

音声合成技術を用いたアバターへのアイデンティティ投影

高道 慎之介* 樋口 啓太† 中村 哲*

概要. ノベルゲームやアバターを使用するゲームの成功の理由の一つは、ゲームユーザのアイデンティティをゲーム世界に投影できる点である。この種のゲームでは、ユーザの見た目や動きをゲーム世界に反映することで、ゲーム世界へのユーザの没入を補助している。本研究では、ユーザの音声情報の反映によりユーザの没入感を加速させることで、ゲームへのモチベーションの維持を図る。具体的には、アバター等のゲームキャラクタの話す音声に対して、ユーザの声色・話し方を付与することで、ユーザのアイデンティティを投影する。本ツールでは、ゲーム開始前にユーザの音声を入力し、ユーザの音声を合成する音声合成システムを構築することで、ゲームシナリオに含まれる任意のテキスト・セリフの読み上げが可能となる。本報告では、音声合成技術をゲームに導入するにあたり、検討すべき課題にアプローチするツールについて述べる。

1 はじめに

アバターとは人間がゲームや仮想空間を体験する際のインターフェースとなる媒介である。ニンテンドーMiiのように、アバターの外観を実在する人間に近づけ、ゲームに取り入れるための仕組みは、現在広く受け入れられている。ゲームへの没入感・満足感を高めるためのアバター生成・表現技術が積極的に開発されている。代表的なものとしては、人間の表情とアバターの表情を同期するためのコンピュータグラフィックス技術や、人間が自らのアバターを生成するためのインターフェース技術 [1] がある。しかし、それらの技術はアバターの外観を似せるものに限られている。

本研究では、ゲームなどで使用するアバターへのアイデンティティ投影を深めるために、ユーザが事前発話した音声を利用して、ユーザの音声を合成する音声合成システムを構築する。本システムの使用時には、任意のシナリオ・セリフを音声合成システムに入力することで音声を合成する。音声合成技術の導入にあたり、a)ユーザの発話への頑健性、b)没入感を加速させる音声の合成 c)音声合成のゲームエンジンへの組み込みを実現するツールを検討する。

2 音声合成技術

音声合成 (TTS: Text-To-Speech) とは、任意のテ

キストから音声を合成する技術である。現在の主流であるコーパスベース音声合成方式は、テキストを読み上げた音声を用いて、データドリブ的に音声合成システムを学習する。学習後は、学習に用いた話者の声質で音声を合成可能となる。コーパスベース方式の中でも、隠れマルコフモデル (HMM) に基づく音声合成方式 [2] は、少量の音声データから合成器を構築できるため、多分野への応用が期待されている。長年の課題であった合成音声の音質劣化問題も、多くの手法により緩和 [3] されつつあり、また、高品質かつ産業的利用の容易な音声分析合成器の登場 [4] により、応用範囲は更に広がると思われる。

3 音声合成技術のゲームへの応用

ユーザの音声を合成する方法について述べる。フロー図を Fig. 1 に示す。まずゲーム開始前に、ユーザは指定された文を発話し、システムに入力する。発話された音声は音声特徴量に分解され、HMM に基づく音声合成システムの学習に使用される。使用時には、ユーザの所望するゲームに含まれる任意のテキスト・セリフを音声合成システムに入力することで、ユーザは、自身の音声情報が付与されたゲームをプレイできる。

音声合成技術をゲームに利用するにあたり、a)ユーザの発話への頑健性、b)没入感を加速させる音声合成について検討する必要がある。

合成音声の高品質化のため、一般的に利用される音声合成システムは、静音環境において収録され、

Copyright is held by the author(s).

*奈良先端科学技術大学院大学, †東京大学

かつ、話者情報・発話様式等が十分にアノテートされた音声データを用いて構築される。一方で本研究では、ゲームユーザによる多様な発話様式・収録環境が予想される。故に、これらの要素に対して頑健な構築法が必要である。

また、ユーザの没入感を加速させるために、どの音声情報（例えば、声色の個性、音質、話し方、感情）が有効的であるか検証する必要がある。また、セリフに応じた表現法（例えば、感情・強調）の制御法も検討する。ゲームデータに感情情報が含まれる場合、回帰モデル [5] 等による制御が考えられる。一方、[6] の様に、他者（例えば、他ユーザやプロの声優）の発話又は合成した音声に含まれる表現を、当該ユーザの合成音声に反映する方法が考えられる。

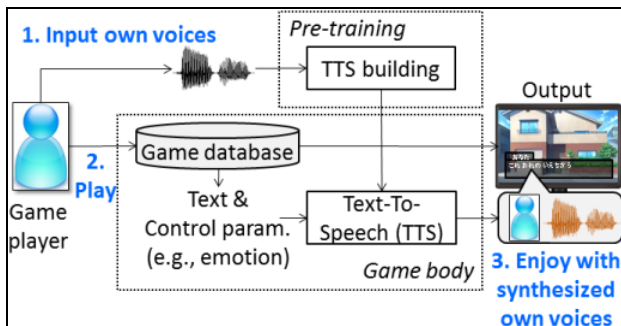


Fig. 1 任意のテキスト・セリフからゲームユーザの声を合成し、ゲームに利用するシステム

4 アプリケーション

本システムはノベルゲームやロールプレイングゲームなど、様々な種類のゲームに取り入れることが可能である。ゲーム開発者は事前に会話のテキストと、テキストに付随する感情情報を登録しておくことにより、ユーザにより生成された音声合成器に任意の言葉を喋らせることが可能となる。また、ユーザ間で音声合成器を共有することにより、登場人物に自分以外の実在する人間のアイデンティティを付与することができる。と考える。

5 今後の展開

本稿では、ゲームなどで使用するアバターへのアイデンティティ投影を深めるために、ユーザが事前ユーザの声を合成する音声合成システムを構築する方法を述べ、検討すべき課題及びアプリケーションを述べた。

今後は実際のゲームエンジンに組み込み可能な音声合成生成・再生システムを開発する。UnityやUnreal Engineなどに搭載可能なプラグインとして開発することにより、より多くのゲーム開発者に本研究の成果を提供する。また、実際にゲームプレイヤーやデザイナーからのフィードバックを通して、必要な性能や機能などを明らかにしていきたい。

参考文献

- [1] Y. Zhang, T. Han, Z. Ren, N. Umetani, X. Tong, Y. Liu, T. Shiratori, and X. Cao. BodyAvatar: creating freeform 3D avatars using first-person body gestures. 2013. Proc. of the 26th annual ACM symposium on User interface software and technology (UIST), 2013.
- [2] K. Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi, and K. Oura. Speech synthesis based on hidden Markov models. Proceedings of the IEEE, vol. 101, no. 5, pp. 1234–1252, 2013.
- [3] S. Takamichi, K. Kobayashi, K. Tanaka, T. Toda, and S. Nakamura, The NAIST Text-to-Speech System for the Blizzard Challenge 2015. Proc. of Blizzard Challenge, 2015.
- [4] 高道 慎之介, 戸田 智基, 森勢 将雅, 中村 哲. HMM 音声合成における音声分析合成器 STRAIGHT と WORLD の比較, 日本音響学会秋季音響研究会, 1-Q-27, 2015.
- [5] T. Nose, J. Yamagishi, T. Masuko and T. Kobayashi. A style control technique for HMM-based expressive speech synthesis. IEICE Trans., Inf. And Syst., vol. E90-D, no. 9, pp.1406-1413, 2007.
- [6] Y. Nishigaki, S. Takamichi, T. Toda, G. Neubig, S. Sakti and S. Nakamura. Prosody-Controllable HMM-based Speech Synthesis Using Speech Input. Proc. MLSLP, 2015.