

conteXinger : 日常のコンテキストを取り込み歌う VOCALOID

西村 綾乃¹ 椎尾 一郎¹

概要: 家事などの日常的な仕事は単調でつまらない作業, いわゆるルーティンワークになりがちである. また, 今日では, 音楽情報処理の技術が発達し, 携帯プレーヤーで音楽を再生するだけでなく, ソフトウェアを利用することで歌詞とメロディをコンピュータに歌わせることも可能となった. そこで, 本稿では位置情報や家電(掃除機, 冷蔵庫, 電子レンジ, 食洗機など)に取り付けたセンサ及びネット(SNS, 天気予報, web ニュースなど)から取得したユーザーの状況(コンテキスト: Context)に合わせた歌詞とメロディを自動的に作成しリアルタイムにVOCALOIDに歌わせることで, 生活に彩りを与えるシステムを提案し, その実装を行った.

conteXinger: VOCALOID takes the context of everyday in singing

NISHIMURA AYANO¹ SHIO ITIRO¹

Abstract: In general, every-day work such as housework is boring and monotonous work, and it tends to be so-called routine work. Work songs have been created and sung by workers to reduce their load of labor. Also, text-to-song synthesizer software such as Yamaha's VOCALOID is commonly used by wide variety of computer music creators. We have developed real-time music synthesizer named "conteXinger" that sings lyric text of the listener's context including home appliance activities (such as a vacuum cleaner, a refrigerator, a microwave oven and a washer) and the Internet information (such as SNS messages, web news, and weather reports). By presenting the synthesized music to a user through a home audio system or head phones, our system will entertain the user who is boring with every-day routine work.

1. はじめに

古くから人々は田植え歌・木挽(こび)き歌・酒造り歌などの仕事歌を歌ってきた。歌を歌うことにより, 労働に対する疲労や苦痛を軽減し, 自分を励まし, 気分の高揚が得られる。現代の日常生活でも, 掃除や料理などの家事の最中に歌を歌ったり, 鼻歌を歌ったりしている。また, 応援歌のように, 他人を励ますために歌われる歌もある。

一方, 携帯プレーヤーの小型化により, 今日ではそれを常に持ち歩き, 歌や音楽を聞きながら移動したり, 仕事や家事をすることが一般的になった。その際に, 歌や音楽によって, 気分が高揚したり, 応援される気持ちになる場合

も多い。また, ソフトウェアを利用することでその場で生成した歌詞をそれに合わせたメロディで歌わせることも可能になった。

更に, ユビキタスコンピューティングの研究分野では, 位置情報やセンサを利用して, 人々や物の状況(コンテキスト: Context)を認識し, 可聴化により情報提示したり [4], 効果音により調理を楽しくする研究がある [3],[1]。本研究では取得した生活コンテキストに依存した歌詞の楽曲を自動生成するシステム「conteXinger」を提案し初期段階の実装を行った。

2. conteXinger

conteXinger(コンテクシンガー)は, センサ及びネット上からユーザのコンテキストを取得することで, それに合わせた歌を自動的に作成し, 音声合成により歌声を合成し,

¹ お茶の水女子大学人間文化創成科学研究科
Ochanomizu University

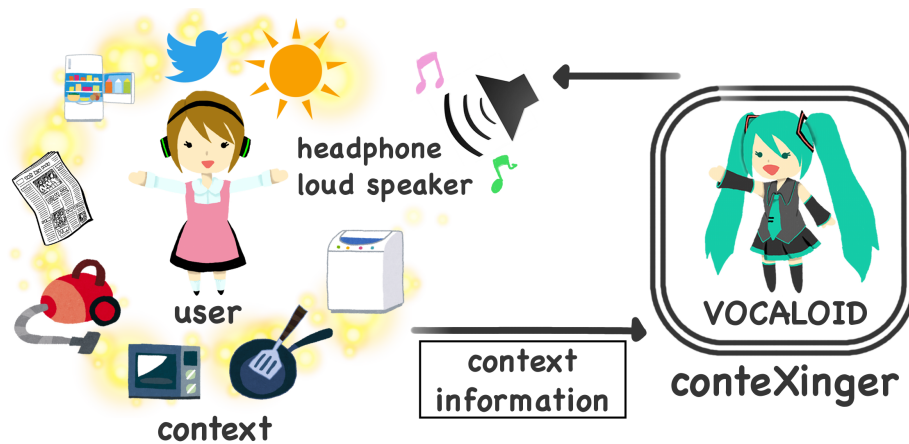


図 1 conteXinger

リアルタイムに歌い上げるシステムである。その構想を図 1 に示す。ユーザを取り巻く家電製品やネット上のコンテキストを取得し、それを元に歌とメロディを作り、音声合成により音楽を生成する。ここで、音声合成には現在広く使われているヤマハ（株）の VOCALOID などを使用することができる。

通常の商業音楽は、一般的に多数のリスナー向けに作詞されている。しかしその中に、たまたまりリスナーの境遇に近い状況が歌い込まれていると、強い共感を覚えることがある。それぞれのリスナーの状況を歌詞に取り込んだパーソナルな音楽が安価に容易に作成できれば、日々の生活に楽しさと潤いを与えてくれるであろう。

また、コンテキストが多く含まれた音楽の例としてミュージカルが挙げられる。ミュージカルの音楽は通常のポピュラー音楽とは異なり、脚本に沿った歌詞とメロディで物語を進めるため、より具体的に主人公の心情や状況、行動などが音楽の中に含まれている。

本システムでは、ユーザをミュージカルの主人公に見立て、センサから取得したコンテキストよりパーソナルな歌詞とメロディを作り出し、VOCALOID を利用して歌い上げることで、日常をミュージカルのような舞台に仕立ててユーザを応援する。これにより、家事などの日常的な仕事への意欲を増幅することを目指している。また、状況を歌に込めることで、天候、メールの着信、SNS の投稿、歩行中ならナビゲーションなどの情報をアンビエントに提示することも可能になる。更には、生成された曲を振り返ることで、日々のフィードバックとして活用することも考えられる。

3. 実装

家庭内におけるユーザのコンテキストを取得し、これから動的に音楽を作り出すシステムの試作を行った。歌詞化することで家事を応援できる可能性のあるコンテキストとしては、掃除、洗濯、料理、皿洗いなどが挙げられる。これに Twitter、天気予報、web ニュース、メール着信など

の状況を加えることで、変化に富んだ歌詞作成が可能になると考えた。

今回は、掃除機、冷蔵庫、電子レンジにセンサを取り付けコンテキストを取得し、さらに Twitter のユーザアカウントへのリプライ、天気予報の情報も取得することで、ユーザのコンテキストから日常に楽しさを与えてくれるような歌を歌うシステムを実装した。

Twitter のユーザアカウントへのリプライが多ければ、より自然で魅力的な歌詞の生成が期待出来る。そこで、他ユーザによるリプライの発信を促すことを目的として、掃除機の on/off、冷蔵庫の開閉時間、電子レンジの on/off についてツイートすることもできる。これらは、ユーザがどのコンテキスト取得時にツイートを投稿するか、アプリケーションのチェックボックスから選択することが可能である。

3.1 全体構成

図 2 に全体構成図を示す。本システムは無線センサネットワーク、Internet、クライアントコンピュータ及びサーバコンピュータによって構成されている。

3.2 コンテキスト取得方法

コンテキストを取得するために、家電である掃除機、冷蔵庫、電子レンジの状況をセンシングし、ネット上の情報である Twitter のメッセージ、天気予報の情報を取得した。

家電のコンテキスト取得にはマイクロコンピュータ Arduino、無線ユニット XBee を使用した。クライアントコンピュータにもペアリングした XBee を接続し、無線通信を行うことで、家事作業の邪魔にならないワイヤレスなコンテキスト取得を可能とした。

3.2.1 掃除機

掃除のコンテキストを取得するために、電源ラインにクランプ型交流電流センサを取り付け、また持ち手には加速度を検出するセンサを取り付けた。クランプ型交流電流センサを利用することにより、家電の電源ラインを切断する

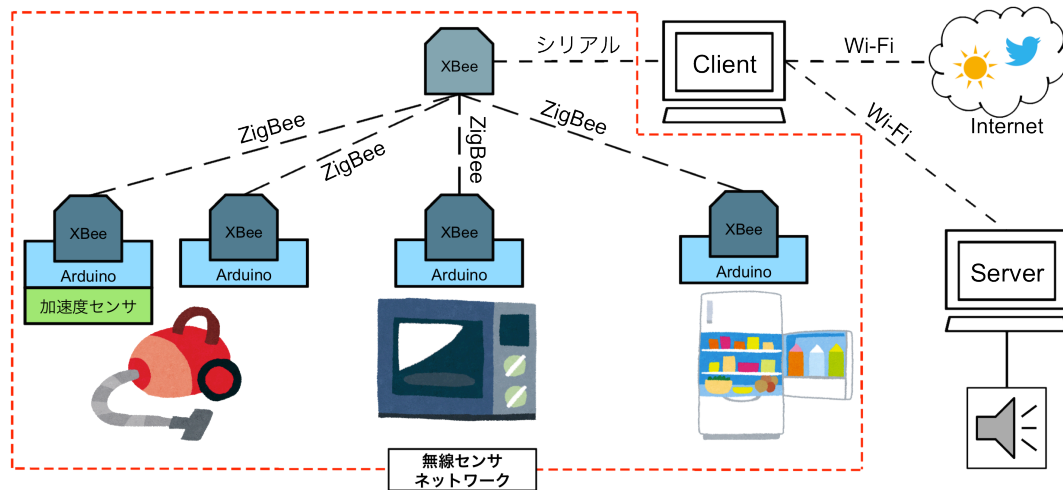


図 2 全体構成図

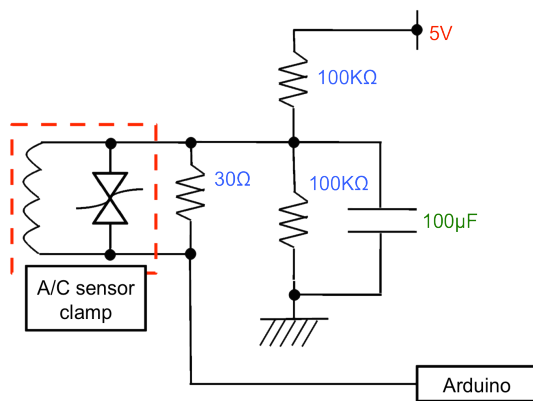


図 3 電源電流センサ回路図

ことなく容易に設置することができる。電流センサ出力を Arduino のアナログ入力に接続し、1秒毎に1ミリ秒間隔で電圧の値を取得し、20ミリ秒間の最高値と最低値を検出する。商用電源の周波数(50Hz)の周期が20ミリ秒間隔のため、最高値と最低値の差を求めることでon/offの取得が可能になる。その回路図を図3に示す。

掃除機には柄の部分に加速度センサを取り付けた。掃除機のスイッチがonの状態、加速度センサの値を取得し、前後運動の激しさからユーザの活動状況を判定する。現時点の仕様では「掃除機の柄を通常で動かしている」「掃除機の柄を速く動かしている」「椅子などの障害物を移動している」の3種類に分類している。

3.2.2 冷蔵庫

冷蔵庫のコンテキストを取得するために、リードスイッチを扉部分に取り付けた。リードスイッチの出力は Arduino のデジタル入力に接続した。これにより、冷蔵庫の扉の開閉を検知する。取得できるコンテキストは冷蔵庫の扉が開いている時間である。秒数によって表1のようにコンテキストが分類される。

3.2.3 電子レンジ

電子レンジのコンテキストを取得するために、掃除機と

秒数	コンテキスト
0~9 秒	通常の冷蔵庫の開閉
10~19 秒	開閉時間が長い
20 秒以上	扉の閉め忘れ

同様に電源ラインにクランプ型交流電流センサを取り付けた。これにより、電子レンジのon/offの検出する。

取得できるコンテキストは「物を温め始めた」「物を温め終わった」「電子レンジが1分以上稼働している」の3種類である。1分以上稼働している場合は、その後、1分毎に電子レンジの稼働が継続していることをコンテキストとしてクライアントに知らせる。

3.2.4 Twitter

クライアントでは、JavaのライブラリであるTwitter4J^{*1}を利用してTwitterの情報を取得している。これにより、12秒に1回の間隔でフォロワーから得られる新規リプライの有無を確認する。このリプライの文章はそのまま歌詞に反映される。

3.2.5 天気予報

天気予報API^{*2}を利用することで、その日の天気と詳細な情報を取得する。現在の仕様では、掃除機がonになっている場合に、天気の情報ランダムで歌う。今後の実装では、洗濯機のコンテキストに関連して天気予報のコンテキストを盛り込む予定である。

3.3 本システムのクライアント

図4にシステムブロック図を示す。クライアントでは各センサ及びネット上からコンテキストを取得し、作詞作曲を行っている。また、コンテキストを取得する際に必要なシリアル通信をJavaで実現するため、RXTXライブラリを利用した。

*1 <http://twitter4j.org/ja/index.html>

*2 <http://www.drk7.jp/weather/>

表 2 家電コンテキストの歌詞

コンテキスト	種類	歌詞
掃除機	掃除機の柄を通常で動かしている	ゆっくりお掃除、のんびりやりましょう
	掃除機の柄を速く動かしている	びっかびかにしてやんよ、この速さについて来られるか
	椅子などの障害物を移動している	頑張れ頑張れ、お手伝いしてあげたいな
冷蔵庫	通常の冷蔵庫の開閉	何を取り出したの、扉をあけるとひんやりしている
	開閉時間が長い	冷蔵庫開け過ぎだよ、電気代がもったいない
	扉の閉め忘れ	冷蔵庫閉め忘れてるよ、冷蔵庫早く閉めて
電子レンジ	物を温める始めた	早く温まらないかな、ほっかほかにしてやんよ
	物を温め終わった	あつあつになったね、熱いから気をつけて
	電子レンジが1分以上稼働している	まだかなまだかな、早くあつあつになーれ

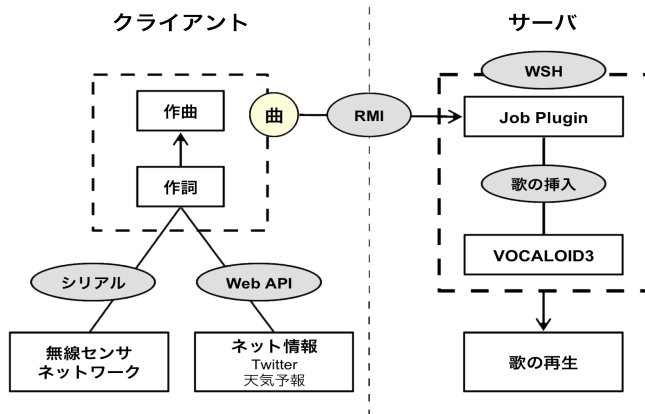


図 4 システムブロック図

センサまたはネット上からコンテキストを取得すると、クライアントは、まず、そのコンテキストに対応する歌詞を作成する。作成された歌詞より、メロディを付加し、サーバコンピュータに曲を送信するまでがクライアントの動作である。クライアント-サーバ間の通信は Java Remote Method Invocation (Java RMI) を利用することにより実現した。

3.4 本システムのサーバ

サーバではサーバプログラムの他に、VOCALOID が稼働している。ここでは、Job Plugin が利用可能な VOCALOID3 (以下 VOCALOID) を採用した。クライアントより送信された曲を受け取り、サーバが VOCALOID の機能である Job Plugin 及び Windows Script Host を起動することによって歌及び伴奏としてドラムビートが再生される。

Job Plugin とは Lua 言語によって書かれており、これを利用することで VOCALOID エディタ上に歌詞とメロディを挿入することが可能になる。また、Windows Script Host によりサーバの動作の一部は自動化されている。VOCALOID の制御が複雑なため、今回はこのようにサーバとして独立させた。

4. 音楽の生成

クライアントは、掃除機、冷蔵庫、電子レンジ、Twitter、

天気予報のコンテキストを取得すると、それに従って歌詞・メロディを作成する。曲は 1 フレーズごとに生成され、サーバでドラムの音と共に再生される。また、取得するコンテキストが無い場合は無音となる。コンテキストを適切に歌詞に反映させる方法は、コンテキストの種類に依存する。ここでは、現在実装されている掃除機、冷蔵庫、電子レンジの「家電のコンテキスト」、Twitter、天気予報の「ネット上のコンテキスト」を例にとって説明する。

また、Twitter や天気予報などネット上から取得した文章及び家電のコンテキストのために用意された文章は漢字まじりのため、それをひらがなに変換する際に Mecab^{*3} を利用した。メロディの音価の決定の際にも Mecab の形態素解析の結果を利用している。

4.1 歌詞

歌詞を VOCALOID に渡す際、漢字、カタカナ及び半角英数は歌い上げることができないため、クライアントプログラム内でひらがなに変換、または削除をする。漢字からひらがなへの変換には Mecab 利用し、カタカナからひらがなへの変換には文字コードの差分を使うことによって実現している。半角英数について、現在の実装では、すべて削除することによって対応しているが、今後は Mecab のユーザー辞書にひらがなの読み方を登録する予定である。

4.1.1 家電のコンテキスト

家電に取り付けたセンサから得られるコンテキストに関しては、あらかじめ歌詞を数個用意しておき、その中から内容に見合った歌詞をランダムで選出している。現在、登録されている歌詞の一例を表 2 に示す。

4.1.2 ネット上のコンテキスト

Twitter のリプライはそのまま歌詞に組み込まれる。これは、人間によって書かれた言葉が加工せずとも自然な歌詞になるということが期待できるためである。

また、天気予報では API リクエストにより返って来た天気、詳細情報の 2 つを歌詞として歌う。天気について歌う場合、「晴れ」「雨」「くもり」などの天気情報だけでは歌詞が短くなってしまい、1 フレーズの音楽生成を満了すこと

*3 <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

表 3 コンテキストによる調号の違い

コンテキスト	調号	コンテキストの詳細
掃除機	C-dur c-moll	掃除機の柄を通常で動かしている, 掃除機の柄を速く動かしている 椅子などの障害物を移動している
冷蔵庫	D-dur d-moll	通常の冷蔵庫の開閉 開閉時間が長い, 扉の閉め忘れ
電子レンジ	A-dur a-moll	物を温め始めた, 物を温め終わった 電子レンジが 1 分以上稼働している
Twitter	G-dur	リプライの内容に関わらず調号は固定
天気予報	F-dur f-moll	天気予報 天気についての詳細情報

表 4 天気予報に付加される歌詞

天気	歌詞
晴れ	ぼかぼか良い天気
くもり	空がどんよりしてるね
雨	ざあざあちやぶちやぶ雨が降る
その他	今日も 1 日頑張ろうね

が出来ないので, クライアント内で言葉を付加している.
 その内容を表 4 に示す.

4.2 メロディ

コンテキストに従い, メロディの「調号」「音程」「音価」を変化させる. 各コンテキストの違いを明確にし, 音によるコンテキストの違いの認知度を向上させることを目的としている. また, VOCALOID の仕様により, 拍子 = 4/4, テンポ = 120 は固定とした.

4.2.1 調号

各コンテキスト種類の違いは「調号」で表現することにした. 長調と短調をそれぞれのコンテキストに当てはめることで, コンテキストの中での違いを表している. その対応を表 3 に示す.

また, これらの調号に対するコンテキストの決定は筆者(西村)の共感覚^{*4}によるものである. それぞれの調号に対して感じられる色彩を一部コンテキストに割り当てた. 例えば, 筆者は G-dur から「青」を感じるため, アイコンが青で表現されている「Twitter」を当てはめた. これにより, 調号とコンテキストのマッピングが記憶しやすくなる. 現在, 各コンテキストに対しての調号は固定されているが, 今後はユーザーの個人の感覚に合わせて自由に設定できるように実装を進めたい.

4.2.2 音程

音程の決定の仕方は全てのコンテキストに共通である. 選択される音は各コンテキストとも中音域 1 オクターブ内に限定されている. これは, VOCALOID の声が聞き取りやすい音域で音程を決定するためである.

^{*4} ある刺激に対して通常の感覚だけでなく異なる種類の感覚をも生じさせる一部の人にみられる特殊な知覚現象をいう.
 (wikipedia より抜粋: <http://ja.wikipedia.org/wiki/共感覚>)

最初の 1 音目は 1 オクターブ内から乱数によって決定される. その次の音からは確率付きの乱数で決定され, 上 2 度, 上 3 度, 下 2 度, 下 3 度にそれぞれ 22.5% の確率で遷移, または, 上 4 度, 上 5 度, 下 4 度, 下 5 度にそれぞれ 2.5% の確率で遷移する. 次の音が前の音と離れすぎていると, 曲として突飛なものになってしまうため, このような確率の値を定めている.

4.2.3 音価

音価の決定の際, Mecab による歌詞の形態素解析の結果を利用している. 特に日本語の中で重要である「名詞」「形容詞」「動詞」の 3 種類の品詞の言葉の 1 文字目が, 音楽の強拍にあたる部分に来るように定めている. 強拍とは小節または拍子の中で強く拍を取る部分のことを言い, 4 拍子の中では特に第 1 拍と第 3 拍を指す.

また, コンテキストによって, 使用される音価は異なる. 掃除機の「掃除機の柄を速く動かしている」コンテキスト, 冷蔵庫の「扉の閉め忘れ」, 天気予報の「詳細情報」では音価を短くするため, 16 分音符, 8 分音符, 付点 8 分音符の 3 種類を使用している. その他のコンテキストは 8 分音符, 4 分音符の 2 種類である. それぞれに割り当てられた音価の中から, 乱数によって歌詞 1 文字につき 1 つが選択される. 調号の違いだけではなく, 音価での違いも出すことで, よりコンテキストの差を明確にすることができる.

最後の 1 音はその小節いっぱい伸ばすように定めた. そうすることで, フレーズの終わりを意識することが可能となる.

4.3 コンテキストの優先順位

本システムでは各コンテキストの演奏に優先順位を設けた. これにより, コンテキストを同時に取得した場合も演奏を可能にしている.

掃除機のコンテキスト演奏が最も優先度が低く, 他に歌うべきコンテキストが無い場合に音楽が作成される. 冷蔵庫の閉め忘れのコンテキストが取得された場合は, ユーザーが扉を閉めるまで歌を歌い続けるが, 途中で他のコンテキストを取得した場合はそちらを優先して歌を歌う.

冷蔵庫の開閉, 電子レンジの on/off 及び稼働時間のコン

テキストに関しては、歌われ待ちの曲が2フレーズ以上溜まっている場合は音楽が作成されない。これは、歌われていないフレーズを多数溜めることにより、ユーザの行動や状況と音楽が一致しなくなることを防ぐためである。

5. 関連研究

人間の行動や状況などのコンテキスト取得を実現するために、多くの研究や実装が行われている。例えばセンサを意識すること無くコンテキストを取得しアプリケーションの開発を平易にするためのミドルウェア [2] や、スマートフォンの加速度センサとマイクのデータを用いて行動認識をリアルタイムに行う研究がある [8]。

また、人間の状況や気分に合わせて音楽を選択または変化させる研究では、ジョギングのペースに合わせて再生している音楽のテンポが変化する研究や [9]、ユーザが受信したメールの内容から抱く感情に合わせた音楽を推薦する研究などもある [6]。

自動作曲及び作詞支援に関する研究では、歌詞の韻律を用いた歌唱曲の自動作曲の手法 [7] や、Ngram 言語モデルをもとに、ユーザの指定した音節数と韻の条件を満たす歌詞候補文を生成し、提示する作詞支援 [5] などがある。

6. まとめ及び今後の予定

本稿では、日常生活のコンテキストを取得し、それによる音楽生成について提案し、実装した。現在は、掃除機の on/off 及び加速度、冷蔵庫の扉の開閉時間、電子レンジの on/off 及び稼働時間、Twitter のリプライ、天気では予報及び詳細情報のコンテキスト取得を行っている。

コンテキストを音楽にする手法は、家事を応援するだけでなく、日常生活におけるアンビエントな情報提示に役立つと考えている。そこで、将来は、携帯音楽プレーヤーに組み込み、周辺の店舗情報の提示や、ナビゲーションなどへの応用も進めていく予定である。

今後は、多様なコンテキスト取得を可能にし、コンテキストを反映した表情豊かな音楽の生成を目指したい。

6.1 コンテキスト

今後の展望として、掃除機の位置を取得することで、掃除をしている部屋に合わせた曲の生成を予定している。また Twitter、天気予報以外に、ニュース、メール、ユーザがブラウザ上で検索したワードに関する情報などのネット情報を歌詞に反映することも考えていきたい。そして、掃除機、冷蔵庫、電子レンジ以外の家電製品や日用品への組み込みも進めたい。

更にはウェアラブルなセンサを利用することで、より詳細なユーザのコンテキストを取得することを考えている。そのために、スマートフォンをセンサとして取り込むことも検討している。

6.2 音楽

現在歌詞は、家電ではあらかじめ用意された数個の中から選択されているが、今後はユーザが自由に歌詞を編集できるように実装を進める予定である。また、名前や誕生日を登録することで、歌詞にユーザの名前を含めたり、誕生日には「Happy Birthday to You」を歌うなどのさらにパーソナルな歌を歌わせることも考えている。

一方で、多数のユーザが本システムを利用することを想定した場合、クラウド上にユーザが考えた歌詞を投稿してもらうことで、より面白い表現や秀逸な歌詞をシェアし、その中から気に入ったものを自由に選択して conteXinger に歌わせるようなサービスも考えられる。更には本システムでは仕様により、音楽が1フレーズずつ演奏されるので、歌詞は従来とは異なる短い言葉での表現が必要とされ、作詞手法が変化することも期待される。

メロディの生成では、現在、どのコンテキストも生成方法が同じなので、今後はコンテキストによって自動作曲アルゴリズムを変化させることを考えている。アルゴリズムを改善する際には、隠れマルコフモデルを用いた旋律生成手法を取り込む予定である。また、伴奏もドラムビートのみとなっているので、音階を持った伴奏を付随させたい。

参考文献

- [1] Halupka, V., Almah, A., Pan, Y. and Cheok, A. D.: Chop Chop: A Sound Augmented Kitchen Prototype, *Advances in Computer Entertainment '12*, pp. 494-497 (2012).
- [2] Salber, D., Dey, A. K. and Abowd, G. D.: The context toolkit: aiding the development of context-enabled applications, *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, New York, NY, USA, ACM, pp. 434-441 (1999).
- [3] 杉野 碧, 岩淵絵里子, 椎尾一郎: 家事と食卓を楽しくする「歌うダイニングキッチン」の試作, エンタテインメントコンピューティング 2007 講演論文集, pp. 67-70 (2007).
- [4] Yataka, S., Tanaka, K., Terada, T. and Tsukamoto, M.: A context-aware audio presentation method in wearable computing, *Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, New York, NY, USA, ACM, pp. 405-412 (2011).
- [5] 阿部 ちひろ, 伊藤 彰則: 統計的言語モデルを用いた作詞補助システム, 情報処理学会研究報告. [音楽情報科学], No. 9, pp. 1-6 (2011).
- [6] 阿部 健一, 藤本 悠, 大原 剛三: ユーザーが受信メールから抱く感情に合わせた音楽推薦システム, *DEIM Forum 2012* (2012).
- [7] 深山 寛, 中妻 啓, 米林 裕一郎, 酒向 慎司, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: Orpheus: 歌詞の韻律に基づいた自動作曲システム (自動作曲), 情報処理学会研究報告. [音楽情報科学], Vol. 2008, No. 78, pp. 179-184 (2008).
- [8] 大内 一成, 土井 美和子: ActivityAnalyzer: 携帯電話搭載センサによるリアルタイム生活行動認識システム, 情報処理学会研究報告. UBI, [ユビキタスコンピューティングシステム], No. 3, pp. 1-8 (2011).
- [9] 帆苺 隼佑, 長安 達也, 北原 鉄朗: ジョギングのペースに再生速度を同期させるスマートフォン用音楽プレーヤー, 情報処理学会第 75 回全国大会講演論文集, pp. 249-250 (2013).