Written by: Meredith Stephens

Learning Pronunciation while Observing the Face

Several years ago, at the end of an undergraduate class, one of the students, Asana, asked me, "Can I sit at the front of the classroom every lesson, please? I have a hearing impairment."

I agreed and resolved to speak more loudly. Despite repeated efforts to project my voice, those at the back often asked me to speak up.

Several years later Asana enrolled in my graduate class. Because there were only two students in the class I was able to get to know her much better than when she had been in the larger undergraduate class. Even as an undergraduate, she had always stood out because of her crisp native-like enunciation. I took the opportunity to ask her about her hearing. It turned out that she had no hearing in her right ear and only 70% in her left.

"How did you learn to pronounce English so well?" I enquired.

"When I was in high school I used to enjoy listening to English language pop music," she informed me.

"What kind of music?"

"I particularly like the Beatles."

I was taken aback that someone so young could be a fan of the Beatles. My own children, approximately the same age as her, had never listened to them. Rather than learning English by listening to pop music I had anticipated that she would tell me that she had learnt to lip read in Japanese and had transferred this skill to English. I asked her about this.

"Yes," she confirmed. "When I was a child I learnt to read my parents' lips. One day I couldn't hear whether they were saying *udon* (noodles) or *budo* (grapes), but by looking at their lips I could see they were talking about *budo*."

I speculate that one of the reasons she articulated so clearly was because she was in the habit of supplementing her partial hearing with observing lip movement. Another reason for her clear enunciation could have been the year that she had spent on exchange in New Zealand during her first year of high school. However, I could not detect even a trace of a New Zealand accent. (Being Australian my ears are attuned to perceiving the differences between New Zealand and Australian English.) Her accent typified the middle ground occupied by English-speaking expats in Japan from around the globe, who accommodate each other as the differences in pronunciation between them may become somewhat attenuated. It sounded like Asana had acquired her accent from her various expat teachers, mainly American, at high school and university in Japan. What Asana may have gained from her year of living in New Zealand may not necessarily have been the accent, but the observation of lip movement every time she listened to someone or engaged in conversation.

My encounter with Asana gave meaning to some research I had done. In my role as classroom teacher, I had spent years advocating for bimodal input in the form of reading-while-listening, rather than mono-modal silent reading. I was mentored by my colleague in Tokyo, Anna Husson Isozaki. Anna regularly directed me to the latest research on reading-while-listening, and in 2017 alerted me to Dominic Cheetham's latest (2017) paper "Multi-modal language input: A learned superadditive effect." This paper provided a strong case for not merely bimodal but also multi-modal input. Cheetham decried the practice of monomodal input, and his paper pointed me to studies of the role of the observation of facial expression (van Wassenhove, 2013) and lip movement (Sekiyama & Burnham, 2008) on listening comprehension.

Van Wassenhove (2013) highlighted the key role of the observation of facial movements in order to understand spoken language. Furthermore, Sekiyama and Burnham (2008) explained that comprehension of spoken English is more reliant on visual cues than is Japanese, because of the former's greater phonetic complexity. English has fourteen or more vowels whereas Japanese has five, and English has the additional complexity of consonant clusters. Arguably, Japanese learners of English may need to devote more attention to lip movement than in their own language. Perhaps a major reason that Asana articulated so clearly was that she had had to devote more attention to observing lip movement than her peers.

Macedonia and Kepler (2013) suggested that mirror neurons may be responsible for making connections between the perception and reproduction of sounds: "Imitation via mirror neurons maps action into sound: the learner sees the motor act performed when the L2-speaker produces language sounds and through mirroring mechanisms trains and maps sound production into his own motor cortex" (p. 9).

They highlight the importance of having a motor representation of sounds: "If pronunciation training is only based on listening, learners store an acoustic pattern


that allows them to recognize the sound when they hear it. However, this is not the motor pattern needed for articulation of the sound. It does not provide the information the speaker would need on how to shape and sequence the airstream, tongue, teeth, lips, and so on. We must pose the question as to whether L2-learners can construct a motor representation of the sound without having seen how to do it?" (p. 9).

In my early days of teaching English in Japan, I used to regularly carry a CD player to the classroom. In my native-speaker ignorance I had not considered that simply having an acoustic representation of English pronunciation could be inadequate. In retrospect I can appreciate that the CD player does not demonstrate the airstream shaping or sequencing, or tongue, teeth, and lip movement described by Macedonia and Kepler (2013). Clearly, as Cheetham (2017) suggested, there is a need for bi- or multi-modal input. This could take the form of not just listening to an audio-recording, but simultaneously observing the airstream and lip movement.

The rationale of providing bi- or multi-modal input is not to help English learners sound like native speakers. Second language learners are unlikely to acquire the pronunciation skills of a native speaker after puberty (Kuhl, 2011). Nor should the aim of L2 English instruction be to foster native-speaker pronunciation, given the role of English as an International Language (Jenkins, 2000). There are more non-native speakers of English than native (Crystal, 2003), and therefore it is likely that English learners will speak English with other L2 English speakers.

The addition of visual input in terms of the observation of lip movement is not to foster the pronunciation skills of a native speaker, but rather to provide optimal opportunities for students to achieve a pronunciation that can be readily understood by a range of international speakers. As Macedonia and Kepler (2013) argued, L2 speakers benefit from observation of the motor act of the airstream and lip movement. English language educators need to be aware of this.
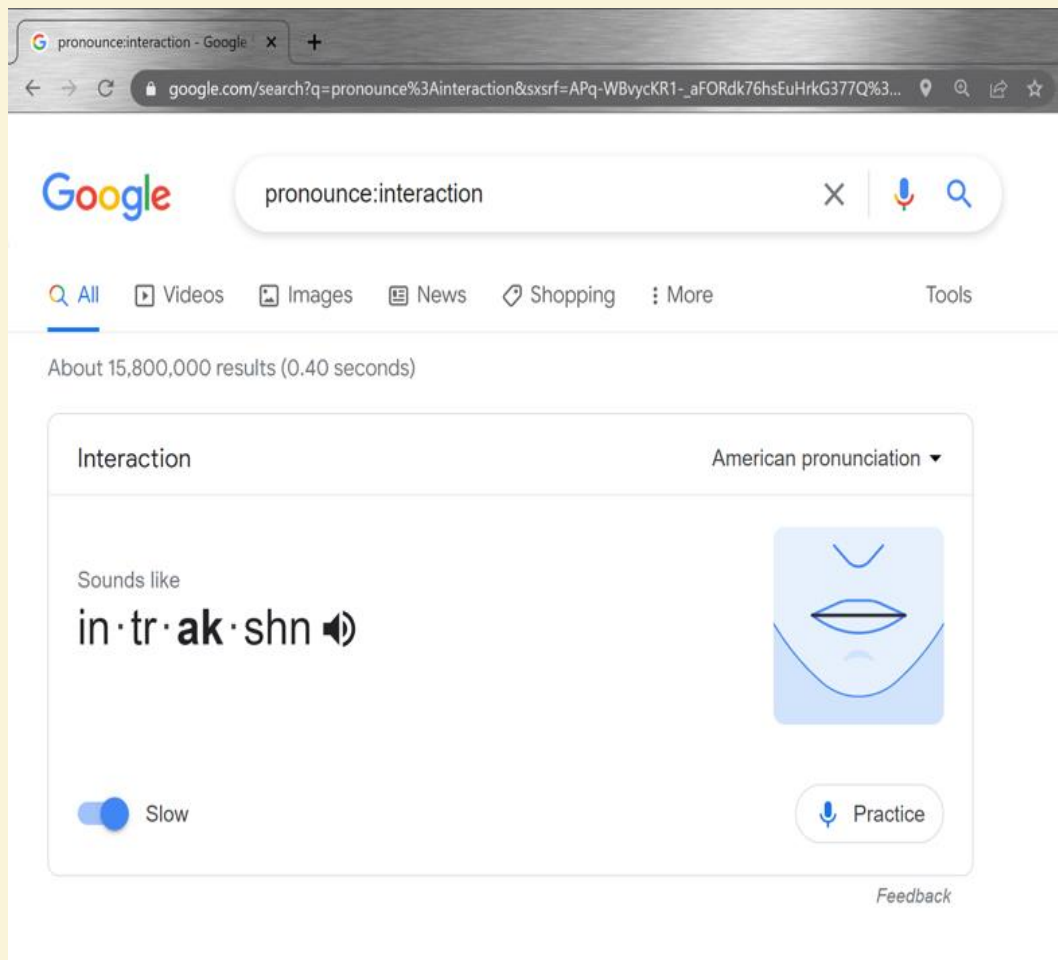
Rather than simply recommend learners listen to audio-recordings, we need to supplement this with observation of the speaker's face, or observation of lip movement on Google Pronounce. Asana taught me how learners with hearing impairments hone their English skills by including the visual factor. She has also opened up a door for all of us, both learners and teachers, that leads to a better way to learn pronunciation.

*What **Meredith Stephens** misses the most from her twenty years working at universities in Shikoku is teaching her seminar students, such as those featured in the accompanying sketch and photos.*

Note

Although this is beyond the scope of this paper, I will mention part of my discussion with Asana while I was revising it. Asana had been watching a [YouTube video](#) about Helen Keller learning to talk, and commented: "She felt the flow of air and even the vibration of the speaker's throat. By watching this, I was surprised that these elements can also be important modalities when learning a language." Learners with visual impairments too may be aided by the provision of input in alternative modalities.

Illustration by Takuma Sasakura

References

Cheetham, D. (2017). Multi-modal language input: A learned superadditive effect. *Applied Linguistics Review*. Advance online publication. https://doi.org/10.1515/applirev-2017-0036

Crystal, D. (2003). *English as a Global Language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511486999

Jenkins, J. (2000). *The phonology of English as an International Language*. Oxford University Press.

Kuhl, P. (2011). Early language learning and literacy: Neuroscience implications for education. *Mind Brain Education 5*(3): 128–142. https://doi.org/10.1111/j.1751-228X.2011.01121.x

Macedonia, M., & Kepler, J. (2013). Three good reasons why foreign language instructors need neuroscience. *Journal of Studies in Education 3*(4), 1-20. https://doi.org/doi:10.5296/jse.v3i4.4168

Sekiyama, K., & Burnham, (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science, 11*, 306-320. https://doi.org/10.1111/j.1467-7687.2008.00677.x

van Wassenhove, V. (2013). Speech through ears and eyes: Interfacing the senses with the supramodal brain. *Frontiers in Psychology, 12*(4), 388. https://doi.org/10.3389/fpsyg.2013.00388.