Lang. Teach. (2010), **43:1**, 84–95 © Cambridge University Press 2009
doi:10.1017/S0261444809990176

Research Timeline

Visual and auditory input in second-language speech processing

Debra M. Hardison Michigan State University, East Lansing, USA
hardiso2@msu.edu

The majority of studies in second-language (L2) speech processing have involved unimodal (i.e., auditory) input; however, in many instances, speech communication involves both visual and auditory sources of information. Some researchers have argued that multimodal speech is the primary mode of speech perception (e.g., Rosenblum 2005). Research on auditory-visual (AV) input has been conducted more extensively in the fields of infant speech development (e.g., Meltzoff & Kuhl 1994), adult monolingual processing (e.g., McGurk & MacDonald 1976; see reference in this timeline), and the treatment of the hearing impaired (e.g., Owens & Blazek 1985) than in L2 speech processing (Hardison 2007). In these fields, the earliest visual input was a human face on which lip movements contributed linguistic information. Subsequent research expanded the types of visual sources to include computer-animated faces or TALKING HEADS (e.g., Massaro 1998), hand-arm gestures (Gullberg 2006), and various types of electronic visual displays such as those for pitch (Chun, Hardison & Pennington 2008). Recently, neurophysiological research has shed light on the neural processing of language input, providing another direction researchers have begun to explore in L2 processing (Perani & Abutalebi 2005).

This timeline on AV input in speech processing is constructed around the following themes based on the studies that have informed the directions this relatively new area of research is taking in L2 acquisition.

- **A** L2 auditory perception studies establishing the groundwork for AV research
- **B** Input from the human face
  - **1** Studies assessing the discernibility of articulatory gestures
  - **2** Native speaker perception
    - **a** Eye-tracking studies
  - **3** L2 learner perception
- **C** Input from computer-animations or talking heads
  - **1** Native speaker perception
  - **2** L2 learner perception
- **D** Electronic visual input (e.g., pitch displays, waveforms) in L2 learning
- **E** Input involving hand-arm gestures in language comprehension
- **F** Neurophysiological studies of AV processing

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

References

Chun, D. M., D. M. Hardison & M. C. Pennington (2008). Technologies for prosody in context: Past and future of L2 research and practice. In J. Hansen Edwards & M. Zampini (eds.), *Phonology and second language acquisition*. Amsterdam: John Benjamins, 323–346.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition. *International Review of Applied Linguistics in Language Teaching* 44.2, 103–124.

Hardison, D. M. (2007). The visual element in phonological perception and learning. In M. C. Pennington (ed.), *Phonology in context*. Basingstoke: Palgrave Macmillan, 135–158.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

McGurk, H. & J. MacDonald (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.

Meltzoff, A. N. & P. K. Kuhl (1994). Faces and speech: Intermodal processing of biologically relevant signals in infants and adults. In D. J. Lewkowitz & R. Lickliter (eds.), *The development of intersensory perception: Comparative perspectives*. Hillsdale, NJ: Lawrence Erlbaum, 335–369.

Owens, E. & B. Blazek (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research* 28, 381–393.

Perani, D. & J. Abutalebi (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology* 15, 202–206.

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (eds.), *The handbook of speech perception*. Malden, MA: Blackwell, 51–78.

DEBRA HARDISON is Associate Professor in the Department of Linguistics and Languages at Michigan State University. Her research program focuses on auditory-visual integration in spoken language processing, applications of technology in perception and production studies, coordination of gesture and prosody, and relationships among learner affect and language use factors, acoustic measures, and rater assessments of the components of speech involving different tasks. She teaches courses in second language acquisition, research methods, and second-language speech. Publications appear in various journals and edited collections. She is guest editor of the October 2009 issue of *Language Learning & Technology*.

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME(S)
1954	Sumby, W. H. & I. Pollack (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* 26.2, 212–215.	This early study of the contribution of visual cues to speech processing noted that the presence of ambient noise is more often the rule than the exception. Speech intelligibility can be substantially enhanced by observation of the speaker. Over the next several decades, research in speechreading (e.g., BINNIE ET AL. 1974), AV processing by monolinguals (e.g., MCGURK & MACDONALD 1976) and L2 learners (e.g., HARDISON 1996, 2003) expanded on this notion.	**B1**, **B2**
1971	Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. *Neuropsychologia* 9.3, 317–323.	**Goto** focused on Japanese learners' difficulty in identifying /r/ and /l/. In an early reference to the potential of visual cues for L2 learners, Goto noted that in the auditory presentation of sounds, 'there was the disadvantage of not being able to read the lips of the speaker' (p. 321). The contribution of visual cues in L2 processing was not investigated until HARDISON (1996).	**A**
1974	Binnie, C. A., A. A. Montgomery & P. L. Jackson (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech & Hearing Research* 17, 619–630.	**Binnie et al**. identified HOMOPHENOUS groups of consonant articulations (categories within which sounds are not visually distinguishable) based on their discernibility in a CV syllable with /a/: /p b m/, /f v/, /θ ð/, /ʃ ʒ/, and /t d n s z k g/ (/r/ and /l/ were omitted). Later studies demonstrated contextual variability (e.g., BENGUEREL & PICHORA-FULLER 1982).	**B1**, **B2**
1976	McGurk, H. & J. MacDonald (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.	This classic paper established the 'McGurk effect' in English – a perceptual phenomenon demonstrating the influence of lip movements on auditory speech perception. Under conditions of AV discrepancy (e.g., visual /ka/ with a dubbed auditory /pa/), observers may report hearing an intermediate consonant (/ta/). Using mismatched AV cues is a way of tagging the information from the two input sources to evaluate their relative contributions to the perceptual outcome. Interestingly, what came to be known as the McGurk effect was, in fact, an accidental discovery. The effect was later explored for speakers of other languages (SEKIYAMA & TOHKURA 1991; MASSARO ET AL. 1993) and L2 learners (HARDISON 1996; SEKIYAMA 1997).	**B2**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

1977	Walden, B. E., R. A. Prosek, A. A. Montgomery, C. K., Scherr & C. J. Jones (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research* 20, 130–145.	**Walden et al.** demonstrated that the information value of visual cues could be improved with training. Results for hearing-impaired adults showed visual recognition of /r/ (in the syllable /ra/) increased the most dramatically of all consonants, from 36% to 89%, after a total of 14 hours of speechreading training. Most of the improvement was reached during the first 5 hours – a time frame that appeared in later L2 perception training studies involving /r/ and /l/ (e.g., LIVELY ET AL. 1993).	**B2**
1982	Benguerel, A-P. & M. K. Pichora-Fuller (1982). Coarticulation effects in lipreading. *Journal of Speech and Hearing Research* 25, 600–607.	In contrast to many speechreading studies (e.g., BINNIE ET AL. 1974), this research considered the influence of vocalic context by testing consonants with the point vowels /i a u/. Contextual variability was addressed later in L2 auditory studies where it was limited to the target sound's position in a word (e.g., LIVELY ET AL. 1993), and in L2 AV processing, which considered both word position and vocalic context (HARDISON 2003).	**B1**, **B2**
1982	Kricos, P. B. & S. A. Lesner (1982). Differences in visual intelligibility across talkers. *Volta Review* 84, 219–225.	**Kricos & Lesner** demonstrated that visual categories (or *Visemes*) vary according to the talker. The influence of talker variability on perception emerged later in auditory (LIVELY ET AL. 1993) and AV perception training studies (HARDISON 2003), challenging the traditional notion of listeners having to normalize the speech signal in order to process the input.	**B1**, **B2**
1983	de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech* 26.4, 331–350.	Earlier studies had advocated the use of speech technology programs to provide displays of intonation for language learners but had not tested their efficacy. **de Bot** demonstrated that such AV feedback (vs. auditory-only) resulted in significantly better production of L2 English by Dutch learners. This research was explored further in HARDISON (2004).	**D**
1987	Hattori, T. (1987). A study of nonverbal intercultural communication between Japanese and Americans – Focusing on the use of the eyes. *Japan Association of Language Teachers* 8, 109–118.	This survey's results provided insight into the language learners' perception of the value of visual cues in speech processing. Japanese students who had lived in the U.S.A. for two years reported they had increased their attention to their interlocutor's face. **Hattori** suggested the practice developed because of American customs and the necessity to 'catch as much information as possible from their interlocutor' (p. 115) to augment their listening comprehension. The nature of this facilitation was examined by HARDISON (1996, 2003).	**B3**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME(S)
1987	Reisberg, D., J. McLean & A. Goldfield (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (eds.), *Hearing by eye: The psychology of lip-reading*. London: Lawrence Erlbaum, 97–113.	SUMBY & POLLACK (1954) had emphasized the value of visual cues because of the frequent presence of ambient noise. **Reisberg et al.** found lip movements significantly improved performance when the auditory signal was clear and intact, but difficult to understand (e.g., shadowing a foreign language). They suggested that speech perception is an AMODAL process.	**B2**, **B3**
1989	Summerfield, Q., A. MacLeod, M. McGrath & M. Brooke (1989). Lips, teeth, and the benefits of lipreading. In A. W. Young & H. D. Ellis (eds.), *Handbook of research on face processing*. Amsterdam: Elsevier, 223–233.	This study investigated the importance of seeing the teeth when speechreading vowels in order to discriminate between those with similar lip shapes. However, seeing internal articulators did not provide an advantage in a later study for L2 learners using a computer-animated face to identify /r/ and /l/ (MASSARO & LIGHT 2003).	**B1**, **B2**
1990	Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (eds.), *New Sounds 90*. Amsterdam: University of Amsterdam, 72–97.	This study served two important roles. It demonstrated computer-based training in the perception and production of Chinese lexical tones by Dutch speakers. It also emphasized the importance of understanding the role of variability. Leather noted: 'too much or too little variability at too early a stage may prevent the learner from discovering with sufficient accuracy the prototypical forms that exemplars expound' (p. 96). Exemplar or episodic models in speech perception and lexical representation emerged in the field in the 1990s and have continued (e.g., LIVELY ET AL. 1993; HARDISON 2003, 2005b).	**D**
1991	Sams, M., R. Aulanko, M. Hämäläinen, R. Hari, O.V. Lounasmaa, S.-T. Lu & J. Simola (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters* 127, 141–145.	This study added to existing behavioral evidence regarding the influence of visual cues on auditory processing (e.g., MCGURK & MACDONALD 1976). **Sams et al.** found that visual input specifically from lip movements influenced activity in the auditory cortex.	**B2**, **F**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

1991	Sekiyama, K. & Y. Tohkura (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America* 90.4, 1797–1805.	Studies exploring the McGurk effect had typically involved only English speakers (e.g., MCGURK & MACDONALD 1976). **Sekiyama & Tohkura** investigated the effect with native speakers of Japanese presented with Japanese syllables. The occurrence of the effect was contingent upon auditory intelligibility and the presence of noise.	**B2**
1993	Lively, S. E., J. S. Logan & D. B. Pisoni (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94.3, 1242–1255.	This study was part of a series investigating the auditory training of Japanese speakers to perceive English /r/ and /l/. It established the criteria for successful perception training, including the use of multiple exemplars produced by a range of talkers (i.e., a HIGH VARIABILITY STIMULUS SET), and served as the foundation for subsequent training studies including HARDISON (1996).	**A**
1993	Massaro, D. W., M. M. Cohen, A. Gesi, R. Heredia & M. Tsuzaki (1993). Bimodal speech perception: An examination across languages. *Journal of Phonetics* 21, 445–478.	In contrast to the findings of SEKIYAMA & TOHKURA (1991), results of this study using computer-animated faces demonstrated the occurrence of the McGurk effect in Japanese, Spanish, and American English.	**C1**
1996	Hardison, D. M. (1996). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect. *Language Learning* 46.1, 3–73.	This was the first study exploring the McGurk effect (MCGURK & MACDONALD 1976) with L2 learners of English (L1s: Japanese, Korean, Spanish, and Malay). Learners were presented with both matched and mismatched CV syllables. With mismatched AV cues, the cue learners could identify the best contributed more to each percept. With matched cues, Japanese speakers identified /r/ significantly more accurately when visual /r/ was present. This visual advantage occurred with /r/ and /f/ for Korean speakers. Results pointed to the potential of visual cues from a talker's face for L2 perception training (see HARDISON 2003).	**B2**, **B3**
1997	Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics* 59.1, 73–80.	In contrast to HARDISON's (1996) findings for L2 learners, a weak McGurk effect was found for Chinese speakers living in Japan when presented with Japanese stimuli. Generally, the longer their stay in Japan, the stronger the effect.	**B3**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME(S)
1998	Vatikiotis-Bateson, E., I-M. Eigsti, S. Yano & K. G. Munhall (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics* 60.6, 926–940.	Although a talker's lip movements are the primary source of linguistic information on the face (e.g., McGurk & MacDonald 1976), other areas may contribute to perception. Eye-tracking data from one Japanese- and one English-speaking observer showed they spent 45%–70% of each AV stimulus presentation in their respective L1s gazing at the talker's eyes. **Vatikiotis-Bateson et al.** suggested phonetically relevant information may be distributed on the face beyond the area around the mouth as a result of changes in facial muscles that accompany articulatory movements. They speculated that observers detect 'well-learned, phonetically correlated events' (p. 938).	**B2a**
1999	Lansing, C. R. & G. W. McConkie (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech, Language, and Hearing Research* 42, 526–539.	Eye-tracking data from a speechreading task involving native speakers of English showed the observers' eye gaze was mostly directed toward the talker's mouth. Gazes in the middle region of the face accounted for 37% of the data – gazes that would have been assigned to one of the eye regions in the study by Vatikiotis-Bateson et al. (1998). Gaze directed toward the nose might represent visual attention to the face as a whole with subsequent shifts to other areas to meet task demands.	**B2a**
2001	Burnham, D., S. Lau, H. Tam & C. Schoknecht (2001). Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. In *Proceedings of AVSP 2001 International Conference on Auditory-Visual Speech Processing, Aalborg Denmark, September 7–9, 2001*. (CD-ROM)	Prior L2 studies investigating visual cues had focused on segmental information (Hardison 1996). **Burnham et al.** found that both tonal (Thai) and non-tonal language (Australian English) speakers were able to discriminate between tone pairs in Cantonese based on visual information when nothing else was available, even when they were unfamiliar with the language. The visual cue may have been neck and/or head movements.	**B3**
2002	Chun, D. M. (2002) *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.	This book on L2 intonation includes numerous screen shots showing displays of pitch contours and waveforms. The companion CD-ROM plays these examples with a cursor that moves through the display, aligned in time with the audio. **Chun** emphasizes that such displays are sources of informative feedback for L2 learners. They continue to be the topic of investigation (e.g., Hardison 2004, 2005a; Levis & Pickering 2004).	**D**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

2003	Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics* 24.4, 495–522.	Based on the findings of HARDISON (1996), this study took the next step by comparing AV and auditory-only perception training for Japanese and Korean learners of L2 English, with a focus on /r/ and /l/. AV training was significantly more effective. Visual input contributed more in the most difficult contexts. There were significant effects of talker, word position, and adjacent vowel. Vowel effects had not been explored in L2 perception training (e.g., LIVELY ET AL. 1993), but their influence was suggested by studies such as BENGUEREL & PICHORA-FULLER (1982). Successful generalization obtained for novel stimuli and a new talker, with transfer to improved production.	**B3**
2003	Hickok, G., B. Buchsbaum, C. Humphries & T. Muftuler (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience* 15.5, 673–682.	Based on results of functional magnetic resonance imaging (fMRI),[1] **Hickok et al.** suggested that there is a predominantly left hemisphere network that enables acoustic-phonetic input to guide the acquisition of language-specific articulatory gestures, and underlies phonological working memory in adults. These findings have implications for future L2 neuroimaging studies involving AV integration.	**F**
2003	Lansing, C. R. & G. W. McConkie (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics* 65.4, 536–552.	In this eye-tracking study **Lansing & McConkie** demonstrated that by the time a talker's face appeared on the screen, observers had often shifted their eye gaze from the center where the nose would appear toward the regions where the eyes or mouth would appear. The talker's eye area attracted attention before and after speaking (EYE PRIMACY EFFECT). In 30%–50% of the cases, the observer's eyes shifted to the region where the talker's mouth would appear (INFORMATION SOURCE ATTRACTION EFFECT: visual attention is drawn to a location they anticipate will offer information). This attraction effect may be stronger in more challenging speech environments (e.g., degraded or missing audio signal), and for more challenged perceivers such as L2 learners. The latter possibility has yet to be explored.	**B2a**

[1] fMRI is a method of capturing and creating an image of the activity in the brain through measurement of the magnetic fields created by the functioning nerve cells in the brain associated with a given task (Hickok et al. 2003).

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME(S)
2003	Massaro, D. W. & J. Light (2003). Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. In *Proceedings of Eurospeech (Interspeech), 8th European Conference on speech Communication and Technology*. Geneva, Switzerland. http://mambo.ucsc.edu/pdf/read_my_tongue.pdf.	Japanese learners of English were trained to perceive /r/ and /l/ using the normal view of Baldi, a talking head, and the view showing movement of the internal articulators. The normal view improved identification and production accuracy, consistent with the use of human faces (HARDISON 2003). When the articulators could be seen, improvement occurred but did not provide an additional benefit. This is in contrast to the findings of SUMMERFIELD ET AL. (1989), which indicated that seeing the teeth facilitated vowel discrimination in speechreading.	**C2**
2004	Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology* 8.1, 34–52. http://llt.msu.edu/vol8num1/hardison.	The intonation study by DE BOT (1983) used delayed video feedback and an imitation approach. In this study, English speakers received training to improve their production of French prosody using a real-time computerized pitch display and native-speaker utterances as feedback. Results revealed significant improvement. Following the criteria for demonstrating successful training (e.g., LIVELY ET AL. 1993), improvement generalized to segmental production and novel sentences.	**D**
2004	Levis, J. & L. Pickering (2004). Teaching intonation in discourse using speech visualization technology. *System* 32.4, 505–524.	HARDISON (2004) used individual sentences in prosody training with computerized displays. **Levis & Pickering** discussed the applications of pitch visualization technology for L2 learners to emphasize the intonational characteristics of discourse-level language.	**D**
2004	Munhall, K. G., J. A. Jones, D. E. Callan, T. Kuratate & E. Vatikiotis-Bateson (2004). Visual prosody and speech intelligibility. *Psychological Science* 15.2, 133–137.	Similar to the study by BURNHAM ET AL. (2001) suggesting the existence of visual cues to tone, **Munhall et al.**, using talking-head animations, found a speaker's head movements contributed to identification of words by native speakers when presented in noise. Because of the correlation between head movement and vocal pitch, visible head motion in conjunction with prosody may aid in segmenting the speech stream, suggesting another direction for L2 research.	**C1**
2005	Hardison, D. M. (2005a). Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal* 22.2, 175–190.	Following the trend toward a focus on discourse-level language in L2 processing (e.g., LEVIS & PICKERING 2004), this study built on the successful improvement in prosody found by HARDISON (2004) for individual sentences, and investigated the feedback potential of visualizations of pitch contours for	**D**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

		discourse-level language. Pitch displays were synchronized with the video of the speech event in prosody training for advanced L2 speakers of English. Discourse-level input with video produced the best transfer to novel natural discourse.	
2005	Hardison, D. M. (2005b). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics* 26.4, 579–596.	Visual cues from a talker's face were beneficial in segmental-level perception (HARDISON 1996) and training (HARDISON 2003). In this study, both visual cues and segmental training improved learners' word identification ability. The AV advantage was especially evident for words beginning with /r/ and /l/. Because articulatory movements precede the associated acoustic signal, visual cues play a priming role in the process for listener-observers in face-to-face situations.	**B3**
2005	Hardison, D. M. (2005c). Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style. *Speech Communication* 46.1, 73–93.	Previous work on L2 spoken word identification used scripted stimuli (e.g., HARDISON 2005b). This study investigated the influence of speech style (unscripted vs. scripted) and visual cues from a talker's face on spoken word identification. Findings highlighted the priming role of visual cues, and the talker-and context-dependent nature of language processing consistent with perception training studies (e.g., LIVELY ET AL. 1993; HARDISON 2003), but did not support a strict conversational–clear-speech dichotomy.	**B2**, **B3**
2005	Hazan, V., A. Sennema, M. Iba & A. Faulkner (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47.3, 360–378.	There is variability in the degree to which visual cues are helpful for L2 learners. In this study, AV training was more effective than auditory-only in improving perception of the /v/–/b/ contrast. However, in contrast to HARDISON (2003), Japanese participants who received AV training did not improve more in perception of /r/ and /l/ than those receiving A-only input. Lipreading performance improved the most for learners who saw a natural face versus a synthetic one (MASSARO & LIGHT 2003). **Hazan et al.** concluded AV training is most effective when critical visual cues are sufficiently salient.	**B3**, **C2**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME(S)
2005	Sueyoshi, A. & D. M. Hardison (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning* 55.4, 661–699.	Expanding on studies focusing on facial cues to spoken language processing (e.g., HARDISON 2003, 2005b; HAZAN ET AL. 2005), this study compared the contributions of a talker's facial cues and co-speech gestures to L2 listening comprehension. For lower proficiency learners, seeing the face and gestures produced the best results; for the advanced group, seeing only the face produced the highest scores. Auditory-only input yielded the lowest scores. Questionnaire responses revealed positive attitudes toward both types of visual cues.	**B3**, **E**
2006	Hardison, D. M. (2006). Effects of familiarity with faces and voices on L2 spoken language processing: Components of memory traces. In *Interspeech 2006: Proceedings of the Ninth International Conference on Spoken Language Processing*. Bonn: International Speech Communication Association, 2462–2465. https://www.msu.edu/~hardiso2/Hardison.ICSLP2006.pdf.	Results of segmental perception training studies supported an episodic view of the memory encoding of speech (LIVELY ET AL. 1993; HARDISON 2003). This study further explored the perceptual details that L2 learners of English retain in memory in the process of word identification by investigating the effects of familiarity with areas of talkers' faces and voices. When the talker was familiar, seeing just the mouth area provided comparable identification accuracy to seeing the entire face. Word identification was also more accurate when the eyes and upper cheek areas were visible compared to just hearing the voice, but only when a talker was familiar. Findings provided evidence for episodic models of lexical representation.	**B3**
2007	Skipper, J. I., V. van Wassenhove, H. W. Nusbaum & S. L. Small (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex* 17.10, 2387–2399.	**Skipper et al.** suggested that visible articulatory movements allow the listener-observer's brain to reduce the set of potential targets the speaker will likely produce. Informative visual cues (e.g., for /p b m/) offer more precise predictions. These findings provide a neurophysiological explanation for the priming role of visual cues and support the transfer of segmental-level perception training to earlier word identification (HARDISON 2005b, c).	**F**
2009	Kelly, S. D., T. McDevitt & M. Esch (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes* 24.2, 313 – 334.	SUEYOSHI & HARDISON (2005) demonstrated that co-speech gestures facilitated listening comprehension by L2 learners. **Kelly et al.** found such gestures also facilitated recall of Japanese words learned by L1 English adults with no prior knowledge of the language. Based on memory tests and	**E**, **F**

https://doi.org/10.1017/S0261444809990176 Published online by Cambridge University Press

		event-related potentials (ERPs),[2] Kelly et al. suggested that co-speech gesture 'deepens the imagistic memory trace for a new word's meaning in the brain' (p. 330).	
2009	Motohashi-Saigo, M. & D. M. Hardison (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology* 13.2, 29–47. http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf.	Previous studies found pitch displays to be user friendly and helpful (e.g., CHUN 2002; HARDISON 2004, 2005). **Motohashi-Saigo & Hardison** found that training involving waveforms to visualize duration differences resulted in significant improvement in the perception of L2 Japanese geminates by English speakers, with transfer to production. Participants offered positive comments on Web-based training.	**D**
2009	Hincks, R. & J. Edlund (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology* 13.3, 32–50.	Instead of pitch contours (e.g., CHUN 2002; HARDISON 2004), the intonation feedback system of **Hincks & Edlund** produces flashing lights of different colors, reflecting the user's degree of pitch variation. Following training with this system, L2 learners of English showed a significant increase in pitch variation in giving an oral presentation and improvement in naturalness.	**D**
In press	Hardison, D. M. (in press). Second-language speech perception: A cross-disciplinary perspective on challenges and accomplishments. To appear in A. Mackey & S. Gass (eds.), *Routledge handbook of second language acquisition*. New York: Routledge.	This chapter examines progress in L2 spoken language processing research. In one of the reported studies, visual cues and sentence context played significant yet independent roles. On average, learners needed 11% less of a target word for identification when visual cues were present, and 9% less with sentence context. Context had long been assumed to help learners but had not been measured. Like visual cues (HARDISON 2005b, c), context serves a priming role in speech processing, reducing the set of potential word candidates.	**B3**

[2] ERPs measure the timing of electrical brain responses to a stimulus. The brainwaves correspond to different types of neurocognitive processes (Kelly et al. 2009).