



Audiovisual integration: an introduction to behavioral and neuro-cognitive methods

Nicholas Altieri*

Communication Sciences and Disorders, Idaho State University, Pocatello, ID, USA

*Correspondence: altinich@isu.edu

Edited by:

Manuel Carreiras, Basque Center on Cognition, Brain and Language, Spain

Keywords: audiovisual speech, integration, brain, speech and cognition, neuroimaging of speech, quantitative methods multisensory speech

Advances in neurocognitive and quantitative behavioral techniques have offered new insights to the study of cognition and language perception. This includes ways in which neurological processes and behavior are intimately intertwined. Examining traditional behavioral measures and model predictions, along with neurocognitive measures, will provide a powerful theory-driven and unified approach for researchers in the cognitive and language sciences. In this topic, the aim was to highlight some of the noteworthy methodological developments in the burgeoning field of multisensory speech perception.

Decades of research on audiovisual speech integration has, broadly speaking, reshaped the way language processing is conceptualized in the field. Beginning with Sumby and Pollack's seminal study of audiovisual integration published in 1954, qualitative and quantitative relationships have emerged showing the benefit of being able to obtain visual cues from "speech reading" under noisy conditions. A pioneering study by McGurk and MacDonald (1976) further demonstrated a form of integration phenomenon in which incongruent auditory-visual speech signals contribute to a fused or combined percept. (One such example is an auditory "ba" dubbed over a video of a talker articulating the syllable "ga." This often yields a combined percept of "da.")

Methods for determining whether "integration" occurs have, for example, involved examining whether a listener is susceptible to the McGurk effect, as we shall in a study by Setti et al. (2013) in the Research Topic. Perhaps a more commonly used assessment tool for determining the presence of "integration" has been measuring the extent to which a dependent variable (accuracy, speed, etc.) obtained from audiovisual trials is significantly "better" than the predicted response obtained from the unisensory conditions. A difference between obtained and predicted measures is thought to indicate a violation of independence between modalities (Altieri and Townsend, 2011; Altieri et al., 2013). In recent years, the neurological bases of these multisensory phenomena in speech perception have been developed largely in parallel with advances in behavioral techniques. Neuroimaging studies have looked at the Blood Oxygen-Level Dependent (BOLD) signal in relation to AV speech stimuli and compared that to the unisensory BOLD responses (e.g., Calvert, 2001; Stevenson and James, 2009). Within the milieu of EEG studies, similar comparisons have been made between the amplitude evoked by audiovisual, vs. auditory and visual-only stimuli. Similar to the fMRI studies, EEG research has contributed to the idea that integration occurs if the AV response differs from the unisensory responses (AV_{ERP}

$< A_{ERP} + V_{ERP}$; see, van Wassenhove et al., 2005; and Winneke and Phillips, 2011).

The application of EEG, fMRI or other imaging techniques in combination with behavioral indexes has therefore enhanced the testability of neural based theories of multisensory language processing. The broader aim of this Research Topic was to investigate the variety of manners in which neural measures of multisensory language processing could be anchored to behavioral indices of integration.

Several pioneering studies appear in this volume addressing a wide variety of issues in multisensory speech recognition. Quite significantly, this research explores integration in different age groups, for individuals with sensory processing deficits, and across different listening environments. First, a study carried out by Altieri and Wenger (2013) sought to rigorously associate the dynamic psychophysical measures of perception—namely the reaction time measure of *workload capacity* (Townsend and Nozawa, 1995)—with a neural dynamics from EEG. Under degraded listening conditions, we observed an increase in integration efficiency as measured by capacity, which co-occurred with an increase in multisensory ERPs relative to auditory-only ERPs. In a much needed review on the rules giving rise to multisensory integration, van Wassenhove (2013) provided an overview of "predictive coding hypotheses." Updated hypotheses were considered, namely concerning how internal predictions about linguistics percepts are formulated. An overview of neuroimaging literature was included in the discussion.

Three reports explored the temporal effects of visual information on auditory encoding. One, provided by Ten Oever et al. (2013), varied the synchrony of the auditory and visual signals to explore the temporal effects of auditory syllable encoding. The results indicated a larger time-window for congruent AV syllables. Second, Moradi et al. (2013) provided a report investigating the influence of visual information on temporal recognition. This study showed that visual cues sped-up linguistic recognition in both noisy and clear listening conditions. Finally, a review and hypothesis article by Hertrich et al. (2013) proposes a brain network explaining how blind individuals, on average, are capable of perceiving auditory speech at a much faster rate compared to individuals with normal vision. Together, these articles will help constrain dynamic and neural-based theories regarding temporal aspects of audiovisual speech perception.

Two studies in this Research Topic also explored the effects of aging and neural development on perceptual skills. Kushnerenko et al. (2013) used an eye tracking paradigm in conjunction with

ERPs to investigate the extent to which these measures predict normal linguistic development in children. Second, Setti et al. (2013) investigated integration skills by looking at whether age is predictive of the susceptibility to the McGurk effect. Interestingly, the authors found that older adults were more susceptible to the fusion than younger ones—ostensibly due to differences in perceptual rather than higher order cognitive processing abilities.

These research and review articles provide a rich introduction to a variety of fascinating techniques for investigating speech integration. Ideally, these research directions will pave the way toward a much improved tapestry of methodologies, and refinements of neuro-cognitive theories of multisensory processing across life-span, listening conditions, and sensory-cognitive abilities.

REFERENCES

- Altieri, N., and Townsend, J. T. (2011). An assessment of behavioral dynamic information processing measures in audiovisual speech perception. *Front. Psychol.* 2:238. doi: 10.3389/fpsyg.2011.00238
- Altieri, N., Townsend, J. T., and Wenger, M. J. (2013). A dynamic assessment function for measuring age-related sensory decline in audiovisual speech recognition. *Behav. Res. Methods*. doi: 10.3758/s13428-013-0372-8. [Epub ahead of print].
- Altieri, N., and Wenger, M. J. (2013). Neural dynamics of audiovisual speech integration under variable listening conditions: an individual participant analysis. *Front. Psychol.* 4:615. doi: 10.3389/fpsyg.2013.00615
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123. doi:10.1093/cercor/11.12.1110
- Hertrich, I., Dietrich, S., and Ackermann, H. (2013). How can audiovisual pathways enhance the temporal resolution of time-compressed speech in blind subjects. *Front. Psychol.* 4:530. doi: 10.3389/fpsyg.2013.00530
- Kushnerenko, E. V., Tomalski, P., Ballieux, H., Potton, A., Birtles, D., Frostick, C., et al. (2013). Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Front. Psychol.* 4:432. doi: 10.3389/fpsyg.2013.00432
- McGurk, H., and MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Moradi, S., Lidestam, B., and Rönnerberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: effects on time and accuracy. *Front. Psychol.* 4:359. doi: 10.3389/fpsyg.2013.00359
- Setti, A., Burke, K. E., Kenny, R., and Newell, F. N. (2013). Susceptibility to a multisensory speech illusion in older persons is driven by perceptual processes. *Front. Psychol.* 4:575. doi: 10.3389/fpsyg.2013.00575
- Stevenson, R. A., and James, T. W. (2009). Neuronal convergence and inverse effectiveness with audiovisual integration of speech and tools in human superior temporal sulcus: evidence from BOLD fMRI. *Neuroimage* 44, 1210–1223. doi: 10.1016/j.neuroimage.2008.09.034
- Sumbly, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 12–15.
- Ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N., and Van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Front. Psychol.* 4:331. doi: 10.3389/fpsyg.2013.00331
- Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial and coactive theories. *J. Math. Psychol.* 39, 321–360. doi: 10.1006/jmps.1995.1033
- van Wassenhove, V. (2013). Speech through ears and eyes: interfacing the senses with the supramodal brain. *Front. Psychol.* 4:388. doi: 10.3389/fpsyg.2013.00388
- van Wassenhove, V., Grant, K., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1181–1186. doi: 10.1073/pnas.0408949102
- Winneke, A. H., and Phillips, N. A. (2011). Does audiovisual speech offer a fountain of youth for old ears. An event-related brain potential study of age differences in audiovisual speech perception. *Psychol. Aging* 26, 427–438. doi: 10.1037/a0021683

Received: 23 August 2013; accepted: 29 August 2013; published online: 17 September 2013.

Citation: Altieri N (2013) Audiovisual integration: an introduction to behavioral and neuro-cognitive methods. *Front. Psychol.* 4:642. doi: 10.3389/fpsyg.2013.00642

This article was submitted to *Language Sciences*, a section of the journal *Frontiers in Psychology*.

Copyright © 2013 Altieri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.