

協調的創造活動支援を目的とした 歌声合成基盤技術の研究開発

森 勢 将 雅^{†1} 河 原 英 紀^{†2} 小 川 真^{†3}

急速に普及した動画共有サイトとコンテンツ制作支援のソフトウェアは、アマチュアクリエイタにも作品を公開する場と技術を与えた。現在では、複数のクリエイタの協調的な創造活動により、高い品質のコンテンツが多数生み出されている。また、クリエイタが利用するためのソフトウェア開発にも協調的な創造活動が行われ、優れたソフトウェアが創出されている。本稿では、歌声合成ソフトウェアをターゲットとし、誰でも利用可能な歌声合成技術を開発・公開することで、歌声合成ソフトウェア開発者間で協調的創造活動が創出されるか社会実験を試みた結果について示す。

Development of a Singing Synthesis Technology for Supporting Collaborative Creation

MASANORI MORISE,^{†1} HIDEKI KAWAHARA^{†2}
and MAKOTO OGAWA^{†3}

Video-sharing websites and advanced software enable amateur creators to upload their music, video and so on. Not only that, collaborative creation by several creators is carried out as an interesting style of creation. Target of collaborative creation is not only the contents but also software to support creators. In this article, a singing synthesis technology is introduced for encouraging collaborative creation of software developers. A social experiment in singing synthesis is carried out, and its results are discussed.

†1 立命館大学
Ritsumeikan University

†2 和歌山大学
Wakayama University

†3 電気通信大学
The University of Electro-Communications

1. はじめに

近年急速に発達した情報技術は、プロフェッショナルだけではなく、アマチュアのクリエイタに対しても多様な創造活動をするためのツールや場を与えている。特に顕著な例としては、歌声合成ソフトウェアである Vocaloid2 初音ミク¹⁾ と、制作したコンテンツを発表する場であるニコニコ動画^{*1} と Youtube^{*2} が挙げられる。コンテンツ制作の幅を広げるためのソフトウェアの発展と多数のクリエイタがコンテンツを発表可能な場の普及は、インターネット上での活動の枠組みを超え、現在では Vocaloid を用いて制作された CD がオリコンのアルバム週間ランキングで 1 位を獲得するなど現実世界にも影響している。

Social Creativity²⁾ と呼ばれる複数人により相補的に発揮される創造性は、動画共有サイトに適した創作の新たなスタイルとして定着しつつある。協調的創造活動そのものも 1 つの研究対象として扱われるほど注目度は高い³⁾。動画共有サイトの普及以来、初音ミクを用いた多数の楽曲がアップロードされ、その曲に対するプロモーションムービーの作成などの制作によりコンテンツとしての完成度が高まっていく様子は、まさに Social Creativity と呼べるだろう。近年の協調的創造活動は、コンテンツ制作だけではなくコンテンツ制作を支援するためのソフトウェアにも波及している。歌声合成ソフトウェア UTAU^{*3} や、3DCG ムービー制作ツールである MikuMikuDance^{*4} はその代表例である。これらのソフトウェアを用いてさらに品質の高いコンテンツが制作され、制作されたコンテンツから得られたフィードバックを活かしてソフトウェアが改良される過程は、動画共有サイトという媒体を通じて行われた新しい創作のスタイルといっても過言ではない。

本稿では、歌声合成ソフトウェアに対して協調的創造活動を促す基盤技術を確立することで、歌声合成ソフトウェア開発やコンテンツ制作に対する協調的創作活動の支援が可能であるか、実験的に検討した結果について報告する。以下では、コンテンツ制作に用いられる歌声合成ツールとそれらの概観、および求められる機能について示す。それらの要求を満足する歌声合成の基盤技術を示し、その技術や機能について説明する。最後に、提案した基盤技術が Web を通じてどのような影響をもたらしたのかについて考察し、提案する歌声合成技術の有効性や求められる課題を明らかにする。

*1 <http://www.nicovideo.jp/>

*2 <http://www.youtube.com/>

*3 <http://utau2008.web.fc2.com/>

*4 <http://www.geocities.jp/higuchuu4/>

2. 歌声合成に関するソフトウェア

ここでは、協調的創造活動に用いられる歌声合成ソフトウェアとして、YAMAHA 株式会社、クリプトン・フューチャー・メディア株式会社から発売されている Vocaloid シリーズ⁴⁾、歌声合成ツール UTAU、Vocaloid のファイルフォーマットである vsq を操作するエディタ Cadencii^{*1}について説明する。

歌声合成ソフトウェアには、入力された歌声を加工する Melodyne^{*2}や Auto-Tune^{*3}も存在するが、本稿では、Vocaloid のように歌詞と譜面から歌声を合成するものとして歌声合成ソフトウェアを定義して用いる。

2.1 Vocaloid シリーズ

Vocaloid は歌詞と譜面を入力とし、それらに基づいて合成された歌声が出力となる歌声合成ソフトウェアである。Vocaloid シリーズでは、1つのソフトウェアに1名の歌手のライブラリが含まれている。譜面情報以外にもピッチや音色の細かい制御が可能であり、それらのパラメタ調整がコンテンツの品質を決定付ける要因となる。初音ミクは Vocaloid2 と呼ばれるキャラクターボーカルシリーズの第一弾として発売されたソフトウェアである。しかしながら、近年では、初音ミクを CM のキャラクタとして起用するなど、それ自身が1つのキャラクターとして地位を確立しつつある。

2.2 歌声合成ツール UTAU

UTAU^{*4}は、Vocaloid と同様に、歌詞と譜面から歌声を合成するソフトウェアとして普及している。元々は音声ファイルを切り貼りする機能しか持たなかったが、ピッチ制御等の歌声加工機能が実装されたことにより、Vocaloid と同様に歌詞と譜面から歌声が合成できるようになった。UTAU は、歌声合成に用いるライブラリに関するフォーマットが明らかにされており、誰の音声でも自由に歌わせることができる。現在 3000 を超える音源が公開されており、特に初期に作られたいくつかの音源は、初音ミク同様にそれ自身がキャラクタとして高い人気を有する。例えば、UTAU 用のライブラリである重音テト^{*5}は、音楽ゲームである「初音ミク -Project DIVA- 2nd」にオプションのキャラクタとして追加すること

ができるほど、キャラクタそのものが注目されている。

2.3 Cadencii

Cadencii は、Vocaloid のファイルフォーマットである vsq ファイルを編集するエディタとして実装されたソフトウェアである。Vocaloid では譜面に基づくピッチに加え、ピッチの微細構造など詳細な情報を有する vsq ファイルに基づいて歌声を合成するため、vsq ファイルを編集できる Cadencii により Vocaloid エディタの代用が可能となる。UTAU はライブラリ、合成エンジン、エディタが独立しているため、Cadencii で制御したピッチ情報を用い、UTAU の合成エンジンを使うことで再生することも可能である。

2.4 歌声合成ソフトウェアの概観と本研究の目的

これらをまとめた歌声合成に関する概観を図 1 に示す。UTAU 用のライブラリ構築に関しても歌声の収録だけではなく音源設定と呼ばれる音素の切り分け作業が存在する。それらを多数の人間で独立して行うことや、音源設定のサポートツールが開発されていることから、ライブラリ構築に関しても協調的創造活動が行われている。歌声合成ソフトウェアの開発に関しては、音声合成に関する基盤技術や信号処理の知識が必要なことから、他の分野に比べて発展速度が遅いことが問題点と考えられる。歌声合成ソフトウェア開発に役立つ信号処理の基礎理論の構築、およびその理論を誰でも使えるように配布することができれば歌声合成ソフトウェア開発にも協調的創造活動が生じる可能性がある。

2.5 歌声合成ソフトウェアに対する要求

Vocaloid、UTAU に共通する機能として、歌詞と譜面から歌声を合成できること、ピッチや音色の詳細な制御が可能であることが挙げられる。また、コンテンツ制作に対する前提として、歌手の個性が崩れずに高い品質で合成できることも挙げられる。前者については、事前に多数の歌声を収録し、音節単位、あるいは数モーラの単語単位で保持しライブラリ化すること、およびその作業を支援するソフトウェアの実現によりある程度解決されている。したがって、ピッチや音色の制御が可能であり、高い品質で歌声合成が可能な信号処理技術を実現することが要求事項となる。

Vocaloid 型の歌声合成フレームワーク (図 2) のうち、本稿では主に (2) に対する新たな信号処理技術を与えることが目的といえる。ただし、合成プロセスに関しては、音声そのものの分析合成だけではなく、譜面とライブラリからどのように音声を接続するべきかという課題も存在するが、本稿では対象としない。

*1 <http://www9.atwiki.jp/boare/pages/18.html>

*2 <http://www.celemony.com/cms/>

*3 <http://music.e-frontier.co.jp/product/at7/>

*4 正式名称は歌声合成ツール UTAU であるが、本稿では単に UTAU と表記する。

*5 ただし、重音テト自体の発祥は 2ちゃんねるであり、UTAU 用ライブラリとして公開されたのはその後である。

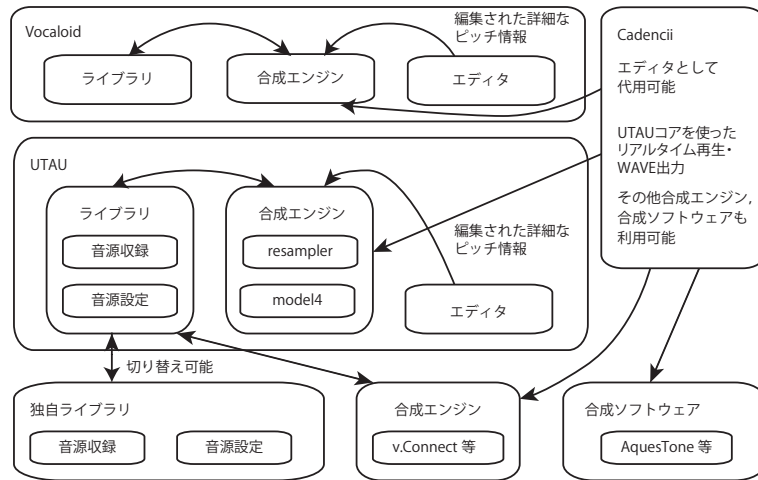


図1 各技術の位置づけ
Fig.1 Overview of the singing software

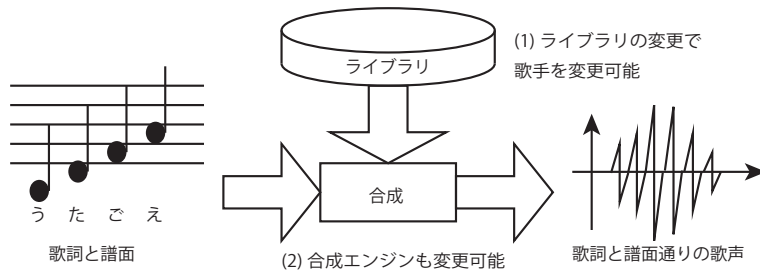


図2 歌声合成ソフトウェアに必要なフレームワーク
Fig.2 Framework required for singing synthesis software

3. 歌声合成に用いる信号処理技術の従来法

歌声に限らず音声を合成する技術は多数存在する．ここでは，前節で示した要求を満足する音声合成システムの従来法である STRAIGHT⁵⁾，TANDEM-STRAIGHT⁶⁾ について説明する．次いで，協調的創造活動を目指した基盤技術である WORLD について示す．

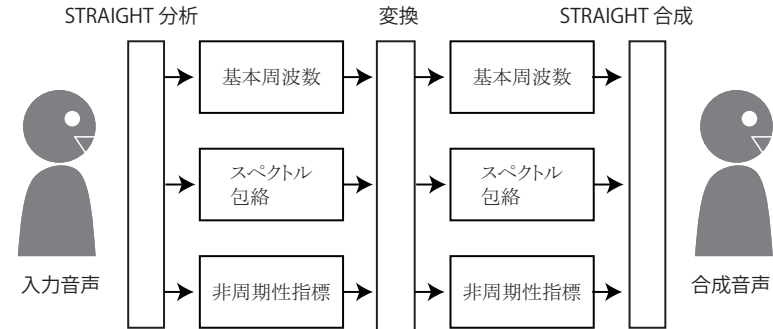


図3 STRAIGHT (TANDEM-STRAIGHT) の処理の流れ
Fig.3 Overview of STRAIGHT (TANDEM-STRAIGHT works the same way)

3.1 STRAIGHT, TANDEM-STRAIGHT

STRAIGHT と TANDEM-STRAIGHT (実質的に等価であるため以下では両方を STRAIGHT と呼称する) は，1939 年に提案された Vocoder⁷⁾ と同一の機構を有する方式であるが，元音声に匹敵する高い品質で音声合成できる利点を有する．図3に示される通り，STRAIGHT は，音声から基本周波数 (ピッチ)，スペクトル包絡 (音色)，非周期性指標 (擦れの程度) を取り出す．推定された各パラメタは自由に操作できるため，歌声合成に求められるピッチのみ，音色のみの制御が実現可能である．

STRAIGHT は，研究用途であれば自由に利用することが可能であるが，商用利用には有償の契約を結ぶ必要があることが問題となる．例えば，STRAIGHT により合成された歌声を含めた CD を販売することが可能かどうか，使用者は容易に想像できない．現在協調的創造活動で用いられるソフトウェアや技術は，全てフリーソフトで制約無く使えることが大きな特徴である．本研究では，権利問題が無く誰でも自由に使える音声合成システムの実現を目指し開発された WORLD⁸⁾ を基盤技術として用いる．

4. WORLD

WORLD は，商用利用を含めて利用が自由なフリーソフトとして，Web を通じて誰でも利用することができる*1．また，ライブラリ化はせずソースコードそのものを配布している

*1 <http://www.aspl.is.ritsumei.ac.jp/morise/world/>

ことも大きな特徴といえる。これは、WORLD を使用する開発者はプログラミングに精通していることが予想されることから、開発者自身で開発するソフトウェアの機能に併せてソースコードを改良する自由度を残す意図がある。

WORLD は、STRAIGHT と同様に基本周波数、スペクトル包絡を以下に示す方法により推定するが、非周期性指標については品質を向上させる目的で別のパラメタを採用している。以下では、各分析法の概要を示す。

4.1 DIO: 基本周波数推定

基本周波数は、有声音に含まれる声帯振動の間隔を示すパラメタであり、一般的な音声の高さとして用いられるピッチと概ね対応する。人間の音声における基本周波数は時間とともに変化するため、周期性を仮定できる短い区間の波形を切り出して、時間軸における相関や、周波数軸におけるスペクトルから抽出された特徴量に基づいて推定する方法が一般的である⁹⁾。高い品質を達成するためには、推定ミスが最小限となるような方法を用いる必要があり、近年では1%以下のエラー率を達成できる方法が複数提案されている^{10),11)}。

DIO (Distributed Inline-filter Operation) は、音声のSNRが十分に高いという制約を満たす音声のみを対象に、極めて高速にF0を推定できる方法として提案された¹²⁾。周期的に生じる声帯振動の波形から求めたスペクトルには、基本周期の逆数 f_0 とその整数倍の周波数にピークを有する特徴がある。一般的な基本周波数抽出では、SNRを稼ぐためにスペクトルにおけるピークの周期性などを評価することで高い精度を達成するが、Vocaloidのような歌声合成用のライブラリ構築に用いられる音声のSNRは十分高いことが期待される。したがって、 f_0 から $2f_0$ の間にカットオフ周波数が存在する低域通過フィルタにより波形を処理することで、 f_0 が推定できることがDIOの基本的な考え方である。

ただし、 f_0 が未知であるため、DIOでは様々なカットオフ周波数を有する複数の低域通過フィルタにより波形を処理する。得られた複数の波形から f_0 の信頼度に相当するパラメタを計算することで、未知の基本周波数に対しても高い精度で推定している。DIOは近年提案された高精度な方法とほぼ等価な精度を達成しつつ計算コストの大幅な削減に成功している。方法の詳細については、文献12)を参照されたい。

4.2 STAR: スペクトル包絡推定

音声のスペクトルには周期的な生じる声帯振動(f_0)の影響とスペクトル包絡とが混在しており、スペクトル包絡推定においては声帯振動の影響を取り除くことが中心的な課題といえる。LPC (Linear Predictive Coding)¹³⁾等が主要な方法であるが、高い符号化効率では達成されるものの高い品質で音声を合成することを旨とした推定法の実現は困難な課題であっ

た。STRAIGHT や TANDEM-STRAIGHT では、ピッチ同期分析¹⁴⁾の考え方に基づき、 f_0 の影響を高精度に分離していた。STAR¹⁵⁾も基本的な考え方はSTRAIGHTと同様であるが、窓関数の種類、長さ、ケプストラム法に基づくスペクトルの平滑化に特殊な条件を与えることで高い精度を達成している。

STARでは、以下の式によりスペクトル包絡 $|S_s(\omega)|^2$ を推定する。なお、 $x(t)$ は入力波形、 $w(t)$ は切り出しに用いる窓関数である。STARでは窓関数の種類と長さに制約があり、窓長が基本周期 T_0 の3倍のハニング窓を用いる必要がある。

$$|S_s(\omega, \tau)|^2 = \exp \left(\frac{2}{\omega_0} \int_{-\frac{\omega_0}{2}}^{\frac{\omega_0}{2}} \log (|Y(\omega + \lambda, \tau)|) d\lambda \right), \quad (1)$$

$$Y(\omega, \tau) = \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-2\pi j\omega t} dt, \quad (2)$$

τ は切り出し時刻、 ω_0 は基本周波数の角周波数表記($= 2\pi f_0$)を示す。なお、 $Y(\omega)$ における負の周波数のスペクトル値は、折り返しにより得られる値をそのまま用いることとする。STRAIGHTとの相違点や高品質音声に適している理由については文献15)を参照されたい。

4.3 PLATINUM: 励起信号抽出

STRAIGHTでは音声に含まれる非周期的な成分を合成結果に反映させるため、非周期性指標と呼ばれるパラメタを用いていた。非周期性指標は、擦れの程度に対応するだけでなく、音声を合成する際の励起信号の制御に用いられるパラメタであり、いわゆる合成音らしさを低減し自然性を向上させることを可能にしている。WORLDでは、非周期性指標ではなく、音声波形から励起信号を直接抽出するアプローチを採用している¹⁶⁾。

詳細は文献16)に譲るが、概要は以下の通りである。基本周期 T_0 で任意の音色を有する有声音 $y(t)$ は、基本周期 T_0 を有する励起信号 $x(t)$ と音色を示すスペクトル包絡から求められるインパルス応答 $h(t)$ との畳み込みで表現される。

$$y(t) = h(t) * x(t), \quad (3)$$

$$Y(\omega) = H(\omega)X(\omega), \quad (4)$$

STRAIGHTでは、励起信号は非周期性指標とスペクトル包絡により決定することが特徴といえる。WORLDでは、収録された音声波形 $y(t)$ が既知であることに着目し、波形から励起信号を直接求める点で大きく性質が異なる。基本周波数とスペクトル包絡は事前に推定されていることと、有声音の合成はスペクトル包絡から求められる最小位相応答と励起信号

とを畳み込みその応答を基本周期 T_0 に基づいて重ね合わせで実施される。これらの特徴から、励起信号 $x(t)$ と最小位相応答 $h(t)$ の畳み込みが波形 $y(t)$ と一致するように $x(t)$ を決定する。詳細は省略するが、STAR により求められたスペクトル包絡は最小位相であり振幅にゼロ点を持たないという逆フィルタ設計における要求を常に満足しているため、単純なスペクトルの割り算で $x(t)$ を求めることが可能である。スペクトル包絡から求めた最小位相応答と励起信号との畳み込みは必ず切り出された波形に一致するため、PLATINUM の考え方は TD-PSOLA¹⁷⁾ を拡張し Vocoder と同様にスペクトル包絡の制御が可能な自由度を与えたと位置づけられる。

また、 $x(t)$ は短時間毎に求める必要があるが、現在の WORLD では、有声音区間で振幅が最大値を示す時間を中心とし、TD-PSOLA と同様に基本周期の 2 倍の長さを有するハニング窓により切り出された波形を $y(t)$ として扱っている。その中心時刻と事前に推定された基本周波数の軌跡から、残りのピーク時間を自動的に算出している。

4.4 WORLD の性能

文献 16) で行われた MUSHRA¹⁸⁾ 評価において、WORLD の合成品質は従来法と比べて優位に高く、実質的に元音声と等価であることが示されている。また、基本周波数の制御が品質に与える影響に関しても従来法よりも少なく、歌声合成のように基本周波数の制御が中心となるソフトウェアに対しては STRAIGHT より適していると考えられる。

WORLD をソースコードを含め全てオープンにし、Web と Twitter を用いて歌声合成ソフトウェア開発者に対して宣伝を行った。これらの活動が歌声合成ソフトウェア開発に対する協調的創造活動の支援に繋がるか、本プロジェクトを開始した 2010 年 4 月から現在までの経過を実験結果として報告する。

5. 歌声合成ソフトウェア開発に対する創造的活動の経過

ここでは、WORLD を用いて制作されたコンテンツ、歌声合成ソフトウェア (UTAU 用の合成エンジンも含む) について述べる。コンテンツやツールの開発者については、本名ではなくニコニコ動画や Twitter のアカウント名のみしか確認されていないものも含まれる。ここでは、それぞれの制作者に直接コンタクトを取り、本稿に載せる許可を頂いたもののみ紹介する。

5.1 UTAU 用の合成エンジン: EFB-GW, world4utau, v.Connect 等

WORLD を用いた UTAU 用合成エンジンとして、筆者らが実装した EFB-GW (Eternal

Force Bresampler Gently Weeps)^{*1}, world4utau, v.Connect が開発され配布されている。Zteer 氏による合成エンジン (名称未定) も現在開発中という報告も受けており、現在サンプルの歌声を聞くことができる。それぞれを用いて公開された歌声のコンテンツもニコニコ動画に投稿されており^{*2}, コメントで合成エンジンにより音色が異なることが指摘されていた。これは、歌声合成において重要となるピッチ制御、時間長の伸縮をどのように実装するかが重要であり、開発者のアルゴリズムが反映された結果だと考えられる。

5.2 独自の歌声合成ソフトウェア: v.Connect-STAND

Vocaloid や UTAU のように歌詞と譜面に基づいて歌声を合成する独自のソフトウェアとして、v.Connect-STAND^{*3}が開発されている。v.Connect-STAND は、Cadencii と組み合わせることで利用する歌声合成エンジンである。Cadencii はエディタであるが、合成エンジンを組み合わせることで Vocaloid と同様に歌声合成を実現できる。v.Connect-STAND は元々 UTAU 用にライブラリ化された音源に基づいて Cadencii をエディタとして用いる合成エンジンであるが、独自のルールでライブラリ化した音源を用いることで、さらに高い品質を達成するための試みがなされている。詳細については本研究会の次の発表を参考にしたい。

6. 考 察

UTAU の合成エンジンは、標準搭載されている resampler のほかに、fresamp, model4, TIPS, bkh01 等が公開されている。WORLD を用いたものうち公開されているのは EFB-GW と world4utau, v.Connect(-STAND) であるが、v.Connect-STAND は UTAU の枠組みを超え、独自の歌声合成システムへ発展しつつある。UTAU 用の合成エンジンは使用者が自由に選択でき、ライブラリ化された音源によって合成エンジンによる相性が確認されることから、合成エンジンを用途に応じて使い分ける自由度が生まれたといえる。

最新の v.Connect-STAND については、独自のライブラリを構築することが必要条件となるが、歌声の強弱をモーフィングにより補間するなど、歌声の品質だけではなく表現力も向上しつつある。爆発的な普及には至っていないが、歌声合成ソフトウェアを発展させるために僅かながら貢献できたと考えられる。音声を対象とした WORLD の基本的な性質は STRAIGHT や TANDEM-STRAIGHT 以上であることは明らかであるが、ピッチや音色

*1 ソフトウェアの名前は Twitter で公募した結果から採用したものである

*2 個別のリンクは出来ないため、各合成エンジンの呼称をニコニコ動画で検索して頂きたい。

*3 ニコニコ大百科 <http://dic.nicovideo.jp/a/v.connect-stand>

の複雑な制御に対する頑健性は現在明らかにされていない。投稿された動画においても品質劣化に対する指摘が少なからずあったことから、歌声合成ソフトウェアとして実施される加工に対する頑健性の評価や頑健性の向上については、今後 WORLD を歌声合成用に特化させるために必要な検討となる。

7. おわりに

本稿では、音声分析合成システム WORLD を Web を通じて配布することによる協調的創造活動の支援を試みた結果を報告した。提案法である WORLD は、特許や利用の制限無く使える技術であり、ソースコードが配布されているため誰でも歌声合成システムを開発することが利点といえる。WORLD を用いた歌声合成ソフトウェア UTAU に対する合成エンジン用のプラグインとして、筆者が実装した EFB-GW だけではなく、UTAU の開発者が実装した world4utau、Zteer 氏が開発した合成エンジンなど、複数の合成エンジンが開発された。さらに、UTAU とは異なる歌声合成ソフトウェアである v.Connect-STAND が開発されている。それらの歌声合成エンジンや歌声合成ソフトウェアによりコンテンツが制作され、動画共有サイトに投稿されはじめている。

今後は、動画共有サイトで得られたコメントからも WORLD の品質改善、速度向上などの改良を目指すことが必要になる。これらの改良についても、協調的創造活動により実現されることを期待したい。

謝辞 UTAU 用の合成エンジンの開発、WORLD により合成された歌声を用いたコンテンツは、多数の開発者、クリエイタにより作られたものである。本稿では、許可を頂いたもののみ開発者の名前を添えて紹介させて頂いたが、特にコンテンツに関してはここでは紹介しきれない量が制作されている。全ての開発者、クリエイタの皆様には、謹んでお礼申し上げます。

参 考 文 献

- 1) 佐々木 渉：仮想楽器をリアルにする「未来 (ミク) の記号」と、VOCALOID で注目される「人の形」「声の形」について、情報処理学会研究報告, Vol. 2008, No. 50, pp. 57-60 (2008).
- 2) Fischer, G.: Symmetry of ignorance, social creativity, and meta-design, Knowledge-Based Systems Journal, Vol. 13, No. 7-8, pp. 527-537 (2000).
- 3) 濱崎雅弘, 武田英明, 西村拓一: 動画共有サイトにおける大規模協調的な創造活動の分析, 情報処理学会研究報告, 2009-MUS-80 (2009).

- 4) 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, Vol. 2008, No. 12, pp. 51-58 (2008).
- 5) Kawahara, H., Masuda-Katsuse, I., and Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, Speech Communication, Vol. 27, No. 3-4, pp. 187-207 (1999).
- 6) Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno H.: TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation, Proc. ICASSP 2008, pp. 3933-3936 (2008).
- 7) Dudley, H.: Remaking Speech, J. Acoust. Soc. Am., Vol. 11, No. 2, pp. 169-177 (1939).
- 8) 森勢将雅, 西浦敬信, 河原英紀: 高品質音声分析変換合成システム WORLD の提案と基礎的評価 ~ 基本周波数・スペクトル包絡制御が品質の知覚に与える影響 ~ , 日本音響学会聴覚研究会, Vol. 41, No. 7, pp. 555-560, (2011).
- 9) Hess, W.: Pitch determination of speech signals, Springer-Verlag, Berlin, (1983).
- 10) Cheveigné, A. and Kawahara, H.: YIN, a fundamental frequency estimator for speech and music, J. Acoust. Soc. Am., Vol. 111, No. 4, pp. 1917-1930 (2002).
- 11) Camacho, A. and Harris, J. G.: A sawtooth waveform inspired pitch estimator for speech and music, J. Acoust. Soc. Am., Vol. 124, No.3, pp.1638-1652 (2008).
- 12) 森勢将雅, 河原英紀, 西浦敬信: 基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法, 電子情報通信学会 論文誌 D, Vol. J93-D, No. 2, pp. 109-117 (2010).
- 13) B.S. Atal, S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Am., vol.50, no.2B, pp.637-655, 1971.
- 14) Mathews, M.V., Miller, J.E., and David, E.E.: Pitch synchronous analysis of voiced sounds, J. Acoust. Soc. Am., Vol. 33, No. 2, pp.179-185 (1961).
- 15) 森勢将雅, 松原貴司, 中野皓太, 西浦敬信: 高品質音声合成を目的とした母音の高速スペクトル包絡推定法, 電子情報通信学会 論文誌 D, Vol. J94-D, No. 7, pp. 1079-1087 (2011).
- 16) 森勢将雅, 松原貴司, 中野皓太, 西浦敬信: 高品質音声合成を目的とした励起信号抽出法の検討, 日本音響学会 2011 年春季研究発表会講演論文集, pp. 325-326 (2011).
- 17) Hamon, C., Moulines, E., and Charpentier, F.: A diphone synthesis system based on time-domain prosodic modifications of speech, Proc. ICASSP89, Vol. 1, pp.238-241, (1989).
- 18) ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems," 2003.