Vol.2012-4
2013/2/13

キャラクターを使用した 自動リップシンクシステムの開発

物部寛太郎[†]　高橋淳也[†]　高橋まどか[†]

近年，「ゆるキャラ」が注目を集めており、多数のゆるキャラが各地に存在し，地元地域だけにとどまらない人気を持つゆるキャラもいる．そのようなゆるキャラであるが，喋らせることで，さらに魅力を高めることができる．キャラクターを喋らせるためには，リップシンクをさせることが有効である．しかし，個人が作成したキャラクターに対してリップシンクをさせることは，困難である．そこで，本研究ではその簡易版として，画像と音声を用意するだけで，誰もが簡単にリップシンクを実現できることを目標とし，自動リップシンク動画生成システムを開発する．

Development of an automatic lip-synch system for a character

Kantaro Monobe[†]　Junya Takahashi[†]
and Madoka Takahashi[†]

In recent years, "yuru-kyara" attracts attention, and much yuru-kyara exists in many places, and yuru-kyara also has a popularity which does not remain only in a local area. Such yuru-kyara can heighten charm further by talking. In order to make character speak, it is effective to carry out lip-synch. However, it is difficult to carry out lip-synch to the character which the individual created. So, in this research, as the simple version, We develop a system which can realize lip-synch easily if there are a picture and a sound.

1. はじめに

近年，「ゆるキャラ」が注目を集めている．ゆるキャラとは「ゆるいマスコットキャラクター」を略したものであり，主に地域のアピールのために使用されているマスコットである．現在，多数のゆるキャラが各地に存在しており，地元地域だけにとどまらない人気を持つゆるキャラもいる．その経済効果は大きく，2011 年にゆるキャラグランプリにてグランプリを獲得した熊本県の「くまモン」は，関連グッズの売り上げだけで 1 年に 25 億円であると言われている[1]． なぜここまでゆるキャラが流行しているのかと言えば，キャラクターの持つ力というものが大きい．愛らしい造形のものには愛着が湧きやすいものである．さらに，ゆるキャラは，基本的にその地域の特徴を模しているため，キャラクターを通して郷土愛も生まれる．キャラクターを前面に押し出すことによって，様々な世代に親しみやすくなる．

そのようなゆるキャラであるが，喋らせることで，さらに魅力を高めることができる．キャラクターを喋らせるためには，リップシンクをさせることが有効である．リップシンクとは，口の動きと音声が合致していることを指す．主に舞台における歌，吹替映像，アニメーション等において使用されている技術である．喋っている内容に合わせて口を動かすことで, キャラクターが喋っていることを実感することができる．

その例として, ディズニーパークのアトラクション「タートル・トーク」があるが，ディズニーのキャラクターとリアルタイムで会話ができるという内容である. これは，裏で従業員が喋っており，その声に合わせて，CG でロの動きを実現する仕組みになっている. また, 人物の写真を使用してリップシンクをさせる研究[2]も行われている．

しかし，個人が作成したキャラクターに対してリップシンクをさせることは，困難である．そこで，本研究ではその簡易版として，画像と音声を用意するだけで，誰もが簡単にリップシンクを実現できることを目標とし，自動リップシンク動画生成システムを開発する．

2. 研究の概要

本研究では，利用者が用意した画像と音声で，自動的にリップシンクを行うシステムの開発を行う．リップシンクは，喋っているときに，口の形を「大きく開いた口」「半開きの口」「閉じた口」の 3 種類を代わる代わる表示させることで行う．これは，主にアニメーションで使用されている手法である．

[†] 宮城大学 事業構想学部 デザイン情報学科
Miyagi University, School of Project Design, Department of Design and Information Systems

出力の種類としては，
・利用者が用意した画像と音声で動画を出力
・利用者が用意した画像とマイクによる音声入力でリアルタイム映像として出力
の 2 つがある．本研究の概要を図 1 に示す．

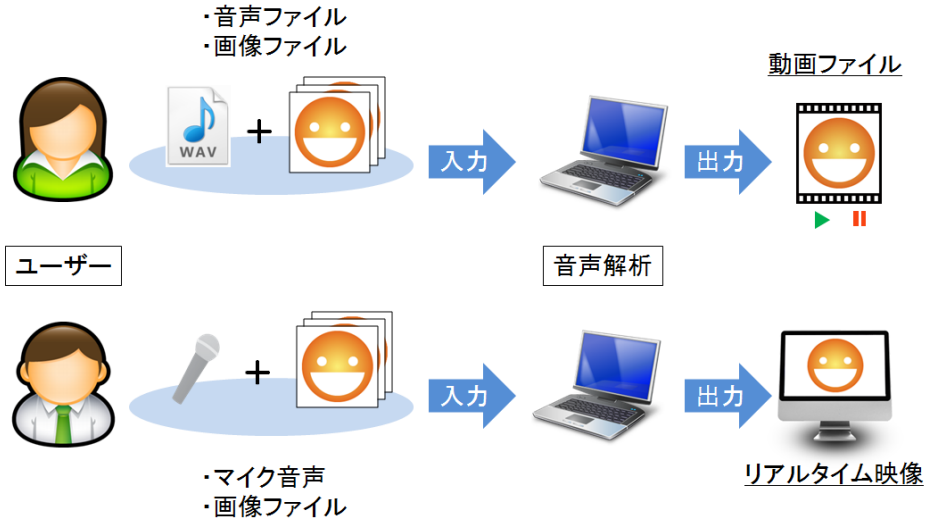

図 1　本研究の概要

3.　システムの概要

本システムは，1) 画像・音声ファイルの読み込み，2) 各種設定，3) 音声の解析，4) 出力の 4 つの処理により実現する．開発環境には Processing[3]を使用した．本システムのフローチャートを図 2 に示す．

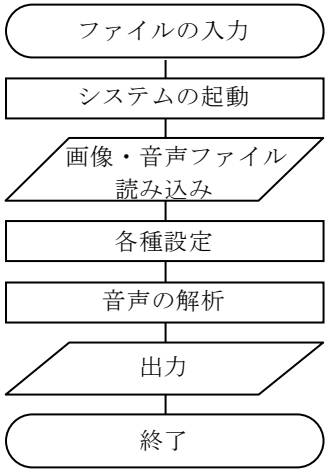

図 2　本システムのフローチャート

3.1　画像・音声ファイルの読み込み

使用する画像は，システム起動前に，指定のファイル名でフォルダに入れておく．画像は，キャラクターの口の部分のみを「大きく開いた口」「半開きの口」「閉じた口」に変更した 3 種類を用意する．用意する画像のイメージを図 3 に示す．

図 3　用意する画像のイメージ

これらの画像は，すべて同じサイズである必要がある．また，図 3 は正面の顔であるが，実際には斜め向きや，横向きの顔の画像も利用することができる．

システムを起動すると，フォルダに入れられた画像のサイズを認識し，その画像のサイズに合わせたウィンドウを自動で開く．

同じく，使用する音声も，事前に指定のファイル名でフォルダに入れておく．また，システムの起動後に音声ファイルを変更することが可能である．

3.2 各種設定

本システムでは，7 つの設定を行うことができる．以下で 7 つの設定について説明する．

3.2.1 モード切り替え

前述したように，本システムには，用意した音声ファイルからリップシンクを行うモードと，マイク入力からリップシンクを行うモードの 2 つがある．そのモードの切り替えを行うことができる．

3.2.2 音量閾値設定

本システムは，音声ファイルの音量を解析してリップシンクを行う．その解析により，キャラクターの口が開閉する音量閾値の変更を行う．小さい音でも口を動かしたい場合には閾値を下げ，大きな音にのみ反応させたい場合には閾値を上げる．この設定により，音声ファイルに雑音が入っている場合でも，音声にのみ反応させることができる．音量閾値の設定イメージを図 4 に示す．

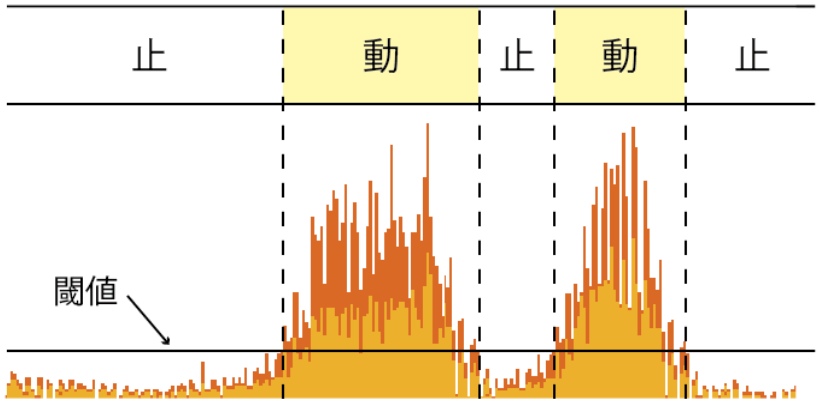

図 4　音量閾値の設定イメージ

3.2.3 音声ファイル選択

ファイル選択ダイアログを開き，使用する音声ファイルを選択することができる．

3.2.4 動画出力開始

事前に用意した音声と画像によるリップシンク動画の出力を開始する．

3.2.5 動画出力停止

事前に用意した音声と画像によるリップシンク動画の出力を停止させる．

3.2.6 音声再生

事前に用意した音声ファイルの再生を開始する．

3.2.7 音声停止

音声ファイルの再生を中止する．同時に巻き戻しを行い，再び音声再生をした際に先頭から再生されるようにする．

3.3 音声の解析

入力された音声ファイルまたはマイクの音量を解析する．その際，設定された音量の閾値に従って，閾値を下回っている場合には「閉じた口」を表示し，閾値を超えた場合にはキャラクターの画像を切り替える．切り替えの法則は，以下の通りである．

本システムでは，1 秒 15 フレームの動画を生成する．その中で，リップシンクを自然に表現する手法 [4]に従って，画像を次の順番で表示する．

「大きく開いた口」3 フレーム　→
「半開きの口」3 フレーム　→
「大きく開いた口」2 フレーム　→
「閉じた口」2 フレーム　→　先頭に戻る

このルールに従った口の動きを図 5 に示す．

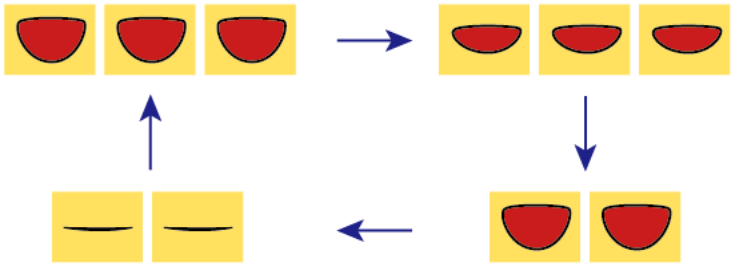

図 5　口の動き

3.4 出力

音声ファイルを使用するモードの場合，動画ファイルを QuickTime (.mov) 形式で出力し，マイク入力を扱うモードの場合，リアルタイムで映像を出力する．出力したイメージを図 6 に示す．

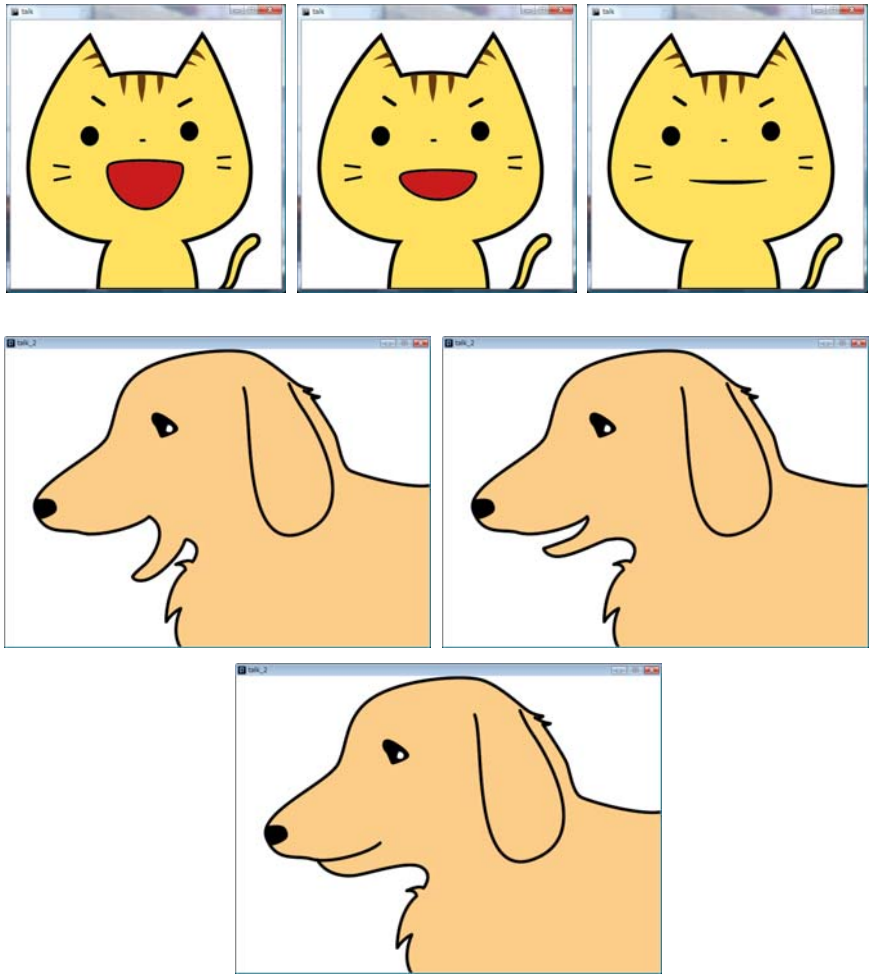

図 6　出力イメージ

4. おわりに

本研究では，画像と音声を用意するだけで，誰もが簡単にリップシンクを実現できるシステムを開発した．
本システムの利用例として，地域のウェブページ上で，ゆるキャラに喋ってもらうことを提案する．地域の PR をする際に，単に文章を掲載するよりも，ゆるキャラが話すことによって，より親しみが湧くと考える．また，近年ではソーシャル・ネットワーキング・サービスによる，顔を知らない者同士の交流が増加している．そこで，本システムによるリアルタイムの音声でリップシンクを実現する機能によって，自分の代わりにキャラクターを喋らせることができる．

今回は，キャラクター画像を使うことを前提とした研究であるため，アニメーションで用いられている技法を参考にリップシンク動画作成を行った．しかし，よりリアルな画像を使用する際には，より高度な音声認識を行うことで，母音を判断して，「あ」「い」「う」「え」「お」「ん」の口の形を再現する動画を作成することが必要であると考える．

参考文献

1) 日本経済新聞：ゆるキャラグランプリの「バリィさん」，愛媛凱旋 (2012). http://www.nikkei.com/article/DGXNZO48840880W2A121C1LA0000/
2) 小林隆夫， 益子貴史：音声およびテキストからの音声同期唇アニメーションの自動生成，平成 10 年度「研究報告」放送文化基金，No.23，pp.45-49 (2002).
3) Casey Reas：Processing をはじめよう，オライリージャパン (2011).
4) A.e.Suck：FLASH アニメーション制作バイブル，オーム社 (2007).