

Generating Intermediate Face between a Learner and a Teacher in Learning Second Language with Shadowing

Yoko Nakanishi*, Yasuto Nakanishi[†]

概要. シャドウイングとは聞いた音を即座に発音する外国語の学習方法である。しかし特に学習の初級者の場合、認知的負荷が非常に高い。そこで我々は、学習者と教師となるネイティブ話者の顔から画像処理により "中間顔" を生成し、顔の表情の動きを分かり易く提示することで、シャドウイングにおける発話の認知的負荷を下げることに取り組む。本稿ではプロトタイプシステムの実装について述べる。

1 Introduction

Following the teacher's movements is an important technique for those who want to learn dance, sports, and language. Studies show differences between the movements of the teacher and those of the learner can teach how to move each body part. While learning a language, learners need to grasp the differences between sounds of native speakers and their own[3]. Shadowing is a language-learning technique whereby a learner attempts to repeat - to "shadow" - what he/she hears immediately.

In addition, note that shadowing face and mouth movements is important for learning a language. Akiyama pointed out that the Japanese are inexperienced at horizontal control of the lips because of their characteristic pronunciation habits[2]. Moreover, Nonaka encouraged the Japanese to be aware that the use of the abdominal muscle, lungs, throat, tongue, lips, mouth, and face are all different when speaking Japanese and English[1].

However, as far as we know, a method integrating sound shadowing and physical movement shadowing has not been proposed. In our research, we suggest a language-learning method incorporating both sound shadowing, and face-and mouth-movement shadowing. In this paper, we describe our prototype system, which enables this new type of shadowing. It generates intermediate faces from 3D meshes and textures, captured with a real-time camera input and captured movie.

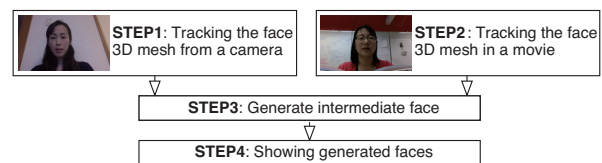


figure 1. The flow of our system.

2 Implementation

Our system comprises a camera and PC. The camera captures the image of the learner sitting in front of it, and the PC recognizes the position of the learner's face without facial markers. Our system recognizes the learner's face from a camera, and the teacher's face, as he/she speaks English, from a movie. Then, it generates two 3D meshes and two texture images and shows two intermediate faces between the learner and the teacher (figure 1).

STEP 1: Tracking a face from a camera

Our system finds the learner's face in the real-time camera input image using openFrameworks add-ons. It recognizes the learner's face as a 3D mesh that includes points of facial features such as eyes, nose, and mouth. The tracked 3D mesh data are sent to STEP 3 in each frame.

STEP 2: Tracking a face in a movie

In this step, our system finds the teacher's face in a movie in the same manner as in STEP1. This step provides 3D mesh data of the teacher's face, and sends the data to STEP 3 in each frame.

STEP 3: Generating intermediate faces

The system generates an intermediate face with the still image of the learner's face and the tracked teacher's 3D mesh (intermediate face A), and another intermediate face with the still image of the teacher's face and the tracked learner's 3D mesh

Copyright is held by the author(s).

* Independent Researcher

[†] Keio Univ., Faculty of Environment and Information Studies

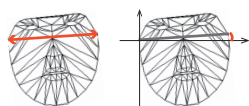


figure 2. Left shows width in the 3D mesh data, right shows tilt in the 3D mesh data.

(intermediate face B). The facial still images are taken beforehand and used as a texture image. At first, the system calculates the width and tilt angle of each face to apply an affine transformation to the 3D mesh data (figure 2). To generate intermediate face A, the system transforms the teacher's 3D mesh data with an affine transformation matrix based on the learner's face width and tilt angle and applies the learner's still image face as a texture of the transformed 3D mesh. To generate intermediate face B, the system utilizes the learner's 3D mesh data, an affine transformation matrix, and the still image of the teacher's face in the same manner. Then, each intermediate face is blurred by image processing to assimilate with the background image.

STEP 4: Showing generated faces

Finally, the learner performs their shadowing with the aid of the audio and generated intermediate faces. In the current implementation, the learner can select the following three modes to show the intermediate faces. The first one shows himself/herself, intermediate face A, and the teacher (figure 3a). The second one shows himself/herself, intermediate face B, and the teacher (figure 3b). The third one shows himself/herself, intermediate face A, intermediate face B, and the teacher (figure 3c).

3 Discussion and Future work

Videos are sometimes used as shadowing teaching material because they can stimulate learner interest more than audios. We prototyped a system to integrate sound shadowing, and face and mouth movement shadowing, using videos and image processing.

It shows faces generated from the learner and the teacher. This would make it easier for the learner to see how different his/her face and mouth movements are from the teacher's compared with when they only watched a video. Most learners who perform shadowing find it cognitively difficult to hear and repeat speech with the correct rhythm and speed, even though all learners



figure 3. The face of the learner in the camera input is located on the left side in all images, whereas the teacher's movie image is located on the right side in all images. a) The center image is intermediate face A. b) The center image is intermediate face B. c) The top center image is intermediate face A, and the bottom center image is intermediate face B.

choose content according to their own English level because it is up to the learner to decide how to reconfigure the sounds[4][5]. This means that appropriate scaffolds bring about more effective shadowing. Showing intermediate faces can work as additional scaffolds because it shows the differences between a learner and a teacher not only with auditory sensations but also with visual sensations. To examine the potential of our approach, we will conduct user studies with the following hypothesis; 1) Learners understand that face and mouth movement shadowing is important for learning languages. 2) Learners understand the differences in face and mouth movements using intermediate faces. 3) Showing intermediate faces with an appropriate layout results in effective shadowing.

REFERENCES

- [1] 野中泉 . 英語舌のつくり方, 研究社, 2005.
- [2] 秋山善三郎. 口輪筋・頬筋, その他顔面表情筋の働きと英語音声. 音声学会会報, 208, pp29-36, 1995.
- [3] Luo, D., e. a. Speech analysis for automatic evaluation of shadowing. In Proc. SLATE (2010).
- [4] 大木俊英, 他. シャドーイング開始期における学習者の復唱ストラテジーの分類. 関東甲信越英語教育学会誌, Vol.25, pp. 33-43, 2011.
- [5] 白木智士, 他. シャドーイングトレーニングによる英語学習意欲向上, コミュニケーション研究叢書 6, pp. 25-36, 2008.