

音高を特徴量とした JPOP 楽曲における印象部分検出手法の提案

The proposal of the impression partial detection technique in a JPOP musical piece

山口 大志 佐々木将章 瀬川 典久 杉野 栄二 澤本 潤*

Summary.

In recent years, music distribution service starts on the Internet and musical audition service is offered. It has been increasing the needs to buy it after trying listening a musical piece. Therefore, in the music distribution service in CD Shop or the Internet, it is more general to offer audition service.

It is almost the case that it can try listening about 30 seconds in a musical piece. Therefore, it is not the parts which a customer wants to hear most in many cases.

The purpose of this research is the conventional technique's raising the rate of detection of the impression parts of a musical piece using a different technique, and improving the convenience of music audition service to a JPOP musical piece domain.

In this paper, the impression partial detection technique which made sound quantity the amount of the features is proposed.

1 はじめに

近年、インターネットで音楽配信サービスが始まり、音楽の試聴サービスが行われ始めた。それに伴い、音楽を購入する顧客の、試聴してから買いたいというニーズが高まってきている。そのため、CD ショップやインターネットでの音楽配信サービスにおいて、試聴サービスを行うことがより一般的になっている。試聴サービスにおいては楽曲中の30秒程度が試聴できることがほとんどである。よって、顧客がもっとも聞きたい部分、すなわちサビ（本研究では印象部分と呼称する）でないことが多い。

産業技術総合研究所の後藤真孝らは、繰り返し回数を元に、楽曲中のサビを検出する手法 RefraiD[1]を開発した。この手法では、4分の楽曲から印象部分を検出するのに1分要し、検出率は80%である。

本研究の目的は、JPOP 楽曲領域に対して従来手法とは異なる手法を用い、楽曲の印象部分の検出率を向上させ、音楽試聴サービスの利便性をよくすることである。近年の JPOP 楽曲では、オリジナルの曲と共にカラオケが提供されていることが多い。つまり、声が入っているものと曲のみの2種類の音楽情報が提供されている。そこで、オリジナルの曲から曲のみの情報を引くことが可能であれば、声成分のみが抽出可能である。そして、その取り出した声成分から音高を抽出し分析することでサビの部分が抽出可能である。

本稿では、検出率向上のために、音高を特徴量と

した印象部分検出手法を提案し、検出率向上を実証するためのシステムを作り、実験・検証を行う。

2 印象部分検出システムの提案

2.1 本システムの考え方

経験的に楽曲の印象部分にはもっとも高い歌声(声高)が使われていることが多い。そこで、本研究では音高を原曲データとカラオケデータから抽出し、その音高データの周波数成分を調べることでサビ区間を抽出する。

本システムは、原曲データとカラオケデータの2種類のデータが用意できることが前提である。1章で述べたように、近年の JPOP のシングル CD には、ほぼ原曲データとカラオケデータが含まれる。

原曲データは、曲構成区間の抽出と歌声データの抽出に利用される。曲構成データは、いわゆるコード進行を表したものである。オリジナルの楽曲データから和音データを取り出すことで、曲のどの区間からどの区間までが、どのコードに当たるのかが判別でき、曲全体に対し、曲の進行と構造を取り出すことが可能になる。

カラオケデータは、歌声データの抽出に利用される。原曲データの周波数成分から、カラオケデータの周波数成分を取り出すと、歌声部分が抽出される。そして、取り出された歌声成分から、どの時間帯がいちばん周波数成分が高い成分を含むかを調べる。この時間帯が、本論文で言うところの声高の部分であり、この部分にサビが含まれていると考える。

得られた声高の部分が、あらかじめ求めておいた曲構成データのどの部分に当たるのかを決定しその部分をサビとして決定する。

Copyright is held by the author(s).

* Taishi Yamaguchi, Masaaki Sasaki, Norihisa Segawa, Eiji Sugino and Jun Sawamoto, 岩手県立大学ソフトウェア情報学部

2.2 システム概要

本システムは、入力部、曲構成進行抽出部、歌声抽出部、音高抽出、印象部分抽出部から構成される。

入力部は、ユーザから原曲データおよび声が含まれていないカラオケデータを与えてもらい、そのデータをシステムに渡す部分である。本システムでは、原曲データとカラオケデータの2種類のデータをWAVE形式で入力する。また、入力されたWAVEデータは入力部で曲の情報を表すテキストデータに変換する。

曲構成進行抽出部は、原曲データから、JPOP特有の曲の進行を調べ原曲に対して、曲の構成を抽出する部分である。

歌声抽出部は、原曲データおよびカラオケデータを利用し、原曲データから楽曲だけを抜きだし、歌声だけを抽出する部分である。原曲データとカラオケデータの周波数成分を比較し、カラオケデータのみに含まれている物を原曲データから取り出すものである。歌声抽出部に関しては、歌声りっぷ [2] を利用する。歌声りっぷとは、楽曲の元データとカラオケデータを利用し、歌声のみを抽出するフリーソフトウェアである。厳密に言えば、歌声成分も取り除かれる可能性が在るが、本研究では、言葉を抽出するのが目的ではなく、その歌声から声高部分を抽出するのが目的であるために、実用上問題ない。

音高抽出部は、歌声抽出部で得られた歌声データを利用し、その歌声データのどの部分が声高なのかを調べ、抽出する。時間軸に対して、周波数成分毎の統計を取り、曲全体に対し、声高とみなされた時間を決定する。

印象部分選択部は、曲構成進行抽出部から得られた曲の構成情報と、音高抽出部から得られた声高と見なされた時間情報を重ね合わせ、サビ区間を決定する。決定したサビ区間が、全体のタイムラインに対してどの部分かをユーザに提示する。

図1にシステムの構成図を示す。

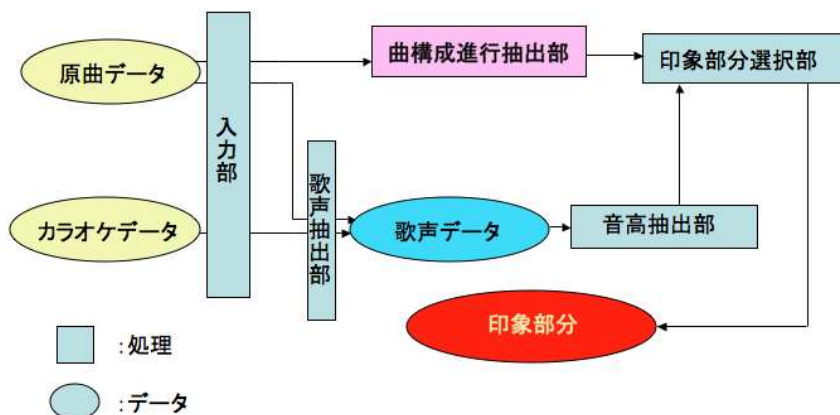


図 1. システム構成図

3 まとめ

本稿では、JPOP 楽曲を対象に、JPOP で提供されている原曲とカラオケデータを用い、声高部分を抽出し、曲のサビを検出するシステムを提案した。

今後、本研究によって、印象部分の検出率が向上できれば、音楽試聴サービスを利用する顧客に、もっと便利でスムーズな楽曲検索、楽曲選択が可能になると考える。すべての楽曲の印象部分が検出できるようになれば、楽曲を購入するときの手助けになるのではと考える。また、速度向上や印象部分自動再生機能の実装が可能になれば、よりよいサービスとして提供できると考えられる。

参考文献

- [1] 後藤 真孝. リアルタイム音楽情景記述システム. サビ区間検出手法, 情報処理学会 音楽情報科学研究会 研究報告 2002-MIUS-47-6, Vol.2002, No.100, pp.27-34, October 2002.
- [2] 歌声りっぷ. www.vector.co.jp/soft/win95/art/se127635.html