



平成11年12月21日

日本電信電話株式会社

究極の音声合成技術「FinalFluet」を開発 —肉声と遜色のない“女声”合成音声を実現—

NTTは、任意の漢字かな混じりテキスト文を高品質な合成音声に変換する音声合成技術「FinalFluet」を開発しました。「FinalFluet」は、音声合成の単位として、「複合音韻連鎖単位（マルチフォームユニット）」を採用することにより、“男声”に比べ音声合成が難しいとされてきた“女声”の合成を、肉声と遜色のない自然さで実現しました。また、合成音声の高低、イントネーション、感情的音声等も簡単に制御できます。

「FinalFluet」は、需要が増大している各種情報案内サービスでの利用、あるいは電子メールやホームページの読み上げ等のインターネット上での利用、各種マルチメディアコンテンツ作成への応用等、幅広い用途が見込まれます。今後、NTTではさらに自然で表現豊かなイントネーションの生成に取り組む予定です。

<開発の背景>

自動バンキングサービスや音楽速報など、合成音声による情報案内サービスが実用化されています。しかし、案内文等のテキスト情報を合成音声にする場合、従来の音声合成では子音-母音連鎖をベースに細分化された音声をつなぎ合わせて合成するため、案内文全体の合成音声の自然性に問題がありました。そのため、テキスト情報から自然な音声を合成する技術の実現が待ち望まれていました。

<主な特徴>

- 1) 肉声と遜色のない自然で明瞭な音質
FinalFluetでは、「自然性」および「明瞭性」に重点をおき、発声した

女性の音声と遜色の無い合成音声を実現しました。多くの情報案内サービスは、定型的なテキストと頻繁に更新されるテキストから構成されます。このような場合に定型文を録音音声、更新テキストを音声合成という組み合わせで行いますが、従来の合成音声では肉声との品質差が大きく、結果として非常に印象の悪いサービスとなりがちでした。

FinalFluetで道路交通情報を提供する場合、定型的文章は肉声であらかじめ収録しておき、「～キロメートルの渋滞」などの表現や道路名・交差点名といった、更新される情報部分を音声合成し、これらを組み合わせ、自然で、かつ運用コストのかからない音声情報案内サービスを実現できます。

2) 声質・発声速度・声の高さ・抑揚などを自由に設定可能

利用するサービスの目的に応じて、声の高さ・抑揚・声質などを自由に設定できます。例えば、あるサービスでは声の高さ・声質を高め、抑揚を強く設定することにより、歯切れの良い明るい感じの音声にでき、逆にこれらの設定を押さえ気味にする事により、落ち着いた感じの音声にすることが可能です。

3) 特定の語彙のイントネーションを事前に登録可能

FinalFluetでは、特定の語彙・フレーズについて、予めイントネーションなどを変えて登録することができます。例えば、特定地方における情報案内サービス用途として、その地方の言いまわしや、フレーズ（例：「おおきに」）に方言イントネーションをもたせ、親近感を抱かせた地域密着型情報案内サービス提供が可能となります。あるいは情報案内サービスの中で、キャッチフレーズ的に利用されるテキストを特に強調した言いまわしとするなどが可能となります。

<技術のポイント>

1) 音声合成単位に独自の「複合音韻連鎖単位（マルチフォームユニット）」を採用

合成音声は、音声合成単位(*1)の組み合わせとしてつくられますが、NTTでは、FinalFluetの音声合成単位に、「複合音韻連鎖単位（マルチフォームユニット）」を採用しました(図)。

従来の音声合成単位の多くでは、子音-母音単位で結合点が発生します。このため、不連続が生じやすくなり、人工的で不自然な合成音声となっていました。FinalFluetでは、音声合成単位の結合点を従来に比較して少なくする「複合音韻連鎖単位」により、自然な合成音声を実現しました。「複合音韻連鎖単位」では、音声合成単位を2つの条件に基づいて

決定します。

- <1>母音の連続はできる限り一つの合成単位として持ち、接続は原則として母音から子音に移る箇所とする。
- <2>同じ音韻の系列においてもイントネーションの異なる複数の合成単位を用意する。

NTTでは、この2つの条件を前提に、日本語データベースから必要にして十分と考えられる約6万の音声合成単位を作成しました。この結果、結合点を従来と比較して少なくすることができ、さらに同じ音でも文頭の声の上がる調子の部分と、文末の声の下がる部分など、音声の高低変化による違いが考慮されており、肉声に近い合成音声出力が可能になりました。

2) 音声デザインツールとして独自開発の「Sesign99」を搭載

FinalFluetは音声の制御、編集機能に、音声デザインツール「Sesign99」を搭載しています。Sesign99は、GUI環境のもと、各種パラメータをマウスで操作するだけで簡単に合成音声の大きさ、声質、ピッチなどを変更できるツールです。また、一度作成したユーザ独自のイントネーションをライブラリ化することもできるため、誰でも手軽に高度な音声編集が可能です。

<用語解説>

* 1 : 音声合成単位

音声合成の元となる音声素片で、合成単位を結合する事によって連続音声を生成する。

- ・ [図：FinalFluetと従来型音声合成技術との合成単位接続比較](#)

本件に関するお問い合わせ先

NTTサイバーコミュニケーション総合研究所
情報戦略担当 萩野・坂本

TEL : 0468-59-2032

e-mail : ckoho@tamail.rdc.ntt.co.jp



[NTT NEWS RELEASE](#)