

歌声合成技術 VOCALOID とその組み込み機器への応用可能性

Singing Synthesizer “VOCALOID” and its possible application to embedded devices

剣持秀紀 吉岡靖雄 (ヤマハ(株) 研究開発センター)
Hideki KENMOCHI, Yasuo YOSHIOKA (Yamaha Corporation))

kenmochi@beat.yamaha.co.jp, yass@beat.yamaha.co.jp

Abstract—This paper describes overview of the commercial singing synthesis software “VOCALOID.” A prototype board where its synthesis engine is ported to a DSP is also shown. Its application possibility as an embedded device is discussed.

1. はじめに

最近、歌声合成ソフトウェア VOCALOID を用いて、多くのクリエイターが音楽制作を行っている。ニコニコ動画」などの動画サイトでは、「初音ミク」を筆頭とする歌声合成ソフトウェア VOCALOID を用いて作成された楽曲が数多く投稿され、クリエイターたちが楽曲制作を日夜競っている。VOCALOID は PC 上での音楽制作に特化した歌声生成ソフトウェアであるが、歌声を合成するという機能を考えると PC 以外の環境でも幅広い応用可能性も考えられる。また、歌声以外の音声についても、韻律を自由自在に操作できるという点はこれまでにない応用を生み出す可能性がある。

本稿では、VOCALOID 歌声合成システムを簡単に紹介し、その合成エンジンを汎用の DSP に移植し、ハードウェアとして実現した試作ボード “VOCALOID-board” について述べる。

2. VOCALOID 歌声合成システム

VOCALOID はヤマハが開発し、ライセンスを行っている歌声合成ソフトウェアである。人間の歌声から取り出した音声素片を連結することによって歌声を合成する。入力された楽譜情報をもとに素片を選択し、接続することで合成を行う。その構成を Figure 1 に示す。

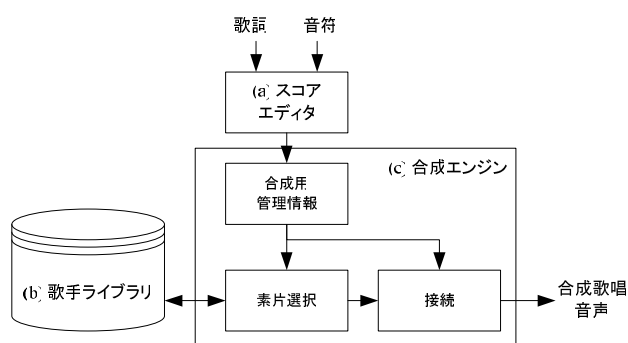


Figure 1 VOCALOID 歌声合成システム

ユーザはスコアエディタ(a)を用いて音符と歌詞を入力する(Figure 2)。歌手ライブラリ(b)には実際の歌手の歌唱データから取り出した音声素片が含まれる。合成エンジン(c)は歌手ライブラリから取り出した必要な音声素片を連結して合成する。

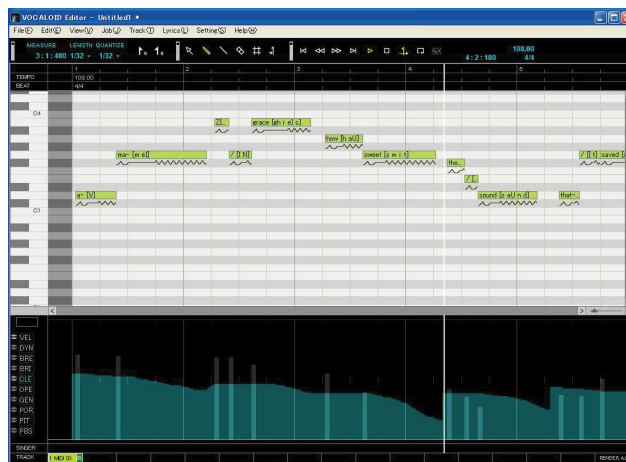


Figure 2 スコアエディタ

合成の際の素片の連結時には、C-V(子音-母音)という素片の V(母音)の位置に音符開始タイミングが合うように素片の位置の調整が行われる。

素片連結時には、単に連結しただけではもちろん歌にならない。素片のピッチを所望のピッチに変換すること、素片接続位置での音色の合わせこみが必要となる。VOCALOID ではこの2つを周波数領域での信号処理にて行っている。すなわち、STFT により求められたスペクトルを周波数軸上でスケールングすることでピッチを変換し、スペクトル包絡が時間的に滑らかになるように調整することで接続位置での音色の合わせこみを行っている。最後に IFFT(および Windowing & Overlapping)を行い合成波形を得る。

VOCALOID は歌声合成を第一の目的としているために、合成エンジンは音符と歌詞を入力としている。しかし一方で、歌声の合成から出発したという特質を活かし、韻律(イントネーションと音素継続長)を直接指定して合成することが可能なインタフェースも VOCALOID-flex という名称で提供されている。すなわち、韻律を与えれば歌声か話し声かを問わずそのまま合成することが可能である。

3. VOCALOID-board

VOCALOID は Windows PC 上で動作するソフトウェアなので、組み込み用途では制限が多い。そこで汎用の固定小数点 DSP に合成エンジンを移植し、小型のハードウェアとして実現したものが VOCALOID-board である。Figure 3. にその外観を示す。

VOCALOID-board の機能は以下の通りである。

- * MIDI 入力により、歌声や話し声を合成
- * 実時間にて合成を行い、出力する
- * 最大負荷動作時 1W 以下の低消費電力
- * 8cmx8cm のボード上に全機能を集約

歌声ライブラリは、PC ソフトウェア用のものがそのまま流用可能であり、SD カードにて提供される。

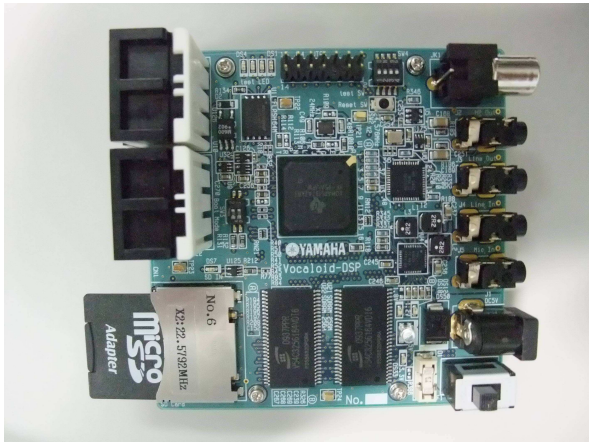


Figure 3. VOCALOID-board

VOCALOID-board は以下の各モードにて動作する。

(a) Playback モード

VOCALOID Editor にて作成したシーケンスデータ (VOCALOID-MIDI 形式) を再生するモードである。

(b) Realtime モード

あらかじめ歌詞を入力しておき、入力される MIDI の Note On/Off メッセージに従って発音するモードである。MIDI キーボード等を接続し、歌声を「演奏」することが可能。歌詞の入力は専用の MIDI メッセージを用いる。

(c) Voicesynth モード

音素長と各時刻でのピッチとダイナミクスを直接指定したものを入力として歌声や音声を出力することが可能なモードである。前述の VOCALOID-flex 機能に対応する。

4. VOCALOID-board の応用

VOCALOID-board は、用途として家電・業務用機器への歌声・音声合成機能の組み込みを想定している。特にエンタテインメントロボット分野で、インタラクティブな歌声の合成、表情豊かな話し声の合成機

能が簡単に実現可能である。この分野で、単なる波形再生や TTS (Text-To-Speech) では不可能な価値を提供していきたい。その価値とは、リアルタイム性、インタラクティブ性、エンタテインメント性である。

リアルタイム性とは、メッセージを受け取ったら直ちに再生可能ということである。機器のメイン CPU に負担をかけることなく歌声や話し声の合成が可能である。インタラクティブ性とは、状況に応じて発話内容を変更可能ということである。エンタテインメント性とは、発話内容そのものが、親しみやすく楽しめるものに成り得るという点である。

5. VOCALOID-board の今後

今後さらに小型化を進めていきたい。また、試作ボードの評価使用を含むアライアンスプログラムの提供により、パートナー企業との協業を通じて用途開発および要求仕様の絞り込みを進めていきたい。

参考文献

- 1) H. Kenmochi and H. Ohshita, VOCALOID - commercial singing synthesizer based on sample concatenation, Proc. Interspeech, pp. 4009-4010. (2007.8).