズ軽減手法では、各スペクトルにおける閾値を、 $max \times rate$  と定めた。そのため、雑音除去を行う際に各スペクトルの最大値に対する  $rate$  の値を手動で設定する必要がある。設定方法としては、元の音を歪ませない程度に  $rate$  の値を大きくし、様々な値で処理を行い、聞きとりやすかったものとした。

#### 3.4 実験結果の評価方法

窓関数の違いによるスペクトル減算法を用いた音質改善法、ミュージカルノイズの軽減法を行った音声の聞こえ方と、ビデオカメラの違いによる処理後音声の違いを評価するために、主観評価を用いた。スペクトル減算法においては、元の音から雑音が除去されているかに対し5段階評価(除去されている場合は5)、元の音声が聞こえにくくなっていないかに対し5段階評価(聞こえにくくなっていない場合は5)を行った。ミュージカルノイズ軽減手法では、スペクトル減算法実行後の音声よりミュージカルノイズが軽減されているかについて5段階評価(軽減されている場合は5)、スペクトル減算法実行後の音声より元の音声が聞こえにくくなっていないかについて5段階評価(聞こえにくくなっていない場合は5)を行った。また、ビデオカメラの違いによる処理後音声の違いについては、評価方法をミュージカルノイズの軽減手法と同じとして、比較を行った。窓

関数の違いによるスペクトル減算法を用いた音質改善法については、実行後の音声に大きい違いがみられなかったため、評価人数を一人とした。また、ミュージカルノイズの軽減手法においては、評価人数を四人、ビデオカメラの違いによる処理後音声の違いについては四人とした。

### 4 実験結果と考察

異なる窓関数を用いたスペクトル減算法による音質改善手法と、先に提案したミュージカルノイズの軽減手法の実験結果を示す。また、最後にビデオカメラの違いによる処理後音声の違いについての比較実験の結果を示す。

#### 4.1 スペクトル減算法を用いた音質改善手法の実験結果と考察

スペクトル減算法を用いた音質改善手法の、窓関数の違いによる比較を行った(表1)。スペクトル減算法を用いた音質改善手法では、窓関数として方形窓を使った場合、音が割れるなどして総じて音質が悪いことが確認できた。主観評価(表1)でも、各サンプルの平均が五段階評価で、元の音声の再現性について1.0、雑音除去について2.9となり、スペクトル減算法を用いた音質改善手法において使用するのには好ましくないと考えられる。

また、各窓関数においては、サンプル、窓関数によって音質が違い若干の聞き取りやすさ、聞き取りにくさはあるが、大きな違いがなく、主観評価(表1)で、各サンプルの平均が4.6(雑音除去について)と2.0(元の音声について)でブラックマン窓が一番良い結果となったが、卓越してブラックマン窓が良いという結果にはならなかった。しかし、ほとんどのサンプル(サンプル4,5以外)において、ブラックマン窓が良い評価となった(表1)。

元の音声が悪くないかについての各サンプルの平均評価を評価が良い順に並べると、ブラックマン窓 > パートレット窓 > ハニング窓 > ハミング窓 > 方形窓となる。これは、振幅特性の副極のエネルギーの大きさが方形窓 > ハミング窓 > パートレット窓 > ハニング窓 > ブラックマン窓であったの(図2)の反対と類似した結果となった。つまり、副極のエネルギーが小さいもの(副極の減衰が大きいもの)のほうが、雑音除去の際のフーリエ変換で用いる窓関数として好ましいということが確認できた。そのため、スペクトル減算法を用いた音質改善手法においては、窓関数としてブラックマン窓を使用することとした。実行結果を図5に示す。

表 1. 窓関数による音質改善法比較の主観評価, A: 方形窓, B: ブラックマン窓, C: ハミング窓, D: ハニング窓, E: パートレット窓

サンプル   窓	雑音除去効果					元の音声				
	A	B	C	D	E	A	B	C	D	E
1	4	4	4	4	4	1	5	5	4	5
2	3	5	4	4	4	1	1	1	1	1
3	3	4	3	3	4	1	2	1	1	1
4	3	5	4	3	3	1	1	1	2	1
5	2	4	4	3	4	1	1	1	2	2
6	2	5	4	3	4	1	2	1	1	1
7	3	5	5	5	4	1	2	1	1	2
8	3	5	5	5	4	1	2	1	1	1
平均	2.9	4.6	4.1	3.8	3.9	1	2	1.5	1.6	1.8

#### 4.2 ミュージカルノイズ軽減手法の比較実験の結果と考察

スペクトル減算法を用いた音質改善手法(ブラックマン窓)の後に二つのミュージカルノイズ軽減手法を行ったところスペクトログラムは図6,7のようになった。また、二つの手法の評価実験の結果を示す(表2)。

時間フレームのスペクトルを用いた雑音除去法では、スペクトログラムで水色、薄い黄色であったところ(図5)が濃い青色になっていることから(図6)、小さい雑音の除去が確認できた。しかし、ミュージカルノイズと考えられる音(0.00秒, 1.08秒, 2.10秒, ... あたりの音)の除去はほとんど行われておらず、音声でもミュージカルノイズに若干の低減はみられるが、まだ残ってし

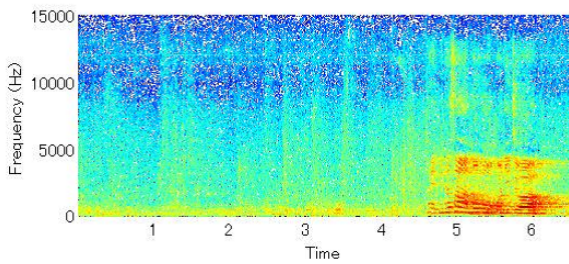


図 4. サンプルのスペクトログラム

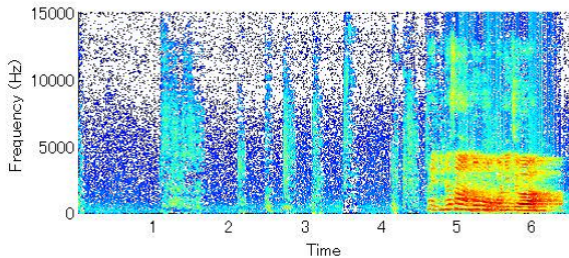


図 5. スペクトル減算法を用いた音質改善手法の実行結果

まっていた。また、実際に必要な音である、人の声の倍音の高周波 (6000Hz 以上の音) にも若干の影響がでていて、高周波であるため耳で聞いたときに大きくわかるわけではないが、スペクトログラム (図 6) から若干歪んでいると考えられる。周波数フレームのスペクトルを用いた雑音除去法では、ミュージカルノイズとみられる音声の低周波 (5000Hz 以下) 部分が大きく除去されているのが確認できる (図 7)。また、人の声などの必要な音についてもほとんど影響がないことが確認できる。音声では、ミュージカルノイズが大きく軽減されていることが確認できた。また、主観評価 (表 2) から、周波数フレームのスペクトルを用いた雑音除去法の各サンプル平均が 3.4 (ミュージカルノイズの軽減について) と 4.0 (元の音声が悪くなっていないかについて) となり、ミュージカルノイズの軽減手法として有用である、と考えられた。

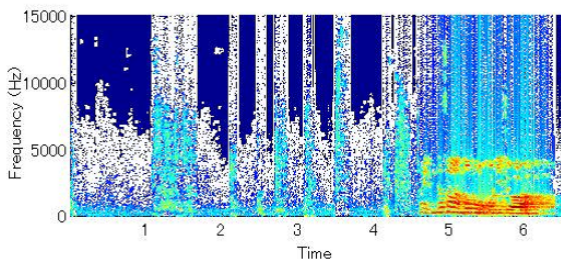


図 6. 処理 1 の実行結果

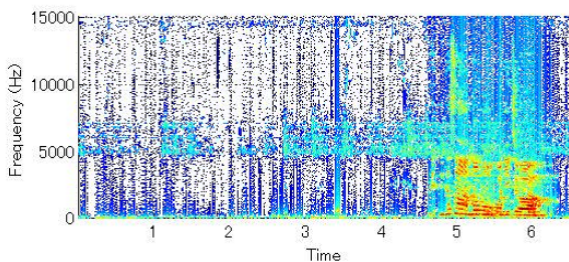


図 7. 処理 1 の実行結果

#### 4.3 ビデオカメラの違いによる比較実験の結果と考察

まず、二つのビデオカメラの違いを示すために、並べて録画した動画から同じ音声区間を抽出したものをサンプルとした。ビデオカメラ 1 とビデオカメラ 2 では音量差があることが確認できた。ビデオカメラの違いによる処理後音声比較の主観評

表 2. ミュージカルノイズ軽減手法の主観評価

サンプル	1	2	3	4	5	6	7	8	平均
処理1	MNの軽減	3.5	2.8	2	2.8	1.5	2	2.8	2.4
	音声	2.5	2.5	2.5	3	3.5	2	3	2.8
処理2	MNの軽減	4.8	3.5	3.5	2.5	3.5	3	3	3.4
	音声	3.5	4.5	4.3	3	4.5	4	3.8	4

価実験結果 (表 3) より、評価結果の誤差平均は 0.66 となり、ビデオカメラの違いによる処理後音声に、大きな違いは見られなかった。これは、ビデオカメラの違いによる音声の違いが、ビデオカメラによる小さい、定常的な雑音であると考えられ、そのためどちらもスペクトル減算法とミュージカルノイズ軽減法により軽減されたためと考えられる。よって、スペクトル減算法を用いた音質改善法とミュージカルノイズ軽減手法は、様々な家庭用デジタルビデオカメラで録画された動画の音声に対し有効であると考えられる。

表 3. ビデオカメラの違いによる処理後音声の主観評価

サンプル	カメラ1		カメラ2		
	7	8	7	8	
処理1	MNの軽減	2	2.8	1.5	2.3
	音声	3	3	3.3	1.5
処理2	MNの軽減	3	3	2	2.8
	音声	3.8	4	3	3.5

## 5 まとめ

今回の研究を通し、スペクトル減算法を用いた音質改善法は定常的で小さい音について有効ではあるが、ミュージカルノイズが発生してしまうこと、また、音声と雑音の音量差が大きく雑音の音量が小さいサンプルにおいては、フーリエ変換を行う際にブラックマン窓を用いるのが有用であることが確認できた。また、スペクトル減算法で発生したミュージカルノイズの軽減手法としては、周波数フレームのスペクトルを周波数軸方向に走査させる手法が有用であると可能性がある、と考えられた。現在のプログラムだと閾値設定が手動 ( $0 \leq rate \leq 1$ ) であり、操作を簡単にするためには閾値を自動で算出するアルゴリズムが必要となるので、今後の課題とする。また、本研究ではミュージカルノイズの軽減に用いる閾値を各スペクトルの最大値を基準として算出した。そのため、各フレームごとに異なる閾値となる。しかし、雑音のスペクトルは各フレームにおいても一定の可能性があり、閾値をすべてのフレームにおいて一定とすることでミュージカルノイズを軽減できる可能性がある。そのため、微弱な信号のフレームにおける一定値を指定し、全フレームの閾値とすることなどの、閾値設定の見直しについても今後の課題とする。また、今回の研究では主に動画から抽出した音声に処理を行ったが、元の動画と処理後音声の合成後の検証を進める必要がある。

## 参考文献

- [1] 古井貞照, 音響音声工学 電子情報工学入門シリーズ 2, 青島伸治 小畑秀文 南谷崇 共編
- [2] Boll.S, 1979, Suppression of Acoustic Noise in Speech Using Spectral Subtraction
- [3] 伊達玄 訳, コロナ社, デジタル信号処理 (上)
- [4] Hirokazu Suzuki Kenji Nakayama Akihiro Hirano, 2005, On Improvement of Voice Quality in Noise Cancelers Based on Spectrum Suppression Approach
- [5] 藤岡一馬 早坂昇 吉田則信 宮永喜一, 2003, スペクトルの時間変化を利用した雑音除去