

タイトル	画像処理によるモノラル円盤レコードからの音声再生
著者	魚住, 純
引用	北海学園大学工学部研究報告, 35: 119-129
発行日	2008-02-00

画像処理によるモノラル円盤レコードからの音声再生

魚 住 純*

Reproduction of Sound from a Monaural Disc by Means of Image Processing

Jun UOZUMI*

Abstract

A method for reproducing sound from old monaural discs is developed on the basis of digital image processing. In this method, the surface of a disc is illuminated with a large illumination angle and a sequence of magnified images is captured by a CCD camera, fed into a computer and processed digitally to yield a sound signal. In the image processing, each color image is first converted into a gray-level image, then transformed from the polar to Cartesian coordinate representations, processed for removing noises, and segmented into individual groove images. From the segmented groove images, sound signals are retrieved, which are connected mutually in the right sequence, processed for the trend removal and frequency filtering to yield finally a sound signal. At present, images containing a groove of 31 rotations were processed and the sound signal of about 24 seconds of good quality was reproduced successfully.

1 はじめに

エジソンがスズ箔蓄音機を発明して以来近年のLP盤に至るまで、録音・再生技術の発達に伴って、形状、材質、記録方式などの異なる様々なレコードが開発され、使用されてきた。そのなかでも、スズ箔蓄音機の改良版として登場し、レコードを聴く楽しみを一般人に開放した蠟管レコードや、その後に登場した円盤レコードの普及版であるSP盤は、庶民が音楽を購入して楽しむことに加えて、音楽家や言語学者などが貴重な音楽や音声を研究目的で記録し、繰

* 北海学園大学工学部電子情報工学科

* Department of Electronics and Information Engineering, Hokkai-Gakuen University

り返し再生することのできる道具としての重要な役割も果たした。そのような歴史的にも貴重な音声記録されたレコードには現存するものも多いが、レコード自体が貴重な文化財であるため、触針式の蓄音機での再生を禁じられているものや、割れや欠損などにより針による再生自体が不可能となっているものも多い。

そのような貴重なレコードから蓄音機を使わずに非接触で音声を再生する方法として、レーザー光を用いる一連の手法が開発されてきた¹⁾。その発端は、レーザー光による蠟管レコードからの音声再生法である²⁻⁴⁾。これは、レーザー光を蠟管表面に照射し、それが反射する方向から音溝の傾斜を読み取って音声情報を得るもので、レーザービーム反射法と呼ばれている。開発の目的は、ポーランドの人類学者であるB.ピウスツキが樺太アイヌの言葉を録音した蠟管の再生にあったが、その後、ブラームスのピアノ演奏や川上音二郎一座の琴の演奏などの貴重な音声の再生にも用いられた⁵⁾。この後者には、蠟管のネガティブ用に開発されたレーザービーム反射法が用いられている⁶⁾。

一方、円盤レコードは、蠟管とは音声の記録方式が異なることからレーザービーム反射法を適用することができない。このため、モノラルレコードであるSP盤を対象として、音溝によるレーザー光の回折現象を利用したレーザービーム回折法が開発された⁷⁾。しかし、円盤レコードでは音溝の横方向への変位として音声情報が記録されていることから、レーザービームが蛇行する音溝を追従する技術に高い精度が要求され、特に傷、錆、割れ、欠損などのあるレコードでは、トラッキングが困難になることが大きな問題となっている。

円盤レコードからの非接触音声再生におけるこのような問題を回避する方法として、本研究では、通常光で照射した音溝の画像にデジタル処理を施すことによって音声を再生する方法を開発することを目的としている。この手法は、まだ実用に供する段階には至っていないが、短時間ながら比較的良質な音声の再生が可能となったことから、この手法の現状と今後の課題について、用いている画像処理手法を中心に報告する。

2 再生装置と音溝の撮像

2.1 再生装置

現在本研究に使用している装置の写真を図1に示す。レコード表面の音溝の撮像に際しては、レコードの中心からの距離およびレコードの回転角度を正確に制御して撮像位置を定める必要がある。このため、X軸パルスステージ（中央精機製，LST-120X）とそのコントローラ（中央精機製，CPC-1D），および回転パルスステージ（シグマ光機製，KSA-160PM）とそのコントローラ（シグマ光機製，AS NET-3）を使用している。回転パルスステージはX軸ステージ上に固定し、その上にレコードを載せ、ハロゲンランプ光源（HOYA-SCHOTT製，MegaLight100）からの光で撮像領域を斜め方向から照射している。これにより、音溝の形状を

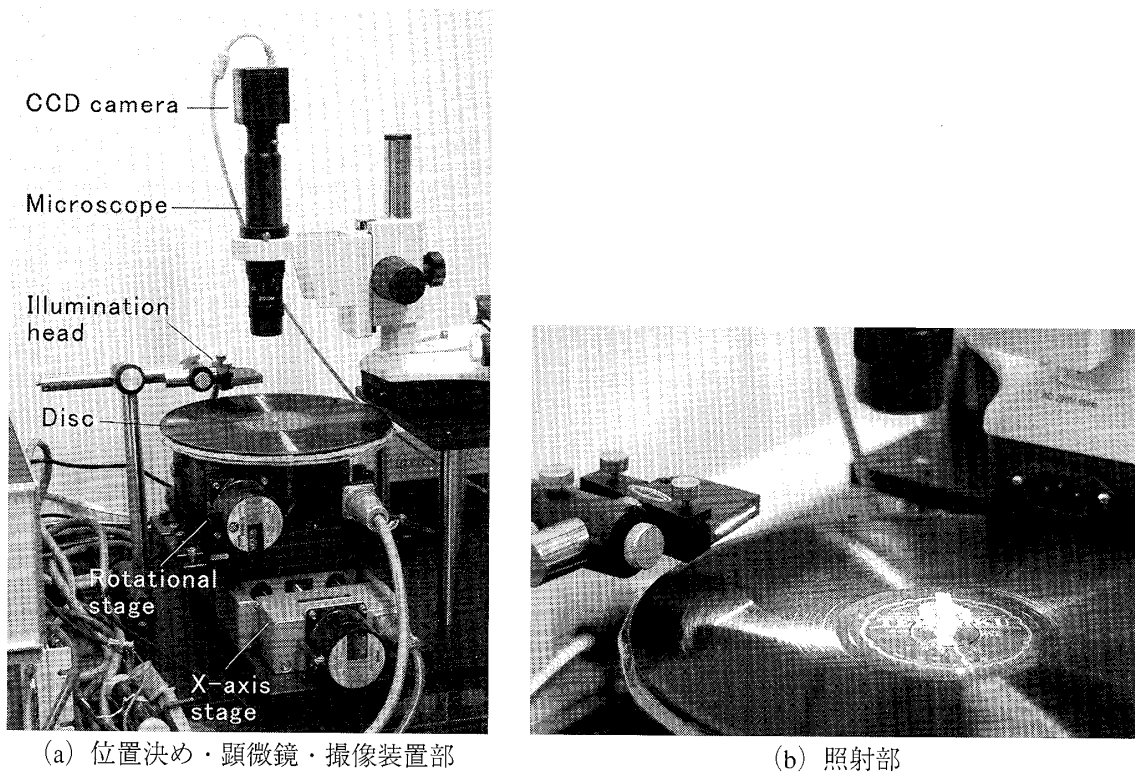


図1 再生装置. パルスステージのコントローラおよびPCはこの左右に設置

反映した陰影を形成し、それを顕微鏡（シグマ光機製，MXZ-2）で拡大して，CCDカメラ（富士写真フイルム製，HC-300Z，画素数1000×1280；ARTRAY製，ARTCAM-500P，画素数1920×2560）により撮像する．画像データは，USBを介してWindows XPベースのPCに転送する．CCDカメラは，平成17年度まではHC-300Zを用い，平成18年度からはより高解像度のARTCAM-500Pを用いている．しかし，現在のところ最も良質の再生音声は前者による画像から得られていることから，双方を併記した．

なお，再生技術の開発用として試験的に用いるレコードは，音楽よりも言葉による語りの方が聴取により再生音声の確認がしやすいことから，現在は落語のSP盤レコード（柳家金語楼，「細君三日天下」，日本帝国グラモフォン社製）を用いている．

2.2 音溝の撮像

音溝部分が明部，非音溝部分（主に表面の平坦な部分）が暗部となるように，約70°の比較的大きな照射角でレコード表面を照射し，その一部を光学顕微鏡を用いて拡大した後，約5.3 mm×4.0 mmの領域をCCDカメラにより撮像する．これを，回転パルスステージによりレコードを一定角度ずつ回転させて繰り返し，約80枚の撮像によりレコードを1周する．その後，X軸ステージによりレコードを径方向に一定距離移動させて，同様の撮像を繰り返す．現在のところ，最大で5周分の画像を取り込んで音声再生処理を行っている．撮像に際しては，隣接す

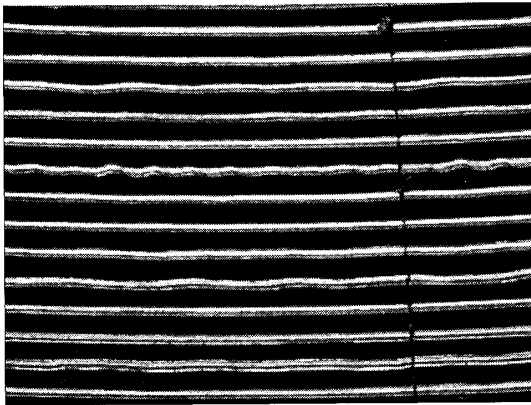


図2 グレーレベル変換後の音溝画像

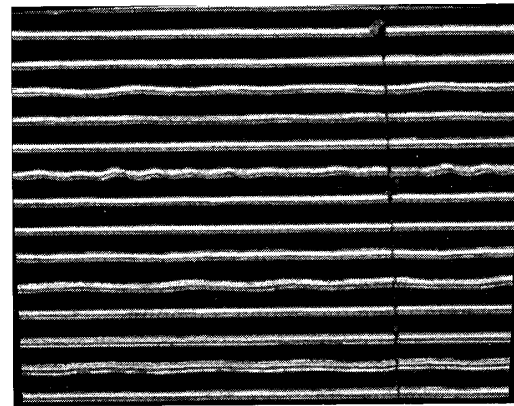


図3 湾曲修正処理をした画像

る画像間の音声情報を結合させる処理を行う都合から、レコードの回転方向の隣接画像、および径方向の隣接画像のいずれにおいても、画像が部分的に重複するように撮像領域の移動量を決めている。

3 画像処理

3.1 座標変換と2値化

3.1.1 グレーレベル画像への変換

使用したCCDカメラはカラー仕様であり、画像はRGBのビットマップ形式でPCに保存される。この画像に対してMATLABを用いた一連の画像処理を施すことにより、音声情報を抽出する。処理には、MATLABのImage Processing Toolbox (IPT) も用いている。

音声再生には色情報を必要としないことから、まずはじめに、これをグレーレベル画像に変換する。図2に変換後の音溝画像の例を示す。SP盤レコードでは、音声情報は音溝の面内の横方向の振動として記録されている。図2の音溝は、場所によって振動の振幅や周期が異なっており、この画像に音声情報が含まれていることが確認できる。なお、縦方向に走る暗線は、レコードの亀裂によるものである。亀裂を持つレコードにも対応可能な画像処理を検討する目的から、このようなレコードを用いている。

CCDカメラARTCAM-500Pは、撮像した画像に幅2ピクセルの黒い枠を付加して出力するため、これを除去した後は、画像のサイズが1920×2560から1916×2556へと変更される。RGB画像からグレーレベル画像への変換には、IPTの関数`rgb2gray`を用いてNTSCの輝度信号を取り出す方法等が考えられるが、レコードの画像が全体にほぼ一様な色彩を持つ画像であること、および計算時間の短縮を考慮して、単純にR, G, Bの3成分の和を求め、それを最大値で正規化することにより、(0,1)の範囲の疑似輝度情報に変換する方法を用いた。

3.1.2 湾曲修正処理

円盤レコードの音溝は、最外周から内側に向かって螺旋状に刻まれているが、溝の間隔が約 $300\mu\text{m}$ と狭いことから、実質的には真円を基準線とした横方向への変位として音声記録されているとみなすことができる。しかし、基準線が真円のままでは変位の計算が複雑になるため、これが直線となるよう、座標変換による湾曲修正処理を行う。

処理に用いた画像は、画素間隔が $\Delta x = \Delta y = 2.08\mu\text{m}$ であり、この間隔の正方格子上の直角座標点における輝度値が得られていることになる。これを、音溝の基準円に適合した座標系である極座標に変換する。すなわち、径方向の間隔を $\Delta r = \Delta x$ とし、角度間隔を $\Delta\theta = 3.6 \times 10^{-5}\text{rad}$ とする極座標格子を設定し、この格子点上の輝度値を x - y 座標点上の輝度値からのリサンプリングにより算出する。得られた r - θ 座標点上の輝度値を、 r を縦軸、 θ を横軸とする直角座標上の画像として表示することにより、直線を基準線とする音溝画像が得られる。円盤レコードが一定の回転角速度で音声を録音・再生していることから、 θ を直角座標系の横軸とすることにより、横軸が時間軸に比例するという利点も得られる。

この処理によって得られた画像を図3に示す。図の左右の下部に見られる黒い部分は、この処理が矩形の原画像を扇形に変換するために生じたものである。この部分は、その後の処理に悪影響を与えるため、次に述べる2値化処理後に切り捨てる処理を行っている。

3.1.3 2値化

座標変換後の画像は、2値化により音溝部分と非音溝部分の区別を明確化する。音溝画像を2値化する際の最適な閾値の条件は、その処理に伴って生成あるいは残存する微細なノイズを最小化でき、かつ音溝形状を可能な限り損なわないことである。しかし、その判定のためにはノイズの定量化等の処理も必要とし、計算が複雑である。そのため、現時点では、目視により最適な閾値を定めている。

図3を4種類のしきい値で2値化した画像を図4(a)–(d)に示す。この図で白い部分が音溝部分、黒い部分が非音溝部分に相当している。閾値を0.1とした場合が、最も音溝の欠損が少なく、かつノイズも抑制されていることから、この値を閾値として採用した。なお、2値化の方法としては、検討した処理法の中で最も計算時間の短かったIPTの関数`im2bw`を用いた。

3.2 ノイズ除去

レコードの表面の細かい傷やほこりは、撮像時の照明の仕方により2値後の画像に様々なノイズを発生させる。これらのノイズは、最終的な再生音声のノイズとなるうえ、音声情報の抽出作業自体の妨げになる場合もあることから、この時点で可能な限り除去する必要がある。本研究では、以下に述べる複数の方法を順次適用してノイズ除去を行っている。

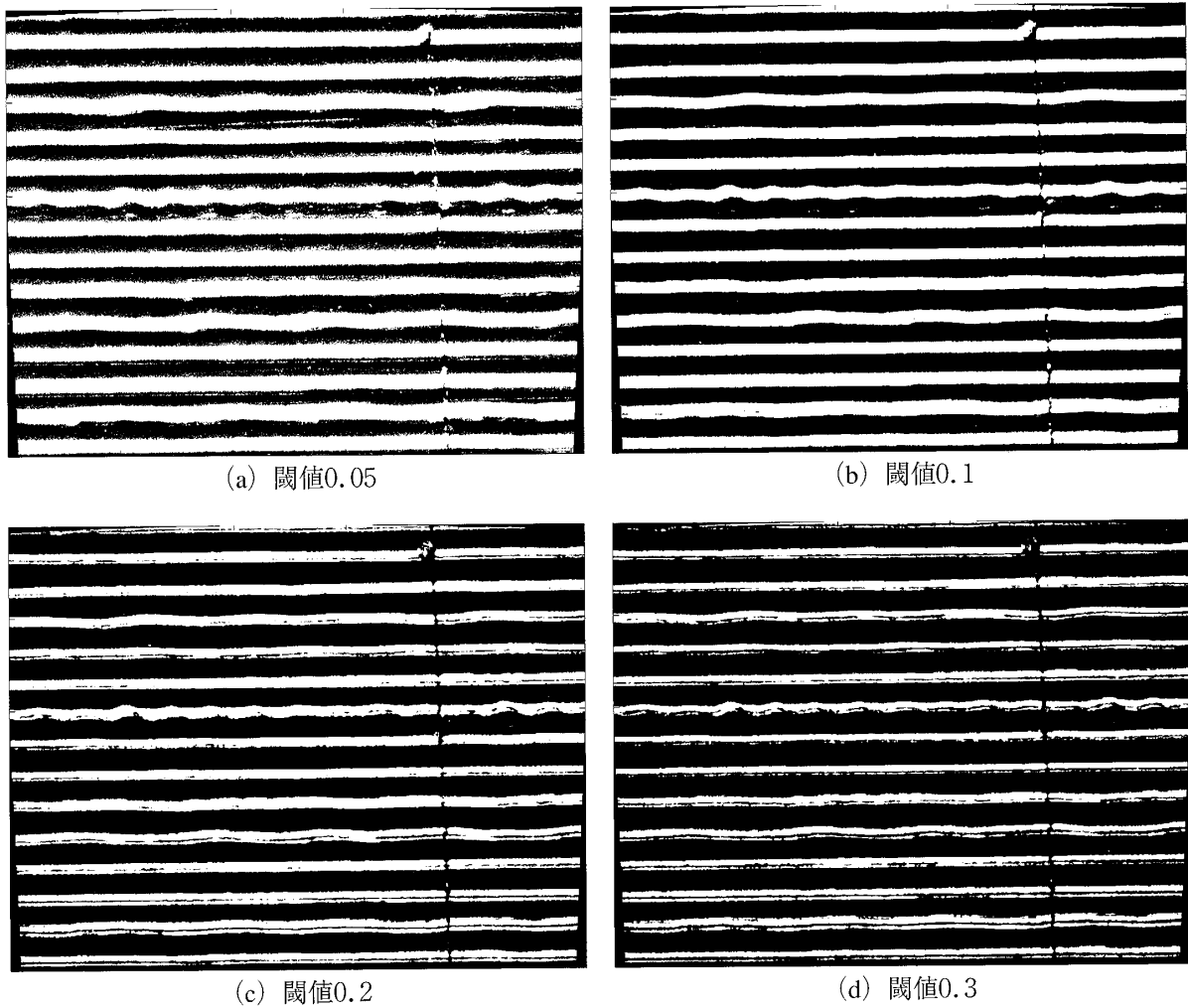


図4 4つの閾値で2値化した画像

3.2.1 小領域ノイズ除去

2値化した音溝画像に見られる最も顕著なノイズは、図5(a)に示すように、白い音溝領域内の黒い孤立領域、および黒い背景部内の白い孤立領域である。このような孤立した小領域を除去するため、IPTの関数**bwareaopen**を用いた。この関数は、2値データ内の指定された面積（3次元では体積）以下の小領域を除去するもので、これを直接適用した後、白黒を反転して再度適用することにより、白と黒の孤立小領域が除去される。

図5(a)の上から12本目の音溝に対して小領域ノイズ除去を行う前と除去後の画像を示したのが図6である。また、図5(a)全体の処理後の画像を同図(b)に示した。

3.2.2 突起部の除去と平滑化

小領域ノイズ除去によって大部分のノイズは除去されるが、図5(b)の4本目の音溝を拡大した図7(a)に見られるような音溝部分とつながっている突起状のノイズは、**bwareaopen**関

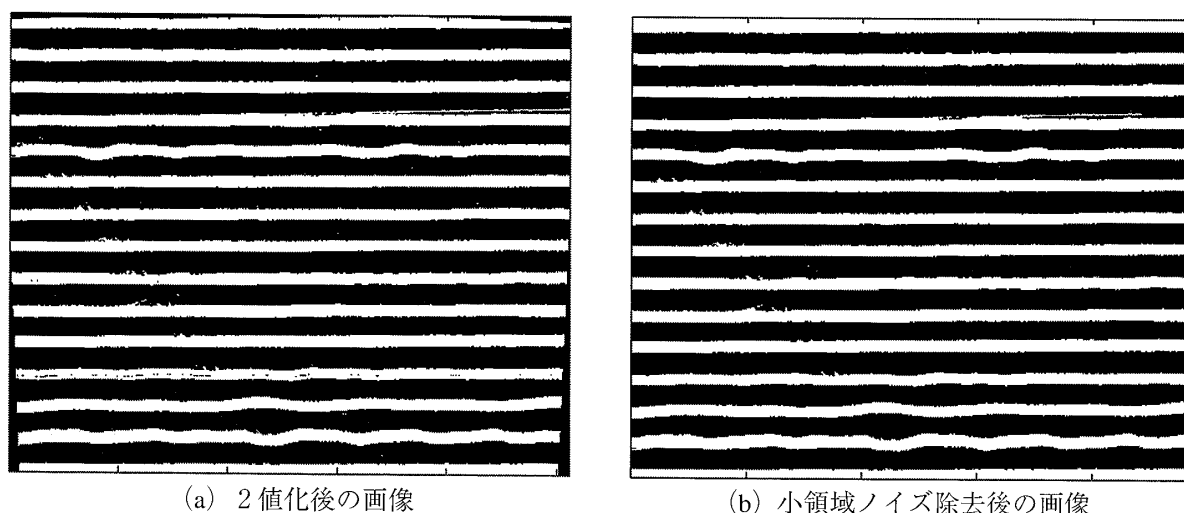


図5 2値化および小領域ノイズ除去後の画像

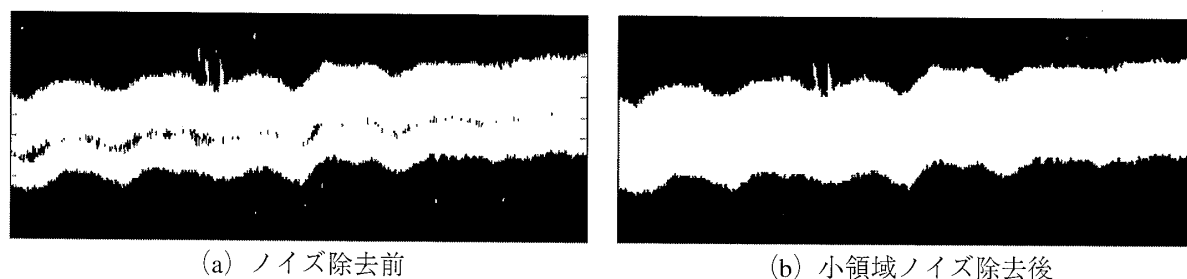


図6 小領域ノイズ除去処理の効果

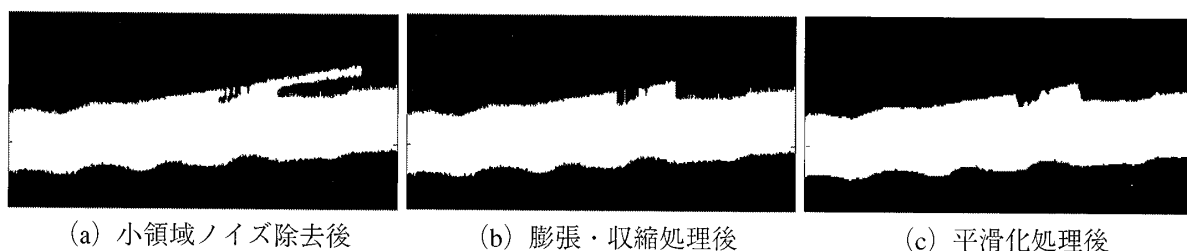


図7 突起部のある音溝画像への膨張・収縮処理および平滑化処理

数では除去されない。このような構造を除去するため、膨張処理と収縮処理を適用する。膨張処理は、白領域を黒領域に向かって1画素分拡張し、収縮処理は逆に、白領域に向かって黒領域を1画素分拡張する処理である。片方の処理を適当な回数適用した後、もう一方を同じ回数適用することにより、白黒の各領域への突起状の部分の除去することができる。ただし、本研究の場合、音声情報を持つ音溝の形状が必要以上に变形することを避けるため、膨張・収縮の方向を上下方向に制限している。この処理には、従来独自に作成したMATLABコードを用いていたが、現在は、IPTの関数imerodeとimdilateを用いている。

図7(a)にこの処理を適用した結果が同図(b)である。突起部は除去されているが、音溝部分と非音溝部分との境界は滑らかな曲線にはなっておらず、非常に細かく振動している。このよ

うな細かい振動も、高い周波数を含むノイズと考えられるため、2次元メディアンフィルタリングを用いて境界を滑らかにする処理を行う。これには、IPTの関数`medfilt2`を用いた。図7(c)は、同図(b)に大きさ 21×21 のメディアンフィルタを適用した結果である。画像中央付近の白く尖ったノイズがある程度低減され、音溝も比較的滑らかな状態になっているのが分かる。なお、メディアンフィルタ処理は、音溝の左右端が丸まる副作用を生じる。このため、処理後に両端を適当な長さ分切り取る処理を施している。

3.3 音声信号の抽出と結合

ノイズ除去処理後の画像から音声信号を抽出するには、まず複数の音溝を含む画像から個々の音溝部分を切り出す必要がある。これには、隣接する2つの音溝の間に水平の切り出し線を設定し、そこで画像を切断する。切り出し線の決定は、画像全体を水平方向に射影して列ベクトルを求め、そのベクトル上の隣接する音溝間の黒領域を一定の閾値以下の暗部として定めて、その中央位置を算出することにより決定した。

切り出された各音溝画像は、図7のように、黒を背景とした帯状の白い音溝部からなっている。音声情報は、この帯の上下の振動として記録されており、その振動を1値の関数として取り出すことを考える。これは、いわゆる細線化処理に相当する。細線化処理は、一般に2次元面全体にわたって幅のある構造を線構造に変換する処理であるが、本研究の目的からは、上下方向にのみ細線化をすればよい。このため、画像の各列毎に、その重心を求める方法を用いた。具体的には、ある列の輝度関数を $p(y)$ とすると、その重心 s を、

$$s = \frac{\int yp_n(y)dy}{\int p(y)dy}$$

として求めた。これは、 $p_n(y) = p(y) / \int p(y)dy$ が非負かつ全面積1の関数であることから、これを確率密度関数と見なし、その平均値を確率重心として求めるものである。この重心 s は列番号の関数であるから、 $s(x)$ として音声信号が得られたことになる。

ただし、ひびなどの影響で音溝を横切るように黒い部分がある場合、その列では重心を求めることができない。したがって、音溝の幅が一定以上狭くなっている、音溝の幅として問題があると判断された場合には、その列の幅を直前あるいは直後の列の幅で置き換える処理を施して、この問題を回避している。

各音溝から得られた信号 $s(x)$ は、隣接する撮像フレーム間の対応する音溝画像から得られた信号と接続させる処理を記録1回転分繰り返して行い、1周した後は、もとの撮像フレーム内の下に隣接する音溝から得られた信号に接続させる処理を行う。これらの接続処理は、音声信号の再構成において重要な部分であるが、本稿では紙面の都合上省略する。

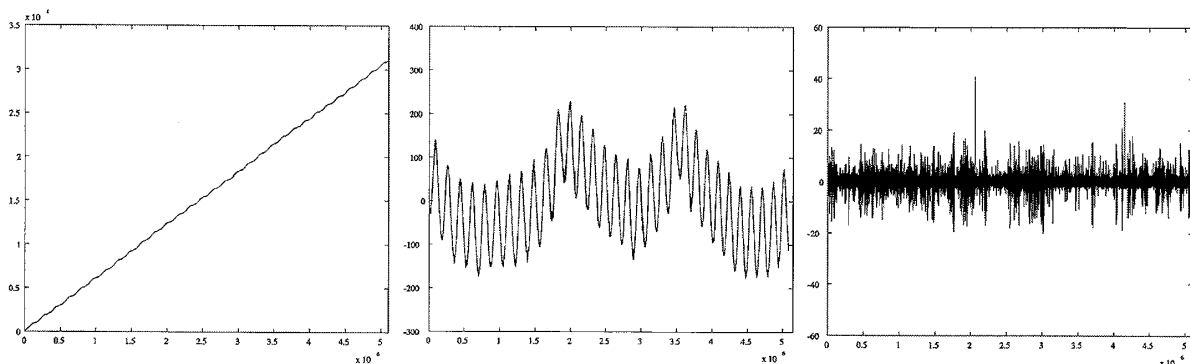
4 音声データ処理

図8(a)は、撮像フレーム3周分、レコード31回転分の音溝画像から得られた信号を接続した結果である。SP盤レコードの回転数は78rpmであるから、これは約24秒間の音声信号に相当する。

この図の音声信号には、右上がりの大きなトレンドが見られる。これは、主にCCDカメラの設置状態の不備によるものである。本来、カメラのCCD素子が捉える画像の垂直軸の延長線がレコードの中心を通るべきであるが、カメラが傾いていたり、横方向にずれて設置されていると、軸外れが生じ、湾曲修正処理後の音溝の基準線が水平線に対して傾きを持つようになる。これを修正するため、線形トレンドを関数`detrend`により除去する。その結果を図8(b)に示す。

図8(b)には、2種類の周期の振動が含まれている。このうち、短い周期の振動は、レコードの回転回数分の周期が含まれていることから、レコードの回転に伴う変動であり、レコードに開けられた穴が実際の音溝の回転中心から僅かにずれる偏心に起因するものと推測される。また、長い周期の成分は3周期分あり、これは撮像フレームの3回転分に相当している。これは、各撮像フレーム毎にレコード中心からの距離を定数に固定して音溝の基準円を算出したことによる誤差であると推測される。

これら2種類の振動は、いずれもレコードに含まれる音声信号に比べて周波数の低い領域にあり、周波数フィルタリングによりほぼ完全に除去することが可能である。また、周波数フィルタリングは、画像処理の過程で行ったノイズ除去処理後にも残存すると思われる高周波ノイズの除去にも有効である。周波数フィルタリングを行うには、FFTを行う関数`fft`により音声信号を周波数領域に移し、可聴音声周波数である20Hz~20kHz以外の周波数成分を削除した後、`ifft`により逆フーリエ変換する方法が基本的である。CCDカメラHC-300Zで撮像した画像から構成した音声信号には、この方法を用いている。一方、より高解像度のARTCAM-500P



(a) 音溝31周分を結合した信号

(b) トレンド除去後の信号

(c) 周波数フィルタ処理後の信号

図8 音声データの処理

で撮像した画像から構成した信号では、信号のデータサイズが大きいことから、使用しているPCシステムではFFTが実行できない。このため、これに相当する処理をウェーブレット変換によるフィルタリング処理として行っている。

周波数フィルタリングを終えた音声信号を図8(c)に示す。これは、ARCAM-500Pを用いて撮像し、ウェーブレット変換により処理を行った結果の信号であるが、HC-300ZとFFTを用いた場合もこの表示スケールでは同じような信号波形が得られている。この音声データは、最終的にWindowsの標準音声ファイル形式であるwavファイルに変換して保存する。これには、MATLABの関数wavwriteを用いる。

再生された音声は、HC-300Zによるものが現状では最も良質であり、柳家金語楼による軽快な落語の語りを明瞭に聞き取ることができる。また、ノイズも比較的低レベルである。一方、ARCAM-500Pによる音声は、まだ残存ノイズが多く、音声の不連続になっている部分もあり、十分な音質が得られていない。主な原因としては、各音溝から抽出された音声信号の結合精度が不十分であること、およびウェーブレットによる周波数フィルタリングが最適化されていないことなどが考えられる。

5 おわりに

高解像度のCCDカメラを用いて撮像したSP盤レコードの音溝画像に、一連の画像処理および信号処理を施すことにより音声信号を抽出することができることを示した。現時点では、31回転、時間にして約24秒の音声と比較的良好な音質で再生することに成功している。

本手法には、まだ改良すべき点も多い。再生音質に最も大きく影響を与える処理として、隣接する画像間での音声信号の接続処理がある。この処理の精度が不十分な場合、極めて短時間の音声信号が不適切に結合されるため、音質は極度に劣化する。この処理は現在最も改良を要するものであり、本論文ではその詳細を割愛している。今後、安定した処理法を確立した後、改めて報告することとしたい。ウェーブレットによる周波数フィルタリングも、使用するウェーブレットの検討も含めて改良の余地がある。また、照射光の照射方法は、画像に反映する音溝形状を決定することから、再生音質に大きな影響を与えるものであり、検討すべき重要課題の一つである。現在は、比較的高コントラストが高く、音溝形状を反映した陰影を生ずる照射角を経験的に定めている状態であり、より系統的な照射方法の検討が必要である。再生時間をさらに伸ばし、レコード全周の再生を実現するには、パルスステージの駆動、撮像、およびデータ処理を連携して実行するシステムの構築が必要である。現在、これらすべてをMATLABにより行うための準備を進めている。

本研究は、過去および現在の学部4年生の卒業研究プログラムと連携して進めてきた。このテーマを担当した高橋瑞穂、佐々木敏文、小林慎平、石田 豪、小林大輔、関崎城矢、金濱伸

秀, 佐藤和弥の各君の熱心な協力に謝意を表する.

参考文献

- 1) J. Uozumi and T. Asakura : Methods for reproducing sounds from old phnograph records, in “International Trends in Optics and Photonics,” ed. T. Asakura, Springer, Berlin, 1999, pp. 409–425.
- 2) 伊福部達, 朝倉利光, 川嶋稔夫 : ピウスツキ録音蠟管レコードとその再生方法, 日本音響学会誌, Vol. 40, No. 3, pp. 167–174, 1984.
- 3) 朝倉利光, 岩井俊昭, 川嶋稔夫 : レーザー光を用いたピウスツキ録音蠟管からの光学的音声再生, O plus E, No. 66, pp. 64–72, 1985.
- 4) T. Iwai, T. Asakura, T. Ifukube and T. Kawashima : Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method, Appl. Opt., Vol. 25, No. 5, pp. 597–604, 1986.
- 5) 魚住 純, 牛坂 健, 伊福部達 : 古ろう管からの名演の復活, 日本音響学会誌, Vol. 54, No. 6, pp. 443–449, 1998.
- 6) 牛坂 健, 魚住 純, 朝倉利光 : ネガティブ蓄音管の光学的再生装置, 北海学園大学工学部研究報告, No. 26, pp. 383–400, 1999.
- 7) J. Uozumi and T. Asakura : Reproduction of sound from old disks by the laser diffraction method, Appl. Opt., Vol. 27, No. 13, pp. 2671–1986, 1988.