6T-06

歌声音源制作支援システムのための自動ミス検出*

加藤 大悟†, 伊藤 克亘,

1 はじめに

音楽を制作する際に、レコーディングという手法がある。その中で、歌を録音する場合がある。歌声の録音に必要なこととして、発声練習・録音・録音後の加工等が挙げられる。このような、歌声を制作する全ての過程のことを VocalProduction という [1]。今回はその中でも、「録音」「コンピング」という過程に着目した。VocalProduction における録音の過程は一回で済むものではない。つまり、何度も似たような音源を聴くことになり、耳が「慣れて」しまい、何が上手いのかの判断がつきにくくなってしまう。

そこで、「歌声制作支援システムの為の自動ミス検出」を提案する。この機能は、入力を音声とし、その音源に対してユーザがミスをしている箇所を自動で検出する。

2 音楽制作におけるミスと自動検出

そもそも音楽制作におけるミスとはどういうものかを述べる。過去の研究として、エレキギターの演奏練習支援システムの為のミス検出 [2] を行ったものがある。しかし、人の声である歌と楽器であるエレキギターでは考慮すべきものが異なる。つまり、ミスも異なるので、別の手法をとる必要がある。

そこで、従来のシステムとして「カラオケ採点システム」がある [3]。従来システムでは主に「音高」について着目して分析している。本研究はこのシステムにはない「音量」についても分析することで機能を実装する。

2.1 音程のミス

ここでいう音程とは楽譜情報における音符の音高である。与えられた音高に対して、一定以上離れた音高を演奏した時を「音程のミス」とする。主に「裏返る」・「一定の音を伸ばしている時に、段々高くなってしまう」・「短く、かつ比較的低い音が正確に出せない」等があげられる。

検出方法として、事前にユーザによって入力された楽譜情報を元に、正解の音高を求める。音声入力された歌声を一定点数のフレームに分割し、フレームシフト処理を行う (1 フレームを 960 点、フレームシフト、48 点:約 1ms)。各フレームに対して、自己相関法を用いて入力音声の音高を推定する [4]。推定結果と本来の音高を比較し、設定された閾値 (50cent) 以上にズレていた場合にミスとして検出する。ただし、一つの音符区間内でも局所的にミスをしてしまう可能性がある。これを考慮して、その音高がある区間の推定結果の平均値を求め、その値も検出に使用する。

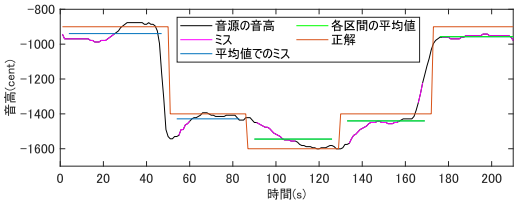

図 1. 音高のミスの検出

図 1 は実際に音程のミスを検出した図である。格納データは図にある通りである。二番目の区間を観察すると、前の区間からの落差により、明らかに歌い出しがミスしている。しかし、平均値だとミスしていないことになってしまっている。そこで、元々の推定音高を見るとちゃんとミスとして検出されていることが分かる。

2.2 タイミングのミス

楽譜情報における音符を演奏する瞬間をタイミングとする。与えられた楽譜情報に対して、人間が知覚できるレベルで演奏がズレてしまうことを「タイミングのミス」とする。主に「歌い出しが遅れてしまう」・「息が続かず、十分な長さ歌うことができず、途切れてしまう」・「途中の歌詞が早口で噛んでしまう」等があげられる。

楽譜情報より、全ての音符の始まりを求める。求められた区間において、いくつかのフレームに分けてスペクトル包絡を計測する。ここで、一般的に音符一つに対して、一音節が与えられる (例外あり)。また、スペクトル包絡は発音に深く関係している。このことより、音符の変わり目はスペクトル包絡が変化することが考えられる。まず、スペクトル包絡 (mfcc,12 次元) を計測する。求められたスペクトル包絡の差を利用して、録音された音源のタイミングを計測。楽譜情報より求められた区間内にタイミングが検出されなかった時、それをミスと検出する。

*: An Automatic Defect Detection for Computer Aided Vocal Production System *Daigo Kato (Hosei Univ.) et al.*

†法政大学 情報科学部

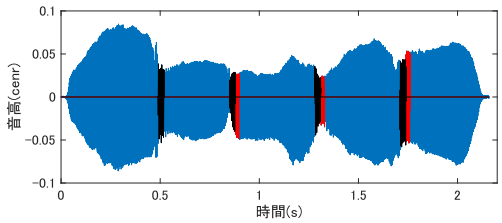

図 2. タイミングのミスの検出

図 2 は実際にタイミングのミスを検出した図である。黒が楽譜情報による正解のタイミング。赤が実際に検出されたタイミングである。この時、一番左のタイミングは赤が表示されておらず、タイミングとして検出されていないことが分かる。

2.3 音量のミス

楽譜には絶対的な音量は定義されていない (ただし、表現として「クレッシェンド」等の音量変化を定義するものはある。)。そこで、人間が聴覚的にメロディーの音量全体を通して、音量変化が著しくある場合をミスとする。主にあるミスとしては、「高音の音量が突出してしまい、うるさく聴こえる」・「低音の音量が小さすぎて十分に聴こえない」があげられる。

今回は音高が変化することが多いため、短時間パワーでは高い音は大きく、低い音は小さいといった問題を解決できない可能性が高い。これより、ラウドネスを用いて検出する [5]6。タイミングの時と同様に、楽譜情報を用いてメロディーの音符ごとの区間を切り出す。それぞれの区間に対してラウドネスレベルを計測し、区間内で偏差の合計を求め、閾値以上の場合ミスと検出する。また、歌声が音の高低に関わらず完全に一定である状態も不自然だと考えられる。その為、区間の周波数に応じてある程度値が異なることも許容する。

3 実験

提案機能の性能を測る為の評価実験を行った。使用データは、歌唱経験のある 20 代の若者男性 1 人が曲のワンフレーズを歌ったものである。曲調の一定化による結果の偏りを防ぐため、テンポが速い・遅いと音高が高い・低いを組み合わせた 4 種類の曲を計 10 曲用意した。さらに曲中での A メロとサビのように、2 種類のフレーズを各曲に対して歌ったので、計 20 のデータである。それぞれのデータに対して、筆者がミスを見つけ、それをミスの正解数とする。同じように、上記の方法で実装した機能によりミスの自動検出を行う。正解数に対する自動検出されたミスの数を検出率として算出した。ここで、音楽における表現技法を誤検出してしまう可能性がある。しかし、その問題は支援システムに「ユーザによるミスでないものの選択」が搭載されていれば解決する。よって、今回の実験では実装されていると仮定して行う。

結果として、音程のミス検出率 0.750・タイミングのミス検出率 0.625・二つの平均値 0.68・ミス全体の検出率 0.700 という結果になった。

音程のミスで検出できなかったもので、「歌い出しのピッチが不安定」がある。これは音高の平均値を用いることによって、不安定である時間が極端に短い場合での影響が少ないことが原因だと考えられる。これに対して、タイミングの検出の結果も利用し、歌い出しの処理だけ別のものにする必要がある。または、音源情報における音符区間内の全ての音高の平均値ではなく、一定数のフレームごとの平均値を求める必要がある。

次に、この機能はフレーム処理を行っている。それに伴って、楽譜情報の音かもフレーム単位に変換される。この処理によって、楽譜情報でのタイミングが分析する際のタイミングに差が生じてしまった。このことより、実際のタイミングではない箇所を正解のタイミングとしていることより、タイミングのミスの検出率が低下してしまったと考えられる。

4 結論

本研究では、音楽制作における最も回数が必要な「録音」とうい過程に着目し、繰り返しの録音による聴き直しを減らす為の機能を提案した。従来のカラオケ採点システムに加えて音量のミスも検出することによって機能を作成した。アルゴリズムの改善及び、歌声を含む音楽制作をしたことがある被験者を用意し、この機能の音楽制作への有用性を検証することが今後の課題である。さらに、発声は歌声の印象に深く関わっている。本研究の提案した機能に発声の良し悪しを判定するものを追加すれば、最早聴き直すことなく音源制作を完了できる可能性が少なからずあると考えられる。

参考文献

[1] Matthew Weiss, The Complete Guide to Vocal Production, 2018.July.30,https://theproaudiofiles.com/vocal-production/

[2] 下尾 波輝, 矢谷 浩司, エレキギター演奏におけるミスの自動検出, 2018 情処全文, 1 号, pp131-132, 2018

[3] 竹内 英世, 保黒 政大, 梅崎 太造, 人の主観評価に近いカラオケ採点法, 電学論 C, 130 巻, 6 号, pp1042-1053, 2010

[4] 竹内 英世, 保黒 政大, 梅崎 太造, カラオケ採点用の高分解能ピッチ抽出法, 電学論 C, 129 巻, 10 号, pp1889-1901, 2009

[5] 曽根 敏夫, 鈴木 陽一, ラウドネス, 音響誌, 44 巻, 10 号, pp1042-1053, 1988

[6] 栗原 信義, 高橋 信夫, ラウドネスレベルメータの開発, Vol.55, no.3, pp364-371, March, 2002