EVALUATION OF A MULTIMODAL 3-D PRONUNCIATION TUTOR FOR LEARNING MANDARIN AS A SECOND LANGUAGE: AN EYE-TRACKING STUDY

Ying Zhou[1,2], *Fei Chen*[2], *Hui Chen*[3], *Lan Wang*[2], *Nan Yan*[2]

[1] School of Information Engineering, Wuhan University of Technology, Wuhan, China

[2] Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[3] Institute of Software, Chinese Academy of Sciences, Beijing, China

Introduction

- Many 3-D pronunciation tutors with both internal and external articulator movements have been implemented and applied to computer-aided language learning (CALL).
- In our previous study, a multimodal 3-D articulation system has been presented that incorporated an airflow model into the 3-D articulatory model to produce the airflow in accordance with articulator movements of Mandarin pronunciation.
- Data of participants' eye-movement behavior showed a correlation between the duration of eye fixations and the depth of learning.
- In the present study, by using eye-tracking methodology, evaluation of the multimodal 3-D Mandarin pronunciation tutor was conducted in comparison with real human teacher to penetrate whether the 3-D talking head can work well during the process of language learning.

- Through analyzing eye-tracking measures, three research questions (RQ) were put forward and investigated:

RQ1: Comparatively, which is more attractive to learners, HF or 3-D?

RQ2: Under different presentation conditions, do learners pay different attentions to the areas of interest (AOI, see *Data Analyses* for more details)? Which is directly related to language learning?

RQ3: During the process of learning, will learners focus on airflow information which is useful to increase the syllable identification accuracy by illustrating the airflow differences between the confusable consonants?

Methods

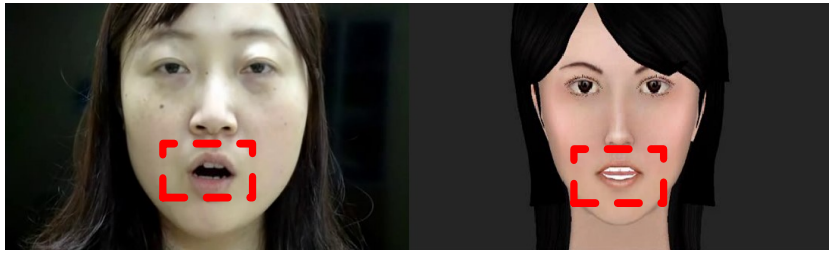
❖ Participants, materials and procedure

◆ **Fourteen** non-Chinese speakers (10 males) between 20 and 31 years of age.

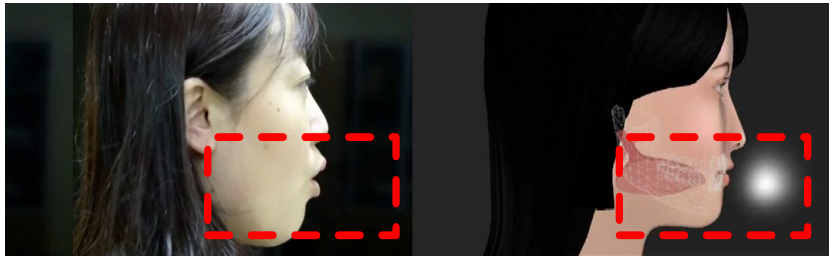
◆ Eye movement data of the participants was collected non-intrusively by **RED 5 Eye Tracker** (SMI Technology, Germany), which was integrated within the panels of the monitor.

◆ The 16 Mandarin syllables were presented under two conditions of **HF** and **3-D**, totaling 32 videos, each with a front view first and then with a corresponding profile view.

(a) Front View



(b) Profile View



▸ *The chosen area of interest (AOI) with front and profile view corresponds to the area inside the dashed red rectangle.*

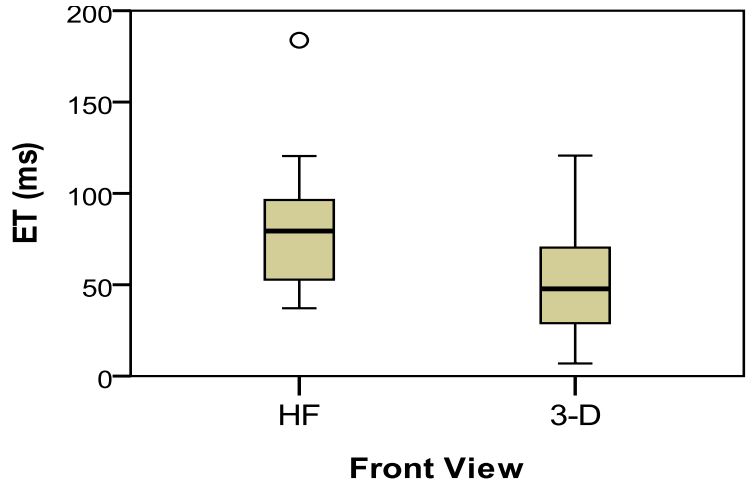
1) 'entry time' (ET), which means the duration from start of the trial to the first hit of one special area. The shorter the ET, the more interest for the area was shown.

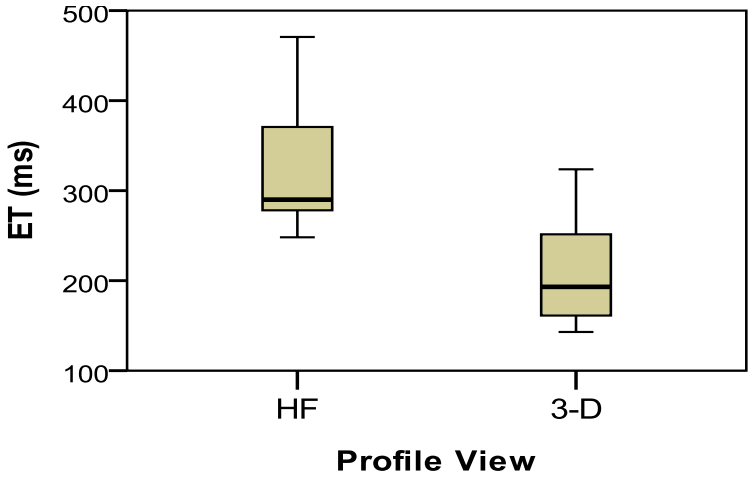
2) 'fixation count' (FC), which is defined as the total number of fixation lasting more than 100 ms inside the area of interest (AOI), reflecting the absolute attention during the process of learning.

3) 'proportion of fixation duration' (POFD), indicating the ratio of fixation duration inside the AOI to the duration of whole video screen, reflecting the relative attention during learning.

Results

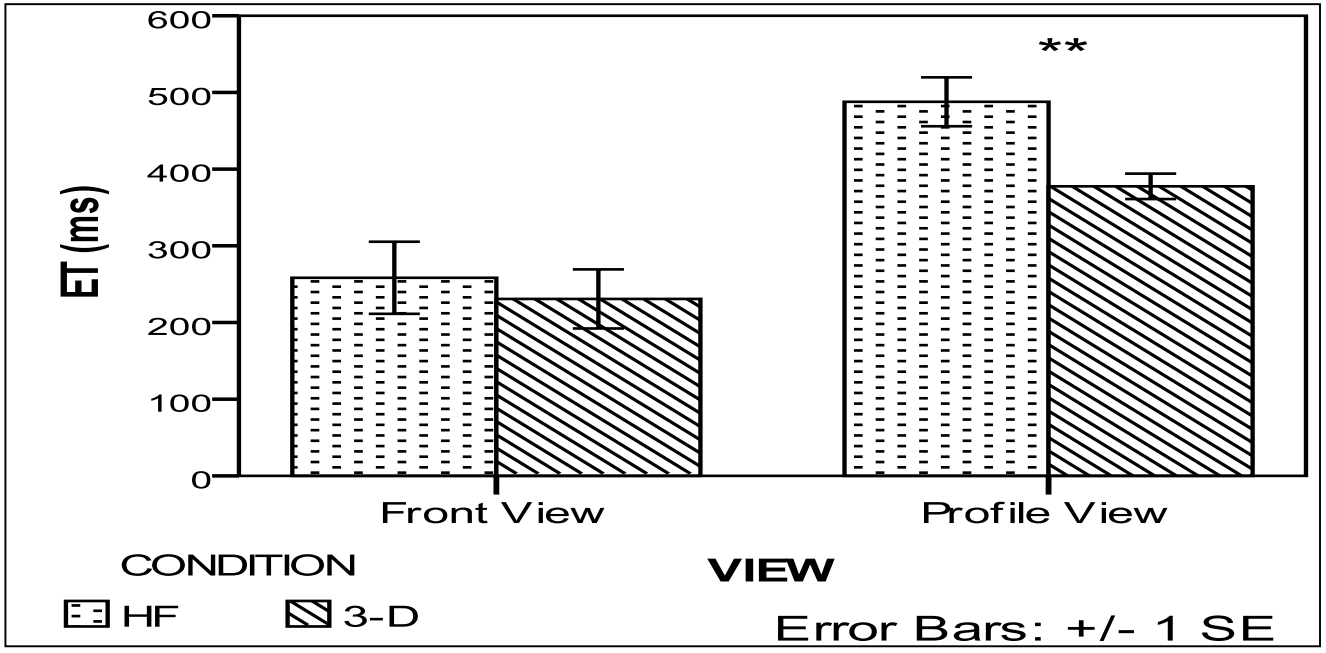
- **Comparison of entry time (ET) into HF and 3-D videos**
  - ET into 3-D condition was significantly shorter than that into HF condition with the front view [F (1, 26) = 4.687; $p < 0.05$], and with the profile view [F (1, 26) = 24.147; $p < 0.001$].
  - The similar results went to the the average ET into different presentation conditions, which could be observed more visually from figure as follows..



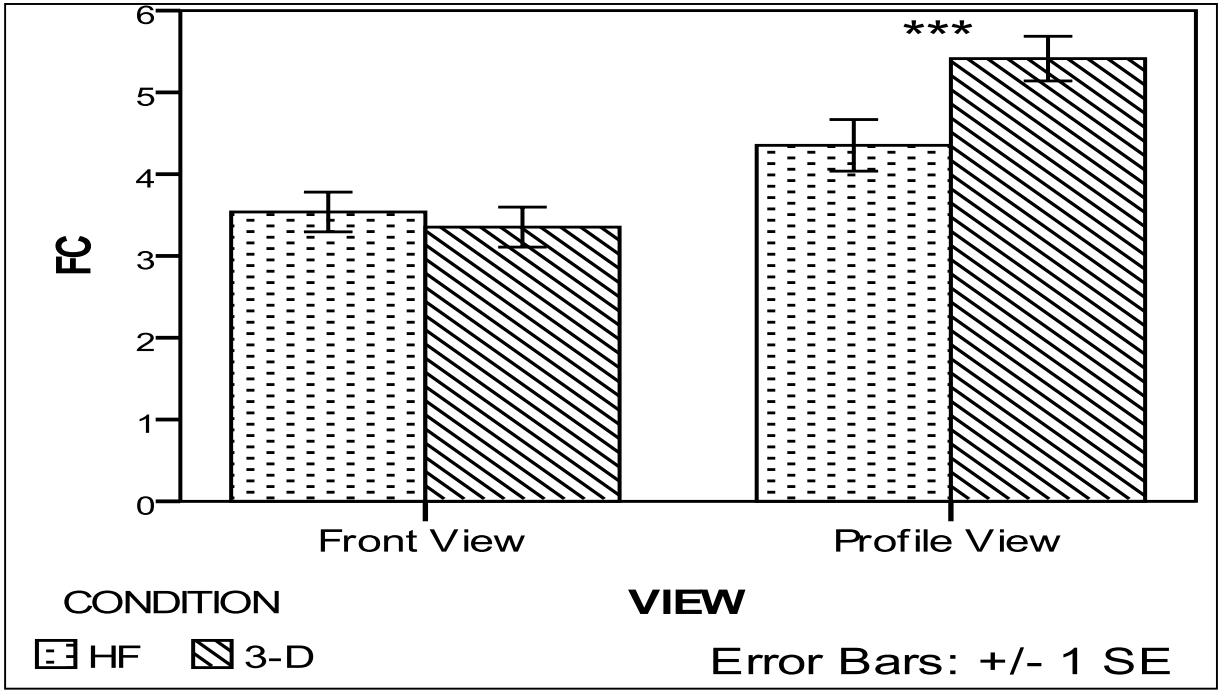


Eye-tracking measures of AOI

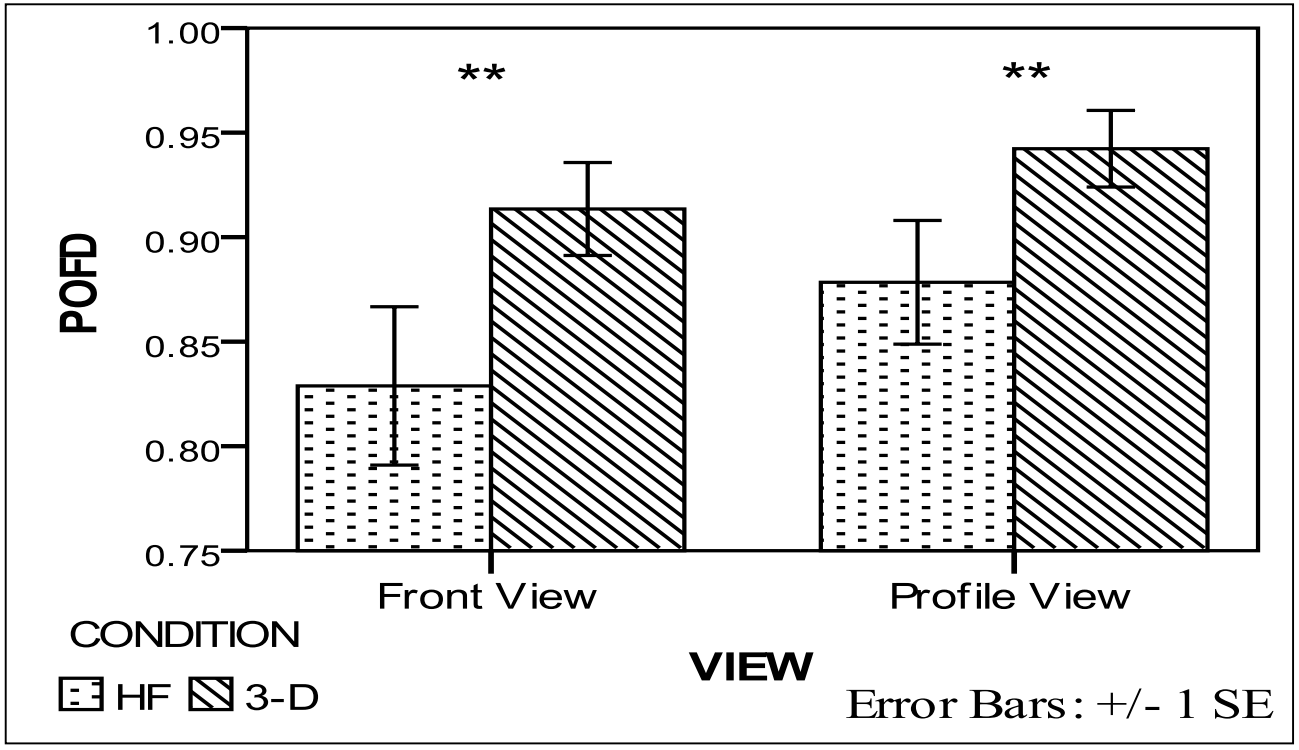
- Post hoc analysis indicated that with the front view, presentation condition showed no significant effect ($p = 0.229$). With the profile view, however, ET was much shorter in 3-D (377.58 ms) than in HF (487.80 ms) condition ($p < 0.01$)



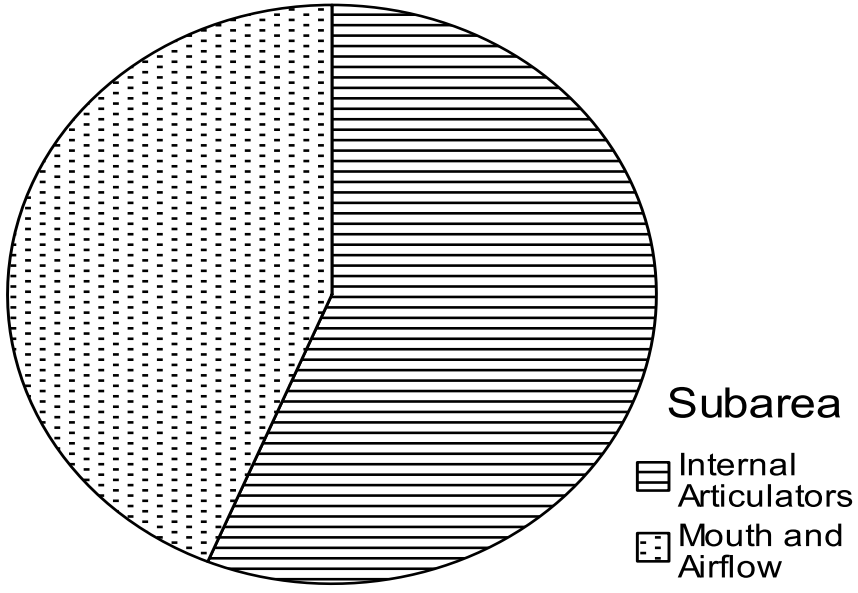
- With the front view, presentation condition showed no significant effect (p = 0.073), while with the profile view, the FC of AOI in 3-D (5.41) was higher (p < 0.001) compared with that in HF condition (4.35) (see Table 1 and Fig. 4)



- Post hoc analysis indicated that the POFD of AOI in 3-D was significantly higher than that in HF condition both with front view and profile view (all *p*s < 0.01)



- Distribution of attention inside the AOI of profile 3-D video
  - The average FC between the two subareas (see Figure) was not different from one another ($t = 1.916$, $p = 0.066$).



Discussion

- Results indeed showed that ET into 3-D videos was much shorter than the HF videos with both the front and profile view, indicating that non-Chinese learners showed more interests in our 3-D Mandarin pronunciation tutor.
- The results showed that learners put more relative attention (i.e., higher POFD) to the lip movement in the 3-D tutors. Moreover, the results of three eye-tracking measures of AOI (i.e., ET, FC, and POFD) all showed that 3-D pronunciation tutor with a transparent profile view triumphed over the real human teacher by effectively delivering articulator movement and airflow information (see Heat map).
- The results of our current study proved that, while watching the 3-D pronunciation tutor, the allocation of absolute attention was equably distributed among the entire AOI with the transparent profile view, rather than concentrating merely on internal articulators

Conclusions

- With a shorter entry time into the 3-D videos, non-Chinese learners showed more preference for our 3-D Mandarin pronunciation tutor.
- Learners observed the lip movement of 3-D tutor for a relatively longer time with a front view.
- With the transparent profile view, our multimodal 3-D pronunciation tutor exhibited greater advantage of delivering articulator movements and airflow information.
- Learning from the 3-D tutor in a transparent profile view, learners kept focusing on the entire AOI (including the internal articulators, mouth and airflow) which is important for Mandarin articulation learning.

The end

Thank you!