

歌声合成ソフトウェア VOCALOID3 と VOCALOID Job Plugin

剣持秀紀[†]

本稿では、2011年10月に発売された歌声合成ソフトウェア VOCALOID の新しいバージョンである VOCALOID3 について、そのポイント(品質面からの改良点、ユーザインタフェースの改良点)を紹介する。また新しく導入された VOCALOID Job Plugin という簡単なスクリプト言語で記述可能な機能拡張のための仕組みについて、その具体的な応用可能性について解説する。

Singing Synthesis Software VOCALOID3 and VOCALOID Job Plugin

Hideki Kenmochi[†]

In this paper, the author would like to introduce VOCALOID3, the updated version of the singing synthesis software VOCALOID, showing its advantages (improvements in synthesis quality as well as user interface). VOCALOID Job Plugin, a newly introduced system for functionality enhancement, will be described as well with its possibility in real applications.

1. はじめに

VOCALOID はヤマハが開発した素片連結型の歌声合成システムである[1]。

図1に示すように、歌詞と音符を入力するための(a)スコアエディタ、歌声の音声素片を集めた(b)歌手ライブラリ、それらを連結して歌声を合成する(c)合成エンジンからなる。

(a)スコアエディタを用いて楽譜(歌詞と音符など)を入力し、それに基づいて(c)合成エンジンは必要な音声素片を(b)歌手ライブラリから取り出し、連結する。その際、音色の合わせこみとピッチ変換の処理を周波数領域で行いながら素片同士を連結する。また、素片の中の母音の位置を音符の位置に合わせる処理も重要である。これを行わないと、タイミングが遅れて聞こえることになる。

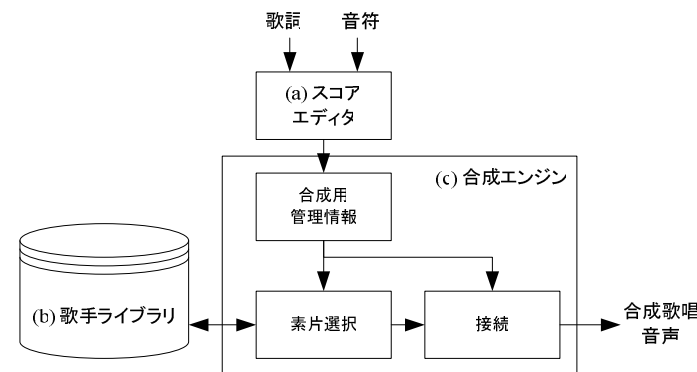


図1 VOCALOID 構成図

2004年に最初のバージョンを発売し、2007年に VOCALOID2 を発表し、これを用いたクリプトン・フューチャー・メディア株式会社の「初音ミク」が大ヒットし、これをきっかけに VOCALOID をはじめとする歌声合成技術にも関心が高まった。

VOCALOID3 は、2012年に発売を開始した新しいバージョンであり、VOCALOID2 を品質面、ユーザインタフェース面でさらに改良したものである。VOCALOID2 までとは異なり本バージョンからは(b)歌手ライブラリとソフトウェア(すなわち(a)スコアエディタと(c)合成エンジン)を別々に販売する形態となっている。

本稿では、VOCALOID3 について、VOCALOID2 に比べてまずは音質面でどのよう

[†] ヤマハ株式会社 yamaha+推進室
yamaha+ Division, Yamaha Corporation

に改善されたのかを解説する。そして、ユーザインタフェース、すなわち(a)スコアエディタの改善点について延べ、最後に新たに導入された機能拡張の仕組みであるVOCALOID Job Pluginについて説明する。

2. VOCALOID3 での改良点～音質面～

2.1 素片間の音色の補間

VOCALOID では、歌声の音色がピッチによって変化することを考慮するため、同一の音韻の組み合わせであってもピッチ方向に複数の素片を持っている。VOCALOID2 までは、ピッチが最も近い素片が選ばれて合成に使われていた。素片が切り替わる箇所では、音色が変化することになるため、切り替わる2つの素片の音色の差があまりない場合は問題にはならないが、音色が大きく異なる場合、明らかに不自然な合成となっていた。このため、VOCALOID3 の合成エンジンでは、ピッチ間での音色の補間を行周波数領域で行うことにより、この問題を解決した。

音色の補間は、スペクトル包絡を表す独自の音色パラメータ \mathbf{e} を補間することによって行われる。時刻 t における所望のピッチを $\hat{p}(t)$ として、 $\hat{p}(t)$ に近い2つの素片の補間する区間(例えば s-a の場合は[a]の区間)を S と S' とする。このとき、求めるべき音色パラメータ $\hat{\mathbf{e}}(t)$ は以下のように表される。

$$\hat{\mathbf{e}}(t) = r(t) \cdot \mathbf{e}_S \left(\frac{D_S}{\hat{D}} t \right) + (1 - r(t)) \cdot \mathbf{e}_{S'} \left(\frac{D_{S'}}{\hat{D}} t \right) \quad (1)$$

$$r(t) = \frac{\hat{p}(t) - \bar{p}_{S'}}{\bar{p}_S - \bar{p}_{S'}} \quad (2)$$

$\mathbf{e}_S(t)$, $\mathbf{e}_{S'}(t)$: それぞれ区間 S, S' の時刻 t における音色パラメータ

D_S , $D_{S'}$, \hat{D} : 区間 S, S' の継続時間長

\hat{D} : 合成時の S, S' に対応する区間の継続時間長

\bar{p}_S , $\bar{p}_{S'}$: 区間 S, S' の平均ピッチ

この時の補間の様子を図2に示す。

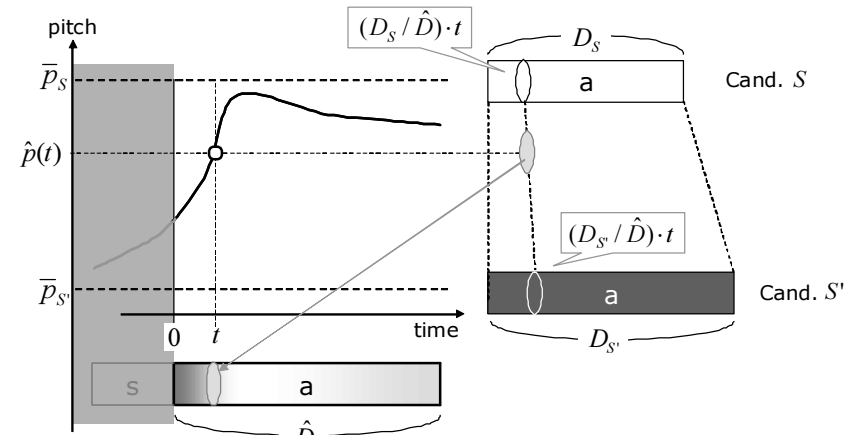


図2 音色の補間

2.2 より大きな合成単位の導入

VOCALOID2 までは、合成単位として VC および CV のみを用いていたが、VOCALOID3 からは、VCV という単位も使用可能となった。これにより、特に C が前後の音韻の影響を受けやすい子音の場合に、合成品質の改善となる。

2.3 子音伸長の改良

VOCALOID2 では、子音の長さを VELOCITY というパラメータを用いてコントロールすることができたが、例えば無声摩擦音の場合はあまり長くすると単に素片を時間伸長するだけでは金属的な音が発生することになるため、VELOCITY の効果を限定的なものにせざるを得なかった。VOCALOID3 では、子音の長さを伸ばしても自然性を損なわないような処理を行った。すなわち、無声子音の場合にはある程度の伸長が行われる場合にはホワイトノイズをソースとして使用し、もともとの素片から得られるスペクトル包絡でフィルタを行うようにした。また、破裂音を伸長する場合には、全体を伸長するのではなく、閉鎖区間のみを伸ばすようにした。

これらの処理により、VELOCITY による変化を大きくすることができるようになった。今までにない歌い方のニュアンスを簡単にらせるようになっていく。

3. VOCALOID3 での改良点～ユーザインタフェース～

ユーザインタフェース面でも VOCALOID3 では大きな改良を行なっている。VOCALOID2 までは、ピアノロール画面だけであり作業中のトラック以外のトラック

は透過的に表示されているだけであったが、VOCALOID3 からは DAW(Digital Audio Workstation)風のトラックエディタ画面を導入した。各トラックは複数の「パート」を持つことができるので、例えば1番を入力した後、2番を入力する場合も、1番をパートとしてコピーすることで作業効率が向上し、楽曲の構造もつかみやすくなった。

さらに、オーディオトラックを2つ(MONO、STEREO 一つずつ)用意し、合成音と同時に再生できるようになったので、例えば伴奏と同時に再生したり、プレス音などの効果音を付加することが簡単にできるようになった。

また、VST(Virtual Studio Technology)ホスト機能も装備し、VST 準拠のエフェクトプラグインを使用できるようになった。これにより、VOCALOID Editor で楽曲を作り込むことが可能になった。

図3に VOCALOID3 Editor のスクリーンショットを示す。

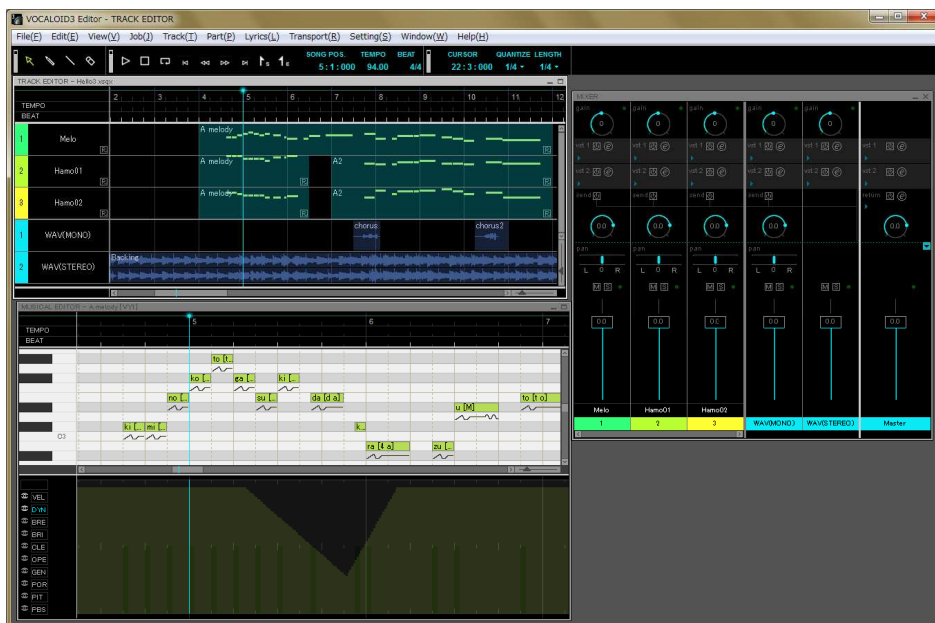


図 3 VOCALOID3 Editor

4. VOCALOID Job Plugin

VOCALOID Job Plugin は、各トラック内のそれぞれのパートについて、その中のシーケンスデータを外部から直接操作することができる仕組みである。例えば、以下のような操作が可能である。

- 音符の基本プロパティ(音程、音長、歌詞)の取得と更新
 - コントロールパラメータ(VELOCITY、DYNAMICS、PITCH BEND などの)取得と更新
 - パートのプロパティの取得と更新

プラグインのスクリプトは Lua 言語で記述する。Lua 言語については公式ホームページ[2]をご参照されたい。

VOCALOID Job Plugin のサンプル(音符をスタック化)の一部を図4に示す。

```
while (retCode == 1) do
    -- 適応範囲内かどうか判定する.
    if ((startTick <= note.posTick) and (note.posTick < endTick)) then
        -- 設定された Duration より長いノートの Duration を変更する.
        if (note.durTick > duration) then
            note.durTick = duration
        end
        -- 併せて velocity を変更する.
        note.velocity = velocity
        note.decay = decay
        note.accent = accent
        retCode = VSUpdateNoteEx(note)
        if (retCode == 0) then
            -- 更新エラー.
            endStatus = 1
            break
        end
    end
end
-- 次のノートを取得する.
retCode, note = VSGetNextNoteEx()
end
```

図 4 Job Plugin のサンプル(一部)

VOCALOID Job Plugin の開発キット(SDK)は、VOCALOID Store[4]にて入手可能である。また、現在のところ VOCALOID Store においてのみ、ユーザが作成した VOCALOID Job Plugin が公開可能となっている。

VOCALOID Job Plugin は、シーケンスデータを外部から自由に変更することが可能なので、ユーザが打ち込んだシーケンスデータに表情を加えるようなプラグインはもとより、自動作曲システムと組み合わせればユーザが歌詞だけを入力するとメロディーを付けてくれるプラグインなど、様々なアプリケーションが考えられる。

また、独立行政法人産業技術総合研究所が研究開発を行った VocaListener[3]についても、VOCALOID Job Plugin として発売される予定となっている。

5. 今後の展開

VOCALOID Job Plugin はエンドユーザにも VOCALOID の機能を拡張できるようになった。Job Plugin にとどまらず、合成エンジンの API の公開も今後行い、エンドユーザやパートナー企業が VOCALOID 合成エンジンを用いた独自アプリを開発できるように進めていきたい。

参考文献

- 1) 剣持秀紀, 大下隼人, “歌声合成システム VOCALOID-現状と課題”, 情報処理学会 研究報告, 2008-MUS-74-9, No.12, pp51-58 (2008)
- 2) Pontificia Universidade Católica do Rio de Janeiro, “The Programming Language Lua”, <http://www.lua.org/>
- 3) 中野倫靖, 後藤真孝, “VocaListener: ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案,” 情報処理学会 研究報告 2008-MUS-75 Vol.2008 No.50, pp.49-56 (2008)