Does what you see affect listening? Multimodality in listening

Marc Jones TU Dortmund/Toyo University marc@getgreatenglish.com 18/05/2025 JALT PanSIG, Kanda University of International Studies



Multimodality & phonology review

Methodology

Results etc.

Listening is multimodal

Acoustics

Visual

Contextual



Does (multi)modality affect phonology acquisition?

Multimodality theory for listening

Dual coding (Clark & Paivio, 1991) information 'reinforc

Cross (2009; 2011) news videos. To an extent, scaffolded vocabulary.

Suvorov (2018) test takers tend not to look at video.

For English, orthography can be confusing: phonemes & graphemes not 1-to-1.

Orthography

Helps (Showalter & Hayes-Harb, 2013) AND hinders speech learning (Mathieu, 2016).

Difficult to unteach orthography-based phonology (Bassetti et al., 2022).

Pictures/graphics?

- aid comprehension (Aldera, 2015)
- may distract (Suvorov, 2008).

Learners may often (Wagner, 2007) or may not (Suvorov, 2015) look at visuals.

Visuals affect speech perception

Articulatory motor cortex: speech production **and** perception (Glanz et al., 2018).



Bundgaard-Nielsen et al., (2012)

"the speaker will perceive the articulatory gestural basis of his or her own productions and the productions of others, and subsequently fine tune the gestures involved into particular speech acts" (p. 645).

Rosenblum (2008)

Multimodal speech perception "gesturally-based so as to inform about articulatory dynamics and support gestural perceptual primitives" (p. 645).

Video of articulation

Work on consonants shows positive effects (Hardison, 2003; Sueyoshi & Hardison, 2005)

For English vowels? Only a laboratory-based study (Masapollo et al., 2017). Impractical to carry out in the classroom.

Experiment with 2 intact classes

Experiment

1st year university students, dominant language Japanese.

Pretest - post-test (N=35) - delayed post test (N=24)

TED Talks using Global Englishes speakers (same as Jones & Blume, 2022)

Groups

Class 3 (n = 19, delayed n=12): audiovisual listening

Class 4 (n = 16, delayed n = 12) audio only listening

to isolated vowels /æ/, /ʌ/, /ɜː/, and /ɔː/

same vowels in utterance context

utterances within TED video edited for length





Post-test gains

Class	Mean Gain	Mean /æ/	Mean /∧/	Mean /ɜː/	Mean /ɔː/
3	3.769	0.923	1.154	0.846	0.846
4	2.714	0.286	1.071	1.143	0.214

Delayed post-test gains

Class	Mean Gain	Mean /æ/	Mean /∧/	Mean /ɜː/	Mean /ɔː/
3	0.250	1.000	-0.167	-0.333	-0.250
4	0.417	0.250	0.250	0.167	0.750

Conclusion

Visual modality works?



BUT not robust post instruction

Audio only gains lower but appear more robust

Balance

Teach with audiovisual media AND audio only so that you can have faster initial development but keep it robust after instruction.



Limitations

Remember

- Vowels only, and only a subset
 - Consonants different?
- Global Englishes
 - One variety better?
- One setting, upper intermediate (~IELTS 6.0)
 - Same with lower proficiency learners?

Aldera, A. S. (2015). Investigating Multimedia Strategies to Aid L2 Listening Comprehension in EFL Environment. Theory and Practice

in Language Studies, 5(10), 1983–1988. https://doi.org/10.17507/tpls.0510.01

Bassetti, B., Cerni, T., & Masterson, J. (2022). The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology. *Applied Psycholinguistics*, *43*(3), 683–705.

https://doi.org/10.1017/S014271642200008X

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated

with improved second language vowel intelligibility. Applied Psycholinguistics, 33(3), 643-664.

https://doi.org/10.1017/S0142716411000518

Clark, J. M., & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review*, 3(3), 149–210.

https://doi.org/10.1007/BF01320076

Cross, J. (2009). Effects of listening strategy instruction on news videotext comprehension. Language Teaching Research, 13(2),

151–176. https://doi.org/10.1177/1362168809103446

Cross, J. (2011). Comprehending news videotexts: The influence of the visual content. *Language Learning & Technology*, *15*(2), 44–68. <u>http://dx.doi.org/10125/44251</u>

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24(4), 495–522. <u>https://doi.org/10.1017/S0142716403000250</u>

Jones, M., & Blume, C. (2022). Accent difference makes no difference to phoneme acquisition. Teaching English as a Second or

Foreign Language Journal--TESL-EJ, 26(3), 1–22. <u>https://doi.org/10.55593/ej.26103a3</u>

Masapollo, M., Polka, L., & Ménard, L. (2017). A universal bias in adult vowel perception – By ear or by eye. Cognition, 166, 358–370.

https://doi.org/10.1016/j.cognition.2017.06.001

Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. Second Language

Research, 32(2), 145–170. https://doi.org/10.1177/0267658315601882

Nishi, K., & Kewley-Port, D. (2005). Training Japanese listeners to identify American English vowels. The Journal of the Acoustical

Society of America, 117(4), 2401–2401. https://doi.org/10.1121/1.4786064

Nishi, K., & Kewley-Port, D. (2008). Non-native Speech Perception Training Using Vowel Subsets: Effects of Vowels in Sets and Order

of Training. Journal of Speech, Language, and Hearing Research : JSLHR, 51(6), 1480–1493.

https://doi.org/10.1044/1092-4388(2008/07-0109)

Rosenblum, L. D. (2008). Primacy of Multimodal Speech Perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (1. publ. in paperback, pp. 51–78). Blackwell Publ.

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon

study. Second Language Research, 29(2), 185–200. https://doi.org/10.1177/0267658313480154

Sueyoshi, A., & Hardison, D. M. (2005). The Role of Gestures and Facial Cues in Second Language Listening Comprehension.

Language Learning, 55(4), 661–699. <u>https://doi.org/10.1111/j.0023-8333.2005.00320.x</u>

Suvorov, R. (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing*, *32*(4), 463–483. <u>https://doi.org/10.1177/0265532214562099</u>

Suvorov, R. (2018). Test takers' use of visual information in an L2 video-mediated listening test: Evidence from cued retrospective

reporting. In G. J. Ockey & E. Wagner (Eds.), Assessing L2 listening: Moving towards authenticity (pp. 145–160). John Benjamins

Publishing Company. https://doi.org/10.1075/Illt.50.10suv

Suvorov, R. S. (2008). Context visuals in L2 listening tests: The effectiveness of photographs and video vs. audio-only format (p.

7029044) [Master of Arts, Iowa State University, Digital Repository]. https://doi.org/10.31274/rtd-180813-16671

Wagner, E. (2007). Are They Watching? Test-Taker Viewing Behavior During an L2 Video Listening Test. *Language Learning & Technology*, *11*(1), 67–86.

Get in touch!

Contacts Website

marc@getgreatenglish.com

@marcjones@scholar.social

http://marcjones.tokyo/