

琉球大学学術リポジトリ

倍音スペクトル削除法を用いた三線演奏の自動採譜

メタデータ	言語: 出版者: 琉球大学工学部 公開日: 2007-08-23 キーワード (Ja): キーワード (En): Pitch distinction, Overtone spectrum, Sanshin, Transcription 作成者: 和宇慶, 琢磨, 山城, 毅, 渡久地, 實 メールアドレス: 所属:
URL	http://hdl.handle.net/20.500.12000/1477

倍音スペクトル削除法を用いた三線演奏の自動採譜

和宇慶 琢磨* 山城 毅** 渡久地 實**

The Automatic Music Transcription of Sanshin Performance Using Overtone Spectrum Elimination Method

Takuma WAUKE* Tsuyoshi YAMASHIRO** Minoru TOGUCHI**

Abstract

In Okinawa, there is an original stringed instrument which is called "Sanshin". These musical instruments are played with unique score book called "Kunkunsi". In this paper, we proposed how to transcript the sound of playing Sanshin to the musical note. As preprocessing we attempt pitch distinction of Sanshin sound. To distinct the pitch we tried to delete the overtone spectrum. In addition, we derive the attacking time of Sanshin sound by sharpness of go up of waveform. The distinction with this method was done from an actual consecutive Sanshin sound, and good results were obtained on the single sound.

Key Words: Pitch distinction, Overtone spectrum, Sanshin, Transcription

1. まえがき

楽器や肉声で演奏された曲を楽譜に書き直す作業を採譜といい、音楽的な専門知識や技術を必要とされる作業である。この作業をコンピュータ上で自動的に行なう自動採譜は、音声認識の中でも重要な研究テーマといえ、これまでも、さまざまな研究がなされている [4],[5],[6],[7],[10]。

沖縄には三線(サンシン)という、本土の三味線に似た独特の弦楽器があり、この三線で演奏する楽曲が数多く存在している。これらの曲は工工四(クンクンシー)と呼ばれる三線独自の楽譜を用いて演奏され [8],[9]、現在も新しい曲が演奏されているが、西洋音楽などに用いられる、五線譜ほどの普及率は無いためと思われる。もし、三線の演奏から五線譜、工工四の両方に自動採譜が行えるならば、より多くの人々が三線を演奏することができるようになり、またアドリブなどの含まれている実際の演奏から、より正確な楽譜を作ることも可能になると考える。

本研究では、三線の実際の演奏から自動採譜を行うことを目的としてきた [1],[2],[3]。自動採譜を行うためには音高判別と発音時点測定を行う必要がある。

音高(ピッチ)判別の方法には、時間軸上の波形に着目した自己相関法、ピークトゥピーク法、ゼロ交差検出法などがある。しかし、この方法では複数の音が重なった、

多重音については対応が出来ない [10]。本研究では、三線演奏が対象となるが、三線曲の多くは演奏者の唄とともに演奏されたり、ときには他の楽器とともに演奏されるため [8]、この方法は適していないと考えられる。

そこで、本研究では音の周波数スペクトル列を用いて音高を判別する方法を用いることにした。周波数成分を用いた音高判別法もさまざまな方法が提案されているが [5],[6],[10]、本研究では周波数スペクトル分布から倍音スペクトルを削除する方法により、音高を判別するものとした。

発音時点の測定については、シフトフレームを用いて一定時間ごとに音高判別を行い、音高が変化した点を発音時点とみなす方法が主流である [5],[6],[7]。しかし、音の持続時間が短く、音の立ち上がりが鋭いという三線音の特徴を考慮すると、処理時間の面でこの方法には無駄が多い。

このため、本研究では時間波形上での閾値を用いた簡素な方法を用いて、発音時点測定を行っている。

本稿では自動採譜の前処理としての音高判別法、発音時点測定法の検討、および実際の採譜への応用結果について述べる。

2. 三線音の特徴

まず、解析対象の三線の調弦(チューニング)について述べる。三線にはいくつかの調弦法があるが、もっとも標準的に使われるのが「本調子」と呼ばれる調弦法である。これは第一弦開放を C_2 (ド)、第二弦開放を F_3 (ファ)、第三弦開放を C_3 (ド) に、それぞれ調弦する方法であるが [5],[6]、調弦は演奏者の声の高さなどによっても、微妙に異なり、必ずしもピアノなどで演奏した上記の音と同じ音高に調弦されるわけではない。

受理: 2001年12月9日

平成13年度電気関係学会九州支部連合大会にて発表。

* 大学院理工学研究科 電気電子工学専攻

(Graduate Student, Electrical and Electronic Eng.)

** 電気電子工学科

(Dept. of Electrical and Electronic Engineering, Fac. of Eng.)

本研究で用いる三線音は、国際標準周波数による12平均率音階に当てはまる基本周波数を持つ音高に調弦するものとする。

なお、国際標準周波数による12平均率音階 [2] では A_3 の音の基本周波数が 440[Hz] となる音階で、この音階での C_2 の音の基本周波数は 130.8[Hz]、 F_3 は 174.6[Hz]、 C_3 は 261.6[Hz] となる。

他の調弦法には一揚調、二揚調、三下調、一二揚調があるが [5],[6]、本稿では、説明は省略する。

三線で演奏できるもっとも低い音が第1弦の開放音であり、工工四では「合」という符号で表記される。前述の本調子で調弦した場合、 C_2 の音高となり、基本周波数は 130.8[Hz] となる。

次にもっとも高い音についてであるが、物理的にはある程度の高音域まで出せるのだが、一般的な演奏における最高音は第3弦の下の方を押弦して弾く、「イ五」という音となり、前述の本調子で調弦した場合、 D_4 の音高と対応しており、基本周波数は 587.33[Hz] である。なお、工工四に用いられる「工」などの符号は三線の指使い(勘所と呼ぶ)に直結しており、どの調でも変化しない音名(C, D, Eなど)ではなく、調性によって変化する階名(ド, レ, ミなど)に近いものである [6]。

「工」などの音が、どの音高に対応するかは前述の調弦で定まるが、上記してある表記は、全て本調子の場合であり、特に断らない限りは今後もそうであるものとする。

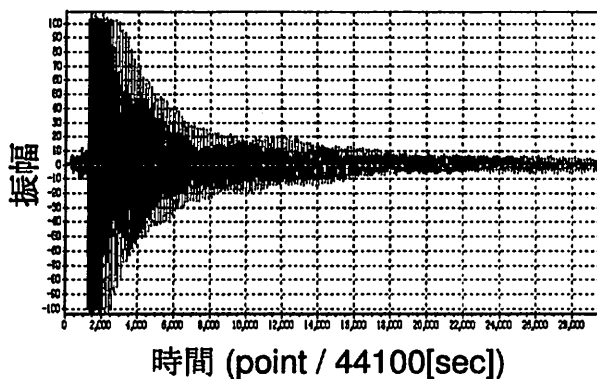


Fig. 1. 三線音 (261.6[Hz]) の時間波形

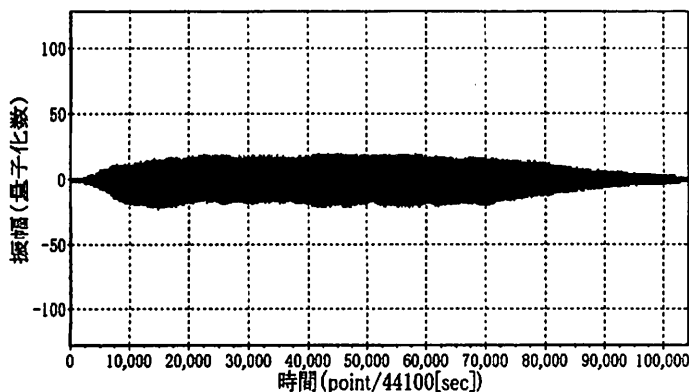


Fig. 2. ヴァイオリン音 (261.6[Hz]) の時間波形

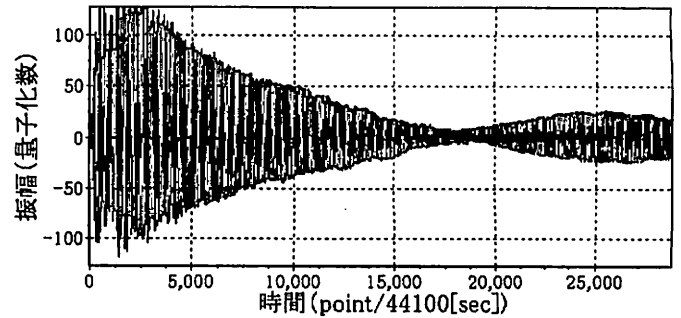


Fig. 3. ピアノ音 (261.6[Hz]) の時間波形

次に、解析の対象とした三線音の時間波形を Fig. 1 に示し、比較のため同じ弦楽器のヴァイオリン音とピアノ音それぞれの時間波形を Fig. 2, Fig. 3 に示す。なお、これらは基本周波数 261.6[Hz] (C_3 , ドの音) の波形である。

これらの時間変化を比較してみると、アタック時から音圧がピークに達するまでの時間は、ピアノ音も三線音も短く、そして減衰していくという波形であるが、持続時間は三線音の方が短い。また、ヴァイオリン音は、発音形態を考えると当然の結果として持続時間が長くなる。

次に、三線音の周波数分布の特徴について述べる。三線音、ヴァイオリン音、ピアノ音の周波数スペクトル分布を Fig. 4~Fig. 6 に示す。これらのスペクトル分布をみると、基本周波数である 261.6[Hz] の整数倍に周波数スペクトルが現われている。

三線音では通常の場合と異なり、基本周波数 (261.6[Hz]) のスペクトルが小さいという特徴を持ち、また倍音成分 (基本周波数の整数倍) は、他楽器に比べ、高次倍音まで豊かに出ている。

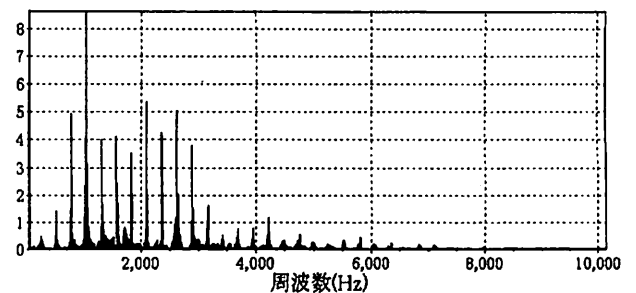


Fig. 4. 三線音 (261.6[Hz]) の周波数スペクトル分布

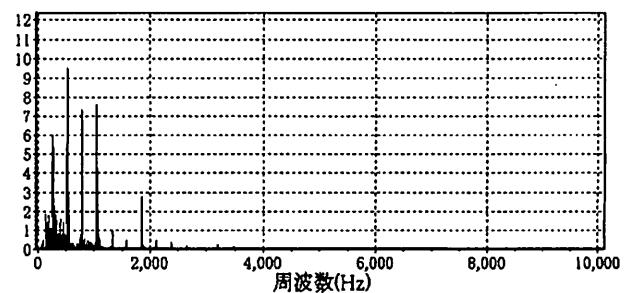


Fig. 5. ヴァイオリン音 (261.6[Hz]) の周波数スペクトル分布

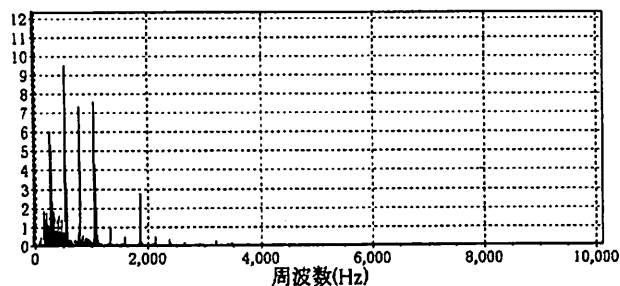


Fig. 6. ピアノ音 (261.6[Hz]) の周波数スペクトル分布

3. 倍音成分による音高判別

楽器に限らず、全ての音の音高(ピッチ)は基本周波数によって決まる。つまり、周波数分布から音高を判別するためには、基本周波数のスペクトルを調べれば良い。しかし、先に述べたように三線音は基本周波数成分が他のスペクトルに比べて非常に小さく、基本周波数にあたる周波数域のスペクトルから基本周波数成分を特定するのが困難な場合が多い。そのため、基本周波数以外のスペクトル、すなわち倍音成分を用いて音高の判別を行う必要があり、次に2つの音高判別の方法を提案する。

4. スペクトル間隔読み法による音高判別

倍音は、基本周波数の整数倍の周波数を持っているから、周波数スペクトル分布の各スペクトルの間隔を知ることによって基本周波数、つまり求めたい音高を判別する事ができると考えられる。

そこでまず、スペクトル分布のピーク値を検出しスペクトルの抽出を行う。普通、フーリエ変換後の周波数スペクトル分布には、離散化や符号化、録音時の環境による雑音などにより、Fig. 7のようにスペクトルの周囲に微小なノイズが生じる。このノイズの影響を軽減するため、周波数スペクトル分布内で最大の値を求め、この最大値の1/10の値でノイズカットを行なう。残った各スペクトルの間隔 f を求め、基本周波数として音高を判別するという方法がスペクトル間隔読み法である。

スペクトルの間隔をそれらの平均から求めると、あるスペクトルが取れない場合、正しい基本周波数の間隔に比べて大きくずれてしまい、正しい値が得られない。このため、間隔 f の頻度を数えて、最も多い間隔 f_{max} を基本周波数とすることにした。

Fig. 7 にスペクトル間隔読み法のイメージを、Fig. 8 にアルゴリズムを示す。

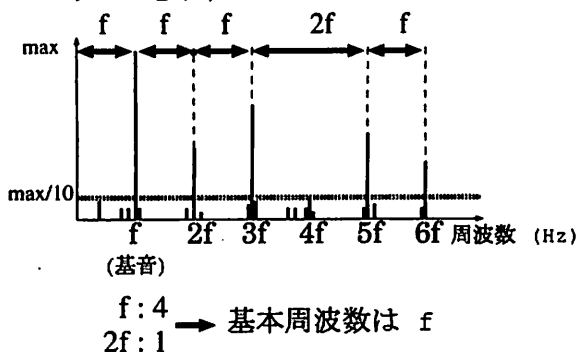


Fig. 7. スペクトル間隔読み法 (イメージ)

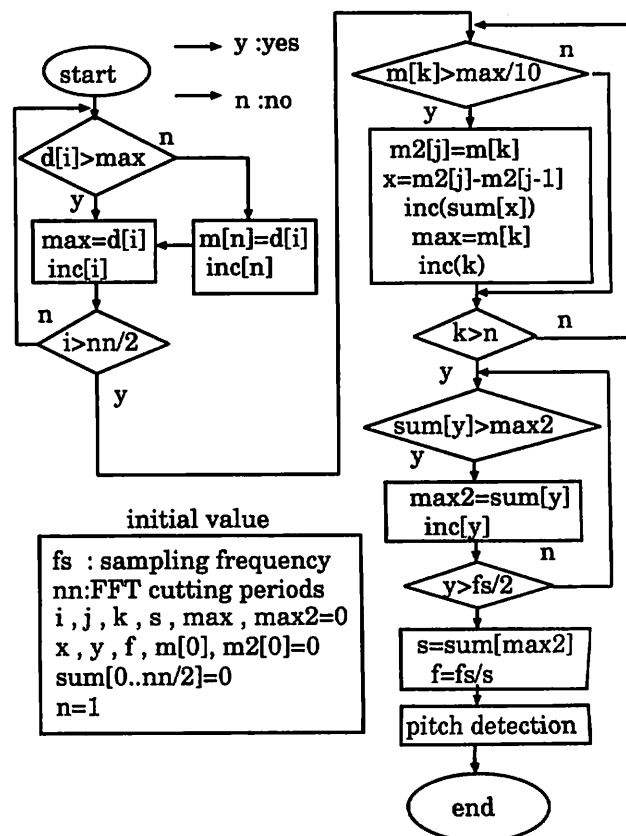


Fig. 8. スペクトル間隔読み法 (アルゴリズム)

この方法を用いて音高判別を行ったところ、単音について倍音成分が豊かに出ている場合には、ほぼ正確に音高を判別することが出来た。

しかし、音域や奏法によって倍音あまり出していない場合には、良好な結果が得られない。

5. スペクトル削除法による音高判別

前節で述べた音高判別法は、単音の場合には、ほぼ良好な結果が得られた。

しかし、前節の方法は、多重音の場合、スペクトルの間隔が基本周波数にはならないことが原因で、音高を判別することができない。

本研究では多重音についても音高が判別できることを目的としている。そこで、多重音に対応できる可能性のある方法として、スペクトル削除法を考案した。

ある音高の楽器音のスペクトル分布から、その音程に対応する周波数とその整数倍の成分(倍音成分)のスペクトルを取り除くと、スペクトルはほとんど残らない。例えば、図4の音のスペクトル分布から261.2[Hz]、および440[Hz]間隔でスペクトル削除を行ったスペクトル分布図をFig. 9, Fig. 10にそれぞれ示す。2つの図を比べてみると、音高の基本周波数に近い周波数間隔で削除したスペクトル分布(Fig. 9)に比べ、基本周波数に一致していない周波数間隔によって削除したスペクトル分布(Fig. 10)の方がスペクトルがより多く残っていることが分かる。

スペクトル削除法は、音高を判別する際、まず各音高の基本周波数の整数倍でスペクトルを削除し、残りのスペクトルを合計する。その合計が最小値となる周波数に対応した音高を、その音の音高と判別する方法である。

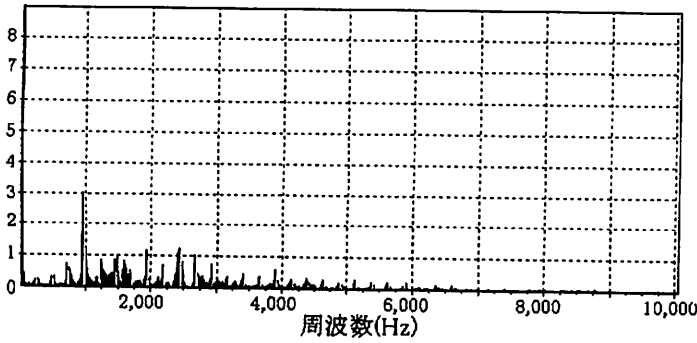


Fig. 9. 261.6[Hz] 間隔でスペクトルを削除したスペクトル分布

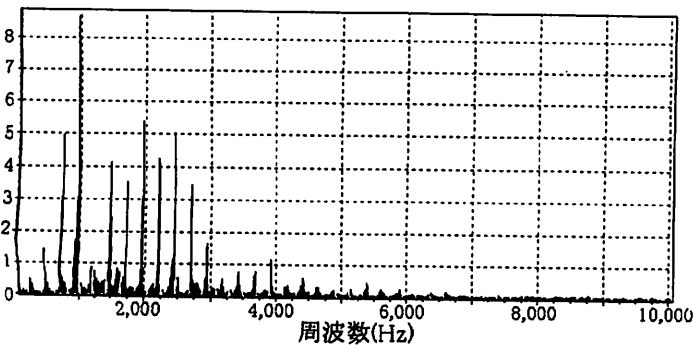


Fig. 10. 440[Hz] 間隔でスペクトルを削除したスペクトル分布

スペクトル削除法のイメージとアルゴリズムを、Fig. 11 と Fig. 12 にそれぞれ示す。

6. オクターブ判別

前節で説明した倍音スペクトル削除法によって音高判別をおこなったところ、オクターブ違いの音高と誤判別することが多かった。

そこで、最初に調べる音高の種類を判別範囲内の一番低い1オクターブに限定し、その後でオクターブ上の音と比較して最終的な音高を判別を行う二段階の方法を用いた。以下に詳しい説明を行う。

ある音の基本周波数を f とすると、1オクターブ上の基本周波数は2倍の $2f$ となり、この音は、 f の奇数倍の成分 $(2i-1)f [i=1...x]$ を持たない。よって、いったん判別された音高の基本周波数を F とし、式1によって奇数倍成分の合計を求め、しきい値 Th によりオクターブ判別を行う。

ただし、 x は倍音の数も考慮し、3から5の間の整数とし、 Th は一定の閾値とする。

$$Sum = \sum_{i=1}^x (2i-1)F \tag{1}$$

$$Th > Sum \begin{cases} YES & F = 2F \\ NO & F = F \end{cases} \tag{2}$$

オクターブ判別を加えて改良した、スペクトル削除法のアルゴリズムを Fig. 13 に示す。

7. 発音時点の測定

自動採譜は、連続音のそれぞれの音について音高を判別

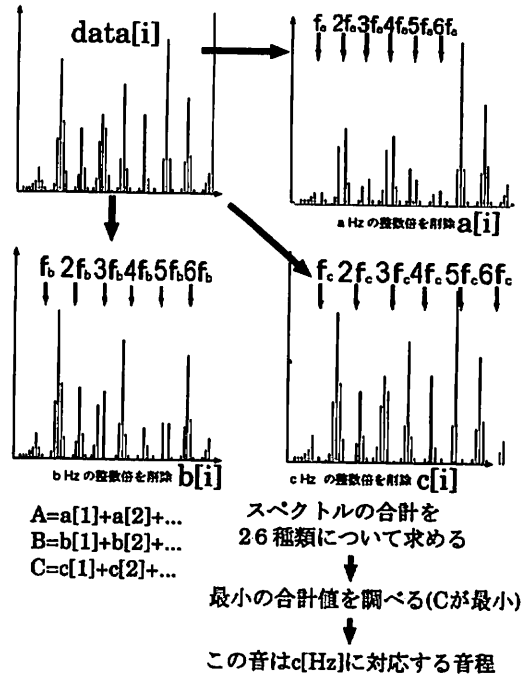
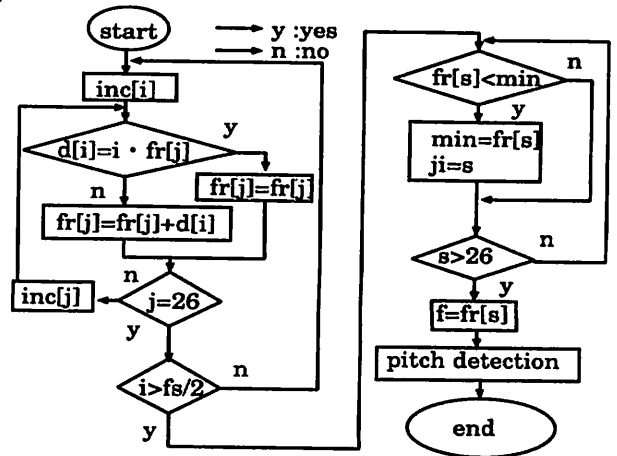


Fig. 11. スペクトル削除法 (イメージ)



initial value

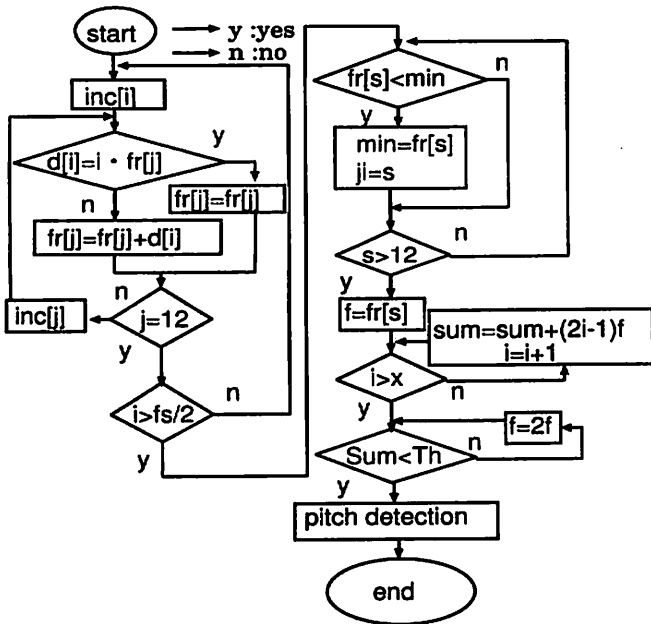
min=100000	fs : sampling frequency
i, s, ji =0	fr[1]=frequency of C2
f, fr[1..26]=0	⋮
j = 1	⋮
nn : FFT cutting periods	fr[26]=frequency of D4

Fig. 12. スペクトル削除法 (アルゴリズム)

することで処理されるので、各音の発音時点(アタック時点)を正しく測定する必要がある。

三線音は、Fig. 1 に示したようにアタック時の音圧がピークに達するのが非常に速い、つまり音の立ち上がりが鋭い。このことを利用し、音圧がある閾値を越えた点を発音時点とした。また、三線音は減衰も非常に速く、0.1[sec] 以内では、ほぼ確実に閾値(音圧ピークの12%)以下になる。

よって、次の発音時点は、最初に発見された発音時点から0.1[sec]後の時点以降を探すこととしたが、既に次の音が始まっていることも考えられる。このことも考慮し、探索開始時点に近い時点で、発音時点が見つかった場合は、時間軸とは逆向きの近傍にもっと振幅の大きい時点を探す



initial value

min=100000	fs : sampling frequency
i, s, j, sum =0	fr[1]=frequency of C2
f, fr[1..12]=0	⋮
j = 1	fr[12]=frequency of B2
nn :FFT cutting periods	th :Threshold
x :integer (3 or 4 or 5)	

Fig. 13. スペクトル削除法 (改良後アルゴリズム)

ことによって修正をするようにした。
発音時点測定の例を Fig. 14 に示す。閾値を超えた丸で囲んだ点が発音時点として測定されている。

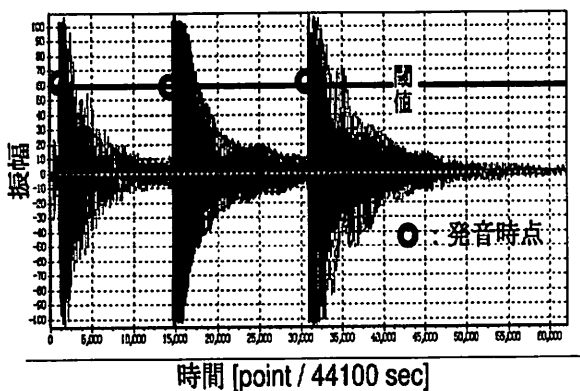


Fig. 14. 連続音における発音時点の測定

8. 処理結果

沖縄民謡である「安里屋ユンタ」の演奏波形の一部を Fig. 15, Fig. 16 に示す。またこれに対して、実際に採譜処理を行った判別結果の一部を TABLE1, TABLE2 に示す。

これは演奏曲中の連続した音に対して、それぞれ勘所 (三線の音階) と西洋音階、発音された時間、平均率に対応した音高を出力させている。

Fig. 15 については全ての音の発音測定、音高認識ができていて (TABLE1) が、Fig. 16 は、矢印の部分の音について発音時点の測定に失敗している。(TABLE2 の7番目と8番目の間に実際は音がある) 原因としては、この音の振幅が非常に小さいことが挙げられる。

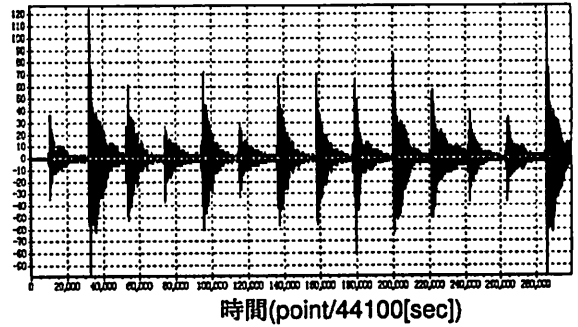


Fig. 15. 安里屋ユンタ (冒頭の一部)

勘所(音階)	発音時間[sec]	音高
中 (ラ)	0.241	A2
工 (ド)	0.738	C3
七 (ファ)	1.22	F3
合 (ド)	1.68	C2
七 (ファ)	2.16	F3
合 (ド)	3.11	C2
七 (ファ)	3.59	F3
七 (ファ)	4.07	F3
五 (レ)	4.55	D2
工 (ド)	5.03	C2
四 (ファ)	5.50	F2
上 (ソ)	5.98	G2
中 (ラ)	6.46	A2

TABLE 1 判別結果 (安里屋ユンタの冒頭部)

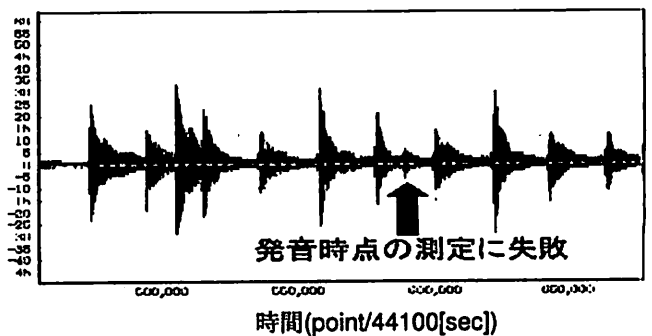


Fig. 16. 安里屋ユンタ (中間の一部)

勘所(音階)	発音時間[sec]	音高
四 (ラ)	10.74	F2
合 (ド)	11.21	C2
四 (ファ)	11.46	F2
上 (ド)	11.69	G2
合 (ファ)	12.16	C2
上 (ド)	12.66	G2
中 (ファ)	13.14	A2
合 (ファ)	13.62	C2
上 (レ)	14.11	A2
合 (ド)	14.59	C2
乙 (ファ)	15.08	D2

TABLE 2 判別結果 (安里屋ユンタの中間部)

この曲の演奏全体での発音された音は241音、そのうち、正しく判別できたものが234音、音を抽出できず判別できなかったものが6音、誤判別が1音であった。判別成功率は約97.1%であり、高い精度で判別できていることがわかる (TABLE3)。

また、TABLE1, TABLE2を見ても分かる通り、本方法では、細かい発音時点の測定が可能であり、一定のテンポではない微小なずれを検出することも可能であると考えられる。

正判別	誤判別	未判別	正解率
234	1	6	97.1%

TABLE 3 判別結果 (安里屋ユンタ全体)

しかし、安里屋ユンタはかなり好結果の例であり、その他の曲についても同様の処理を行ったところ、成功率にかなりのばらつきが見られた。平均するとおよそ、70%程度である。

9. むすび

今回、三線曲の自動採譜に必要な三線音の音程判別と発音時点を調べる手法を提案し、実際に演奏された曲に対して、自動採譜処理を行い、音高、発音時点ともある程度高い精度で判別できることを示した。

しかし、まだサンプルが少なく、結果にばらつきも多いため、今後は多くのサンプルに対して処理を行い、その有効性をさらに検証していきたいと考えている。

また、振幅の小さい音の発音時点が測定できないという問題についても、発音時点測定法を改良する必要がある。

最終的には実際の楽譜を作成するほか、当研究室の琉球サンシン自動演奏ロボットへの入力データとしての応用も考えている。

参考文献

- [1] 和宇慶 琢磨, 山城 毅, 渡久地 實: "スペクトル削除法を用いた三線演奏の自動採譜", 平成 13 年度電気関係学会九州支部連合大会論文集, pp.489, (2001).
- [2] 和宇慶 琢磨, 山城 毅, 渡久地 實: "倍音スペクトル削除法による三線音の音高判別", 2001 年電子情報通信学会総合大会講演論文集, D-14-6, (2001).
- [3] 和宇慶 琢磨, 山城 毅, 渡久地 實: "倍音スペクトル削除法による三線音の音程判別", 平成 12 年度電気関係学会九州支部連合大会論文集, pp.750, (2000).
- [4] 井口 征二: "音楽情報の処理—電子計算機を用いた自動採譜", 計測制御, vol.19, no.3, pp.314-319, 1980.
- [5] 梅本 敏孝, 青島 伸治: "採譜システムの構築と音高抽出精度の比較", 計測自動制御学会論文集, vol.29, no.10, pp.1227-1231, (1993).
- [6] 三輪 多恵子, 田所 喜昭, 斎藤 努: "くし形フィルタを利用した採譜のための異楽器音中のピッチ推定", 電子情報通信学会論文集 (D-II), J81-D-II(No.9), pp.1965-1982, Sep(1998).
- [7] 三輪 明宏, 守田 了: "ステレオ音楽音響信号を用いた三重層に対する自動採譜", 電子情報通信学会論文集 (D-II), J84-D-II(No.7), pp.1951-1960, Jul(2001).
- [8] 山内 秀吉: "琉球音楽の研究 -さんしん-", 山内秀吉琉球研究会, (1996).
- [9] 祖慶 剛: "琉球古典音楽 -野村流稽古本-", サン印刷, pp1-pp10, (1962).
- [10] 長尾 真, 宇津呂 武仁, 島津 明, 匂坂 芳典, 井口 征二, 片寄 晴弘: "文字と音の情報処理", 岩波書店, pp.163-189, (2000).