

骨伝導技術を用いたカラオケシステムの提案

川村 一輝[†] 及川 俊^{††} 荒川 淳平[‡] 太田 和夫^{‡‡}

[†]東京工業大学 機械知能システム学科 ^{††}, [‡]電気通信大学 情報通信工学科

^{‡‡}株式会社インフォクラフト

1. はじめに

1.1. 背景

現在のカラオケシステムではモニタに表示される歌詞を見ながら歌うのが通常である。このことにより、以下のような問題点があげられる。

- ・視力が弱い者には負担を強いる。
- ・聴衆と向き合って歌うことができない。

現在、騒音下での聴覚による情報伝達を可能にする技術として骨伝導技術があるので、この技術と組み合わせる事で上記の問題をクリアすることができないかと考えた。

1.2. 研究目的

骨伝導技術を用いて耳から歌詞情報を取得すると、歌いやすくなるかを確認するために、骨伝導技術を用いたカラオケシステムを試作する。

2. プロトタイプシステムの概要

2.1. 設計コンセプト

今回試作したシステムの概要図を図 1 に示す。

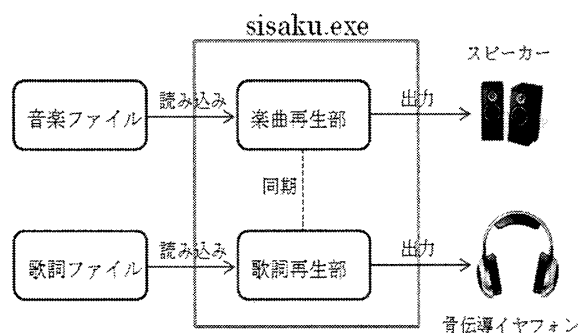


図 1. システム概要

A Proposal of KARAOKE System by Bone Conduction

[†] Kazuteru KAWAMURA

(kawamura.k.af@m.titech.ac.jp)

^{††} Takashi OIKAWA

(oikawa@ice.uec.ac.jp)

[‡] Jumpei ARAKAWA

(arakawa@infocraft.co.jp)

^{‡‡} Kazuo OHTA

(ota@ice.uec.ac.jp)

Dept. of Mechanical and Intelligent System
Engineering, Tokyo Institute of Technology. ([†])
Dept. of Information and Communication
Engineering, University of Electro-
Communications. (^{††}, [‡], ^{‡‡})
infocraft, inc. (^{‡‡})

WindowsPC 上で動作する sisaku.exe は、カラオケの BGM である音楽ファイルと、歌詞情報が記述された歌詞ファイルを読み込み、再生を開始してから楽曲が終了するまで同期しながら処理を行う。

スピーカーからは歌い手の声と BGM が出力され、骨伝導イヤホンを用いる事で歌詞情報は歌い手のみに伝わる。歌詞情報を取得するための骨伝導イヤホンは耳を塞がずともよいので、スピーカーから出力される BGM を聞くのにも不都合はないと期待できる。

2.2. 機能

sisaku.exe の楽曲再生部、歌詞再生部について説明する。

2.2.1. 楽曲再生部

多くの楽曲を再生可能とするために、幅広い拡張子に対応した WindowsMediaTechnology を利用した。

2.2.2. 歌詞再生部

歌詞を耳から伝えるのに以下の 2 つの方法を考えた。方式 1 … 曲に合わせて人間が歌詞を喋った物を録音してそれを再生する方法。

方式 2 … 歌詞を再生する時刻と内容をテキスト形式で記録して、発音は音声合成で行う方法。

既存のカラオケシステムには歌詞情報が記録されている事から、現在のカラオケシステムへの組み込みを考えた時に後者の方が自動化できる部分が多いと考え方式 2 を選択した。

(1) 歌詞再生デバイス

歌詞音声の出力先は、PC にインストールされている再生デバイスの一覧の中から選択する。また、BGM はその PC 上で再生デバイスの既定値に設定されているデバイスから再生される。

(2) 音声合成エンジン

sisaku.exe では音声合成エンジンとして Aquest 社の AquesTalk を用いた。他の候補として Microsoft の SpeechAPI, YAMAHA の VOCALOID 等があった。このうち AquesTalk は軽量で多くのリソースを必要とせず、インターフェイスが公開されている事から AquesTalk を選択した。なお、日本語の SpeechAPI は WindowsVista において不具合を生じる可能性があったことから選択しなかった。

(3) 聖徳太子モード

これは曲のテンポが速い、歌詞の文字数が多いといった事が原因で歌詞の読み上げに十分な時間が取れない場合を想定して今回工夫した機能である。

通常の再生では中央から歌詞が聞こえるのに対し、聖徳太子モードでは歌詞を前半と後半に分けて、右から前半の歌詞を、左から後半の歌詞を再生する。また、左から聞こえる歌詞は右から聞こえる歌詞よりわずかに遅れて再生される。

3. コンテンツの作成方法

試作システムの利用者は次の手順で音楽ファイル、歌詞ファイルを準備する。

3.1. 音楽ファイルの作成

音楽ファイルは普段 PC 上で音楽を聴く時に使っているファイルである。通常はカラオケ用のボーカルがカットされているものを使う事を想定しているが、ボーカルがカットされていなくても問題はない。

3.2. 歌詞ファイルの作成

歌詞ファイルは図 2 に示すような.txt 形式のファイルである。

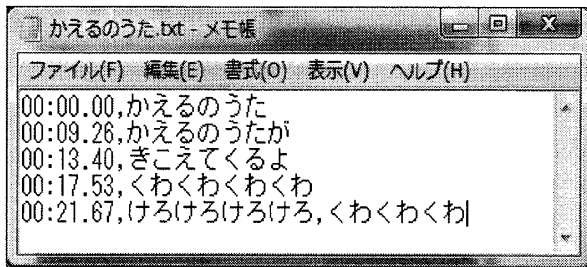


図 2. 歌詞ファイルの例

図 2 は「かえるのうた」の例である。歌詞を再生する時刻と再生する歌詞を[,]で区切って記述する。現段階では歌詞はひらがなで記述する仕様になっている。聖徳太子モードで歌詞の読み上げを行いたい場合は歌詞の二つに分割したい位置に[,]を入力する。図 2 の最終行は聖徳太子モードの例であり、右から「けるけるけるける」、左から「くわくわくわ」と聞こえる。

4. 実証実験

4.1. 目的

試作システムの有用性を検証し、改善点を洗い出す為に実証実験を行う。具体的に以下の 4 つの点について検証を行う。

- (1) 歌詞の聞き取りを正確に行えるか。
- (2) 正確なタイミングで歌唱できるか。
- (3) 歌詞を思い出す助けになるか。
- (4) 歌詞を見るのと聞くのでは違いはあるか。

4.2. 方法

- ・ランダム歌詞評価(目的(1), (2)を検証するため)

誰もが知っていると思われる「かえるのうた」の歌詞を異なる物にして、正確に発声できた文字数で評価を行う。

- ・既知曲評価(目的(3), (4)を検証するため)

歌詞の記憶が不完全な曲を歌って貰った後に、その曲と同程度に歌詞を記憶している曲を歌詞を見ながら歌って貰う。評価は被験者の感想(主観評価)で行う。

4.3. 結果

表 1 にランダム歌詞評価の実験結果を示す。

表 1. ランダム歌詞評価結果(文字数:25)

	見る	聞く
平均間違い数	1.5	18.8
間違い確率	0.06	0.75

表 1 から骨伝導イヤフォンを用いて音声合成で読み上げられた歌詞音声を正確に聞き取るのは容易ではない事がわかる。

既知曲評価では、被験者 16 名の内、骨伝導が助けになったという肯定的な意見が 6 名、否定的な意見が 7 名、無回答が 3 名となった。

また、被験者の自由記述意見の中で興味深かった意見として以下のようなものがあった。

- ・歌詩を聴くのに集中して、歌を歌えない。
- ・音声合成の質が悪い。(子音が聞き取れない)
- ・歌詞を囁くタイミングが悪いと歌い難い。
- ・歌詞につられる。

なお、歌わずに知っている歌詞を聴くという事は容易であった。

5. 考察

今回の実証実験の結果から、次の 2 つの事が重要であると言える。

- (1) 歌詞タイミングの最適化。
- (2) 歌詞音声の聞き取りやすさの追求。

(1) について、歌詞の囁きが歌いだしに被ると歌い難いという意見が多い。しかし歌い易い「囁きタイミング」には個人差があるようだ。よって、ある程度一般性があると思われるタイミングを自動生成し、実際に使用しながら、使用者が任意にタイミングを微調整できるという様なプログラムへの改良が考えられる。

また、歌詞タイミングの自動生成について、現在カラオケで使われている MIDI データがあれば可能でないかと考えている。何故ならば MIDI データは楽譜のような物であり、カラオケでも歌メロディの一音一音に対して歌詞が割り当てられている。音符と音符の時間間隔はテンポから計算でき、歌詞の読み上げに掛かる時間も計算できるため、この二つを比較すれば良いからである。

(2) について、歌詞音声の質を高めるためには、AquesTalk 以外の異なる音声合成エンジンを試してみたい。

6. 今後の課題

実際に歌唱しながらでは、歌詞を聞き取りにくい事が判明した。自分の歌声を聴くのに無意識に注意を奪われてしまうからだと推察する。

そこで、歌詞を聴く作業と歌を歌う作業を、どのようにしたら両立させることが出来るかがこのシステムを有効に出来るかのポイントになると思われる。

参考資料

AquesTalk:

<http://www.aquest.com/products/aquestalk.html>

UEC ソフトウェアレポジトリ:

<http://www.nishino-lab.jp/repository/index.html>

及川 俊: 「骨伝導技術を用いたカラオケシステムの実証実験とその考察」, 平成 21 年度電気通信大学情報通信工学科卒業論文 (2010 年 1 月)