

実写動画コンテンツを用いた音声対話システムの検討

國部 渚†, 元村 隼†, 庄司 直紘†, 安部 恵一†

神奈川工科大学 創造工学部 ホームエレクトロニクス開発学科†

1. はじめに

近年, AI(Artificial Intelligence)技術の発展により, 音声対話システムが普及している. 宅内で使用する音声対話システムとして, AI スピーカーやヒューマロイドタイプがある. しかし, 利用者によっては外観が気にいらなかったりした場合, キャラクターを大幅に変更することはできないため利用者によって会話が弾まなくなる課題がある. この課題を解決するため, 我々は過去に CG(Computer Graphics)キャラクターを用いた音声対話型ホームシステムの提案 [1][2]を行ってきた. 我々の提案した音声対話システムはリップ動作の CG アニメと待機時 CG アニメの2つの GIF アニメを用意し, 会話に合わせて切り替えることで CG キャラクターのリップ動作を行う手法である. また CG キャラクターは FaceRig などのアバターキャラクターを使用しているため開発が容易になる.

従って, 本研究ではこの過去の CG 音声対話システム [1]の原理を利用して, CG キャラクターを実写の動画コンテンツに変えることで, 現実では会話が難しい人物, 例えば亡くなってしまった人や憧れのスターなどと自然対話が行える音声対話システムの実現を試みたので, その詳細について述べる.

2. 関連研究

近年, AI 技術及びCGなど駆使して, 亡くなった人 [3]や有名スター [4]を蘇らせる技術がある. 例えば, 有名スターの美空ひばりの復活プロジェクト [4]や, 韓国のテレビ番組で紹介された亡くなった娘と一度会いたい母親のために AI, CGそしてVR(Virtual Reality)の技術を駆使して再現された技術がある [3]. これらの既存技術は製作期間及び製作費にコストがかかるため容易に作成できない課題がある. そこで, 本研究では従来の既存技術と比べて低コストで現実では会話が難しい人物, 例えば亡くなった人や憧れのスターなどと自然対話が行える音声対話システム技術について提案する.

Study of Voice Dialogue System using Live-Action Video Contents.

†Nagisa Kokubu, †Jun Motomura, †Naohiro Shouji, and †Keiichi Abe.

†Department of Home Appliance Engineering, Faculty of Creative Engineering, Kanagawa Institute of Technology.

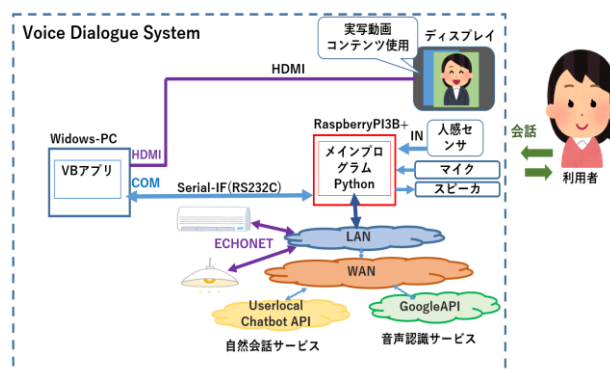


Fig.1 我々が提案する VDS-LVC の概要 [1]

3. 提案システム

3.1 提案システムの概要

我々が提案するシステムは, 我々が過去に開発してきた CG キャラクターを用いた音声対話システム [1]の原理を利用して, CG キャラクターを実写の動画コンテンツのキャラクターに置き換えることで行う. 本稿では実写動画コンテンツを用いた音声対話システム (Voice Dialogue System using Live-Action Video Contents)を VDS-LVC と呼ぶ.

我々が提案する VDS-LVC のシステム概要を Fig.1 に示す. VDS-LVC では Fig.1 に示すように Windows-PC と Raspberry PI3B+(以下 RPI と呼ぶ)で分散処理を行う. RPI では自然対話機能(音声入力, 自然対話, 音声合成)及び人感センサ検出, ECHONET-Lite による宅内の各種家電の遠隔操作を行う処理を担当する. 一方, Windows-PC 側では利用者と対話を行う実写人物の表示切替を担当する. PC 側に実写人物の動画として, 待機時と会話時の2つを用意し, この動画を会話に合わせて切り替えることで自然なリップ動作を実現させる. RPI 側の処理プログラムは Python2.7 で開発した. 利用者との音声の入出力はマイクとスピーカーで行っている. 本システムでは自然対話を実現するため, 音声認識に Google Speech API サービスを, 自然対話に User-local 社の Chat Bot API を, 音声合成は Open JTalk を使用した. Windows-PC と RPI とのデータ通信は RS232C のシリアル通信で行う. RPI が本プログラムのメイン処理を担当し, 必要なタイミングで RPI から Windows-PC へ任意のリクエストを送信することで実写人物の表示を切

り替えるシステムとした。この処理は Visual Basic2019 で開発した。

3.2 実写動画コンテンツの作成

我々が提案する実写動画コンテンツの作成は Fig. 3 に示すように Web カメラで、人物の実写動画として待機時動画と会話時動画の 2 つを作成する。自然な表示を行うためそれぞれ 10 秒以上の長さの動画で、かつ撮影時の背景は何もない背景の方がよい。また、本提案手法は生前に上記のような実写動画コンテンツを GIF ファイルで作成しておく必要がある。

本提案手法では、元となる動画の用意ができれば実写人物を作成するのは容易である。作業に慣れていれば約 5 分で作成することが可能であるため、既存技術と比較すると、低コストで実写動画コンテンツを作成できる利点がある。

3.3 音声コンテンツの作成

今回我々は実写人物に似た音声をフリーの Open JTalk[5] で調整したものを使用したが、あまり再現度は高くないと感じた。他の手法として、商用となるが、自分の声に似た音声合成サービス[6]を行うものが存在する。

3.4 生活情報の提供サービス[1]

本システムの生活情報サービスは、過去の我々のシステム[1]と同様、検索サービス、お天気情報、時刻などがある。また、宅内のネットワークに接続されている ECHONET-LITE 対応家電(エアコン、照明など)を利用者の音声を介して遠隔操作を行えるサービスを実装した。

3.5 音声対話の選択画面アプリケーション

Fig. 4 に今回開発した実写人物選択画面アプリケーションを示す。画面上から利用者の好みで実写キャラクターをマウスで選択すると、Fig. 4 に示すように画面が大きくなり、実写人物のキャラクターと自然会話ができるようにした。

4. 結論

本稿では従来の既存技術と比べて、低コストで現実では会話が難しい人物(例えば亡くなった人や憧れのスターなど)と自然対話が行える音声対話システム技術について提案した。今後は、作成した実写人物の会話がどれだけ違和感なくできるかなどアンケートなどを通じて本提案システムの評価を詳細に行っていく予定である。また、提案手法は亡くなった人と会話する場合、生前に実写動画コンテンツを作成しておかなければならない課題がある。また、亡くなった遺族の方がこのようなシステムを利用する場合の心のケアについても倫理的・宗教的観点からも総合的に検討していく必要があると考えられる。

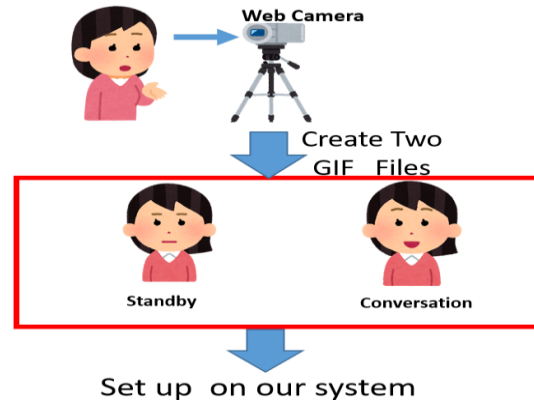


Fig.3 実写動画コンテンツの作成



Fig.4 実写人物キャラクターを選択できるアプリケーション画面

<参考文献>

- [1] DAISUKE HIRUMA, KEIICHI ABE *et al*, "Proposal of Interactive Home System Using Computer Graphics (CG) Characters ", IEEE International Conference On Consumer Electronics 2019(ICCE2019), pp.1-4, Jan. 2020.
- [2] Naohiro Shoji, Takayo Namba, Keiichi Abe, "Proposal of Spoken Interactive Home Doctor System for Elderly People", 2020 IEEE 9th Global Conference on Consumer Electronics Proceedings (GCCE2020), pp. 587-589, Oct. 2020.
- [3] “亡き娘にもう一度会いたい死者に VR で再会する番組の波紋 (Newsweek)”, <https://www.newsweek-japan.jp/stories/woman/2020/07/vr-1_2.php>, (入手日 2020 年 12 月)。
- [4] “AI 美空ひばり”の紅白新曲「あれから」が音楽業界で注目の理由”, <<https://www.businessinsider.jp/post-205093>>, (入手日 2020 年 12 月)。
- [5] “OpenJ Talk”, <<http://open-jtalk.sp.nitech.ac.jp/>>, (入手日 2020 年 12 月)。
- [6] “コエステーション”, <<https://coestation.jp/>>, (入手日 2020 年 12 月)。