MIRRORING BEAT GESTURES: EFFECTS ON EFL LEARNERS

Noriko Yamane[1], Masahiro Shinya[1], Brian Teaman[2], Marina Ogawa[1], and Soushi Akahoshi[1]

[1]Hiroshima University, [2]Osaka Jogakuin University
yamanen@hiroshima-u.ac.jp, mshinya@hiroshima-u.ac.jp, teaman@wilmina.ac.jp,
b152151@hiroshima-u.ac.jp, b154851@hiroshima-u.ac.jp

ABSTRACT

It has been shown that apexes of manual gestures such as deictic gestures are often temporally coordinated with the intonation peak [3, 7], and beat gestures help planning utterances and rhythm of prosodic constituents in English [10, 9]. However, what is not known is how mirroring someone's beat gestures would affect vocalization of L2 English speech. Japanese college students participated in speech training sessions and a comparison was made between an experimental group mimicking beat gestures of a TED Talk speaker and a comparison group shadowing audio of the same speaker. The speech sounds and gestures (Kinect One, Microsoft) were measured in pre- and post-training sessions. The experimental group tended to produce a more target-like prosody, suggesting that gesticulation facilitates an English-like rhythm. The temporal relation between the stressed syllables and the gestural apexes will also be discussed.

Keywords: Kinect, beat gesture, pitch range, error.

1. INTRODUCTION

Recently a growing body of experimental evidence suggests that intonation peaks and apexes of deictic gesture are often temporally coordinated in highly controlled focus conditions [3, 7]. These findings suggest exploring the utilization of gestures in order to deliver messages effectively. In EFL contexts in Japan, where the use of English is extremely limited, an urgent need is to give them an experiential learning to mimic a Ted Talk speaker, while enhancing students' pronunciation and oral presentation skills. Mirroring a speaker's "total expressive system" [1] can be a breakthrough for L2 English learners, as gestures enhance perceptual and production prominence [6], and a study of mirroring a model speaker of Ted Talk [12] reports it helps pitch range expansion. Frequently used in academic contexts as well as public speeches are beat gestures, which help planning utterances and rhythm of prosodic constituents in English [10, 9]. Inspired by this previous research, we explore questions such as: i) Do beat gestures help L2 learners enhance the difference between stressed and unstressed syllables? ii) How are EFL learners' beats aligned with stressed syllables? This short paper reports the procedures of the experiment, methods, and the preliminary results.

2. PROCEDURE

2.1. Participants

An English class for second-year students at a Japanese university was used for this study, and thirty-three students signed the consent form. They were divided into Group 1 (n=11, 7 males and 4 females) and Group 2 (n=12, 4 males and 8 females). Fifteen students' data had to be excluded from the research due to non-participation in either the pretest, posttest or pre/post-questionnaires which were part of this study. Therefore the total number of analysable data was sixteen; eight students (5 males and 3 females) who received mirroring training, and eight students (4 males and 4 females) who received shadowing training.

2.2. Oral Reading Script

The material of the main oral reading test was adapted from a Ted talk [11].

> I used to think the whole purpose of life was pursuing happiness. […] Eventually, I decided to go to graduate school for positive psychology to learn what truly makes people happy. But what I discovered there changed my life. […] And according to the research, what predicts this despair is not a lack of happiness. It's a lack of something else, a lack of having meaning in life. (194 words in total)

We prepared three kinds of reading materials:

(i) A4 paper (both sides, 2 pages), typed in Arial, 12pt, Bold, with a Japanese translation
(ii) 7 pages of PowerPoint slides, Arial, 28pt, Bold, English only
(iii) 9 pages of PowerPoint slides, Arial, 28pt, Bold, English only

Item (i) was used for the preparation session before the pretest. To facilitate the understanding of the content, a Japanese translation was added below each

sentence. Items (ii) and (iii) were used for pretest and posttest respectively. In order to facilitate the readability/view of the texts, the texts were divided into several chunks consisting of a few sentences, which were laid out on each slide. For (iii), a few novel sentences were added at the end of the text (ii), in order to examine look at how they behaved on novel unrehearsed speech.

2.3. Instructions

An experiment was performed using a pretest/posttest design. Students belonged to either the Mirroring group or Shadowing group.

2.3.1. Pretest

In the pretest, they were handed an A4 paper, item (i) above, given 5 minutes to get themselves familiar with the text, and told that they are going to read the paragraph as if they were making a speech. In this short preparation session they were allowed to read it aloud, and ask an assistant for the pronunciation of unknown vocabulary.

Each person was guided to a classroom one by one. They were handed an ear-hook microphone (Andoer B011C6FTK6) and asked to wear it on their right ear. Its cord was secured to a part of their shirt with a clip to keep the microphone from dropping or moving suddenly. They were asked to stand behind a yellow tape line on the floor that was prepared in advance. In front, participants see a projector screen which shows the script of the speech from item (i) that they already saw in the preparatory session. They were asked to imagine they were going to make a speech in front of an audience of about 20 people. They were told that their hands and body can move in a natural way while speaking, but not to hold their hands behind their body.

In the classroom, one of the assistants was in charge of the PowerPoint slide (ii), and hit a keyboard to advance the slides according to the speech rate of each participant. Another assistant was in charge of recording audio and motion. Another researcher was in charge of controlling an ipad to record their video. While the participant was being recorded, the latter two people were out of sight of the participant to minimize the participant's stress of being monitored.

2.3.2. Training Session

The training sessions continued over 3 weeks, with one session per week. In order to get students engaged in practice for a short period, the story was divided into 3 sections in advance. Each section (consisting of a few sentences) was presented to them each week.

In each sentence, students were given an explanation about the following points:

- Location of stressed vs. unstressed syllables, pauses and phrasing, focus words, and its relation to the loudness, duration and pitch
- Pronunciation of challenging words the students may not know (e.g., "gnawing")
- Phonological change due to linking, reduction, and deletion

This was done to raise their phonological awareness, and to help them to feel ready to read it fluently. They were also told to underline the location of all stressed syllables.

The students were divided into two groups based on their position on the student roster. The first half of the students on the roster were moved to a different classroom, and given a shadowing practice session. A .wav file of the script, which was extracted from the same Ted Talk, was played through a classroom audio speaker. First, they were asked to listen carefully to the whole audio. Second, an assistant stopped the audio at the end of each sentence which they were to repeat audibly. Finally, they were told to shadow the audio as closely as possible, and continue to do it.

The second half of the students remained in the first classroom, and were given a mirroring session. First, the video of the Ted Talk was played and the students were asked to watch the speaker's facial expression and manual gestures in the relevant part carefully. They were encouraged to annotate the speaker's movement on the handout, and at this stage they came to be aware that the location of the beat gestures - a rapid up and down movement of the hands - often matches with stressed syllables. Second, they were asked to mimic beat gestures without vocalizing, while watching the video. Third, an instructor paused the video at the end of each phrase/sentence, and students were asked to read the script and match the beat strokes with stressed syllables. Finally they were told to read the script aloud and move their hands as they watched and continue to do it without any pausing of the video. Both session consisted of about 20 minutes.

2.3.3. Posttest

Four weeks later, participants took the posttest using the same procedure as in the pretest. They were shown some new material following the familiar script, but asked to read it as well as they could:

> But that raised some questions for me. Is there more to life than being happy? And what's the difference between being happy and having meaning in life? Many psychologists define happiness as a state of comfort and ease, feeling good in the moment. Meaning, though, is deeper. (47 words of continuation of the pretest)

3. METHOD

3.1. Acoustic data

The pitch of the recorded speech was measured using Praat [2]. The pitch unit was set to "semitones re 1Hz" to normalize individual differences. The measurement was done for the following sentences for comparison:

- **Pretest & posttest**: Eventually, I decided to go to graduate school for positive psychology to learn what truly makes people happy.
- **Extended part**: Many psychologists define happiness as a state of comfort and ease, feeling good in the moment.

The sentence "Eventually…" was chosen because the speaker used the most beat gestures for this than any other sentence. The extended part was chosen as above, because among sentences in the extended passage, that has a relatively similar number of syllables as the test sentence.

Our acoustic analysis focuses on 7 students (4 males and 3 females) from the mirroring group, and 7 students (3 males and 4 females) from the shadowing group, who participated in all three training sessions. (Data M16 from shadowing group had noise issues in the audio file, and data M13 from mirroring group did not follow the instructions while recording, so these two files were removed.)

3.2. Kinematic data

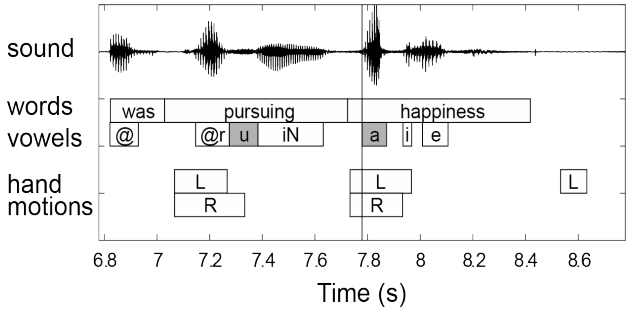
The participants' kinematics were measured by using Kinect One for Windows (Microsoft), which is a marker-less motion capture solution that estimates the location of 25 anatomical landmarks including the head and both hands. The Kinect sensor was placed on the horizontal surface of a 0.45 meter high desk 3 meters in front of the participants.

Kinect measured the kinematics at approximately 30 Hz but sometimes it slowed down because of a small jitter associated with Windows OS and other software installed in the computer. The recorded time series of the kinematics were re-sampled in 30 Hz and smoothed using a zero-lag Butterworth digital low-pass filter (4th order, 10 Hz of cutoff frequency). The square of the right and left wrist was calculated. Then, the time periods when the square velocities exceeded 10% of the maximum value were defined as the periods when the left and right hands are in motion (Figure 1).

3.3. Speech-gesture data

The participants were asked to clap their hands after the recording started but before they started speaking to create a point of time to be utilized for synchronizing the sound and kinematic recordings. The maximum intensity in waveforms around the clap was identified in the audio tier, and the midpoint of zero motion while both hands met was identified in the motion tier. These specific starting points were aligned in Camtasia, a video-editing application. Using Matlab, all the acoustic data and kinematic data were time-aligned, and arranged vertically (Figure 1). The example shows a part of the sentence "I used to think the whole purpose of life was pursuing happiness" uttered by one participant (M11).

Figure 1: The example of acoustic and kinematic data from a participant in the mirroring group.



From the top, the tiers are (1) waveform, (2) words, (3) vowels, (4) left hand motion, and (5) right hand motion. The numbers on the bottom show timesteps in seconds. In tier (3), stressed vowels and unstressed vowels are coloured differently. (In the vowel tier, @ denotes [ə], and iN denotes [ɪŋ]. For typographical reasons, some IPA symbols were replaced by other symbols.) Boundaries reflect the segmentation of vowels. If it was hard to detect the boundaries between a vowel and a consonant (e.g, *-ing*), the adjacent consonant was left attached to the vowel.

The kinematic tiers ("L" and "R" boxes in Figure 1) should generally include the *preparation* phase and the *stroke* phase of beat gestures. Beat gestures adopted here can be considered as multiple events – (i) the movement of one or two hands toward a higher position (preparation phase), (ii) a static phase at the highest position (pre-stroke hold), (iii) a quick downward motion (stroke phase), (iv) a post-stroke hold, and (v) retraction/recovery movement [4, 5].

Thus, one beat gesture should usually consist of two sets; a *preparation* phase occurs between 7.1 and 7.3, and the *stroke* phrase happens between 7.7 and 8. A static phase should be the cut-off point of one beat and the wrists should be held in a higher position between 7.3. and 7.7. So in this case, the learner's *preparation* phase starts before [@r] of "pursuing," and stretches to [u], while the *stroke* starts before [æ] of "happiness" and the motion of the left wrist stretches over to [i].

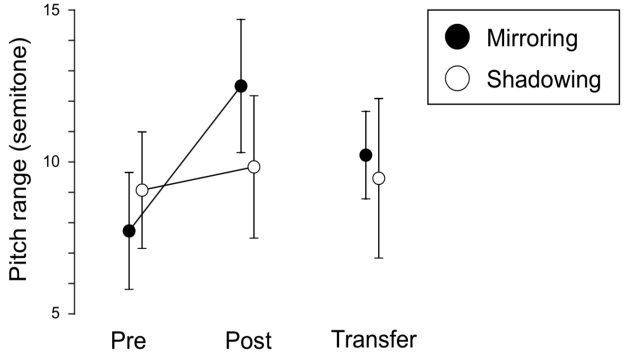
We examined where gestural time lag occurs, which could feed into error analyses of EFL learners of different proficiency levels. The gestural lag was defined as a case where the beginning of the stroke phase comes after the beginning point of a stressed vowel.

4. RESULTS

4.1. Acoustic samples: Pitch range expansion

A two-way repeated measures ANOVA (training: Pre/Post, and group: Mirroring/Shadowing) was performed to test the differences in pitch range data. A significant main effect of training was observed ($F(1, 12) = 15.3$, $p = 0.002$, $\eta 2 = 0.43$). Importantly, a significant interaction was observed ($F(1, 12) = 8.0$, $p = 0.015$, $\eta 2 = 0.23$), which indicates that the effect of Mirroring training on the pitch range was larger than that of Shadowing training.

Figure 2: Pitch range (semitone) of pretest and posttest, and transfer stage (to see whether participants were able to transfer their pitch expansion to new sentences.)



To test the generalizability of the training (i.e, skill transfer), another two-way repeated measures ANOVA (training: Pre/Transfer, and group: Mirroring/Shadowing) was performed. A limited generalization was revealed by the statistics ($F(1, 12) = 4.2$, $p = 0.063$, $\eta 2 = 0.17$).

4.2. Speech-gesture coupling: Case study of M11

Data of M11 was chosen because he made the greatest change in pitch range among the participants. Data from M11 shows that most of his beat *strokes* start and end before the stressed syllables, which is compatible with the previous findings [8]. However, some beat strokes start and end after stressed syllables, as given below.

- ... ever feeling ful**fill**ed, I felt anxious, and adrift.
  (Start lag: 0.1, End lag: 0.2)
  (stressed [ɪ]: 24.3–24.4, stroke: 24.4–24.6)
- … chasing happiness can make people un**ha**ppy.
  (Start lag: 0.1, End lag: 0.1)
  (stressed [æ]: 57.8–58, stroke: 57.9–58.1)
- … don't have to be clinically depressed to f**ee**l it.
  (Start lag: 0.4, End lag: 0.4)
  (stressed [i]: 97–97.2, stroke: 97.4–97.6)

These gestural lags are also noticeable by looking at the video. What is interesting is that such lags have not been reported in previous studies. Similar delays are observed in some other participants as well, which might be caused by the interference of their L1 Japanese. This point will require further investigation. Nonetheless, M11, who underwent mirroring training, expanded pitch range, and showed fewer pronunciation errors.

5. CONCLUSION

The results suggest mirroring helps L2 learners expand pitch range. In spite of time lags in the strokes, acoustic features crucial for speech intelligibility improved in the posttest. Further investigation should identify the effects of mirroring on other acoustic features, the duration of pause, its relation to the syntactic boundaries, types of errors, emotion, etc.

6. REFERENCES

[1] Acton, W. 1984. Changing fossilized pronunciation. Tesol Quarterly, 18(1), 71–85.

[2] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.40, http://www.praat.org/

[3] Esteve-Gibert, N., Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. Journal of Speech, Language, and Hearing Research, 56(3), 850-864.

[4] Kendon, A. 1972. Some relationships between body motion and speech. In: Seigman, A., Pope B. (eds.), Studies in dyadic communication. Oxford, England: Pergamon Press. 177–210.

[5] Kita, S. 1993. Language and thought interface: A study of spontaneous gestures and Japanese mimetics (Doctoral dissertation), University of Chicago.

[6] Krahmer, E., Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. Journal of Memory and Language, 57, 396–414.
[7] Krivokapić, J., Tiede, M. K., Tyrone, M. E. 2017. A kinematic study of prosodic structure in articulatory and manual gestures: results from a novel method of data collection. Laboratory phonology, 8(1). 1–26.
[8] Leonard, T., Cummins, F. 2011. The temporal relation between beat gestures and speech. Language and Cognitive Processes, 26(10), 1457–1471.
[9] Loehr, D. 2012. Temporal, structural and pragmatic synchrony between intonation and gesture. J. Lab. Phonol. 3, 71–89.
[10] Shattuck-Hufnagel, S., Ren, A. 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. Frontiers in psychology, 9. 1–13.
[11] Smith, E. 2017. There's more to life than being happy. TED Talk. https://www.ted.com/talks/ emily_esfahani _smith_there_s_more_to_life_than_being_happy
[12] Tarone, E., Meyers, C. 2018. The Mirroring Project. Speaking in a Second Language, 17, Amsterdam/ Philadelphia: John Benjamins. 197–223.