frontiers in
PSYCHOLOGY

OPINION ARTICLE
published: 03 June 2014
doi: 10.3389/fpsyg.2014.00427

Speech is not special... again

Kathy M. Carbonell and Andrew J. Lotto*

Department of Speech, Language and Hearing Sciences, University of Arizona, Tucson, AZ, USA
**Correspondence: alotto@email.arizona.edu*

Edited by:
Kaisa Tiippana, University of Helsinki, Finland

Reviewed by:
Jean Vroomen, University of Tilburg, Netherlands

Keywords: sensorimotor effects on perception, multisensory integration, speech perception, auditory processing, Motor Theory

THE "SPECIALNESS" OF SPEECH

As is apparent from reading the first line of nearly any research or review article on speech, the task of perceiving speech sounds is complex and the ease with which humans acquire, produce and perceive these sounds is remarkable. Despite the growing appreciation for the complexity of the perception of music, speech perception remains the most amazing and poorly understood auditory (and, if we may be so bold, perceptual) accomplishments of humans. Over the years, there has been considerable debate on whether this achievement is the result of general perceptual/cognitive mechanisms or "special" processes dedicated to the mapping of speech acoustics to linguistic representations (for reviews see Trout, 2001; Diehl et al., 2004). The most familiar proposal of the "specialness" of speech perception is the various incarnations of the Motor Theory of speech proposed by Liberman et al. (1967; Liberman and Mattingly, 1985, 1989). Given the status of research into audition in the 1950s and 1960s, it is not surprising that speech appeared to require processing not available in "normal" hearing. Much of the work at the time used relatively simple tones and noises to get at the basic psychoacoustics underlying the perception of pitch and loudness (though some researchers like Harvey Fletcher were also working on some basics of speech perception, Fletcher and Galt, 1950; Allen, 1996). Liberman and his collaborators discovered that the discrimination of acoustic changes in speech sounds did not look like the psychoacoustic measures of discrimination for pitch and loudness. Instead of following a Weber or Fechner law, the discrimination function had a peak near the categorization boundary between contrasting phonemes—a pattern of perceptual results that is referred to as Categorical Perception (Liberman et al., 1957). In addition, the acoustic cues to phonemic identity were not readily apparent with similar spectral patterns resulting in different phonemic percepts and acoustically disparate patterns resulting in identical phonemic percepts—the problem of "lack of invariance" (e.g., Liberman et al., 1952). The perception of these varying acoustic patterns was highly context-sensitive to preceding and following phonetic content in ways that appeared specific to the communicative constraints of speech and not applicable to the perception of other sounds—as in demonstrations of perceptual compensation for coarticulation, speaking rate normalization and talker normalization (e.g., Ladefoged and Broadbent, 1957; Miller and Liberman, 1979; Mann, 1980).

One major source of evidence in favor of a Motor Theory account of speech perception is that information about a speaker's production (anatomy or kinematics) from non-auditory sources can affect phonetic perception. The famed McGurk effect (McGurk and MacDonald, 1976), in which visual presentation of a talker can alter the auditory phonetic percept, is taken as evidence that listeners are integrating information about production from this secondary source. Fowler and Deckle (1991) have demonstrated a similar effect using haptic information gathered by touching the speaker's face (see also Sato et al., 2010). Gick and Derrick (2009) reported that perception of consonant—vowel tokens in noise are biased toward voiceless stops (e.g., /pa/) when they are accompanied by a small burst of air on the skin of the listener, which could be interpreted as the aspiration that would more likely accompany the release of a voiceless stop.

In addition, there have been several studies that have demonstrated that manipulations of the *listener's* articulators can affect perception, which are supportive of the Motor Theory proposal that the mechanisms of production underlie the perception of speech. For example, Ito et al. (2009) obtained shifts in phoneme categorization resulting from external manipulation of the skin around the *listener's* mouth in ways that would correspond to the deformations typical of producing these speech sounds (see also Yeung and Werker, 2013 for a similar demonstration with infants). Recently, Mochida et al. (2013) found that the ability to categorize consonants can be influenced by the simultaneous silent production of these consonants. Typically, these studies are proffered as evidence for a direct role of speech motor processing in speech perception.

Independent of this proposed motor basis of perception, others have suggested the existence of a special speech or phonetic mode of perception based on evidence of neural and behavioral responses to the same stimuli being modulated by whether or not the listener believes the signal to be speech or non-speech (e.g., Tomiak et al., 1987; Vroomen and Baart, 2009; Stekelenburg and Vroomen, 2012).

THE "GENERALITY" OF SPEECH

Since the early work by Liberman and colleagues and the development of the Motor Theory, there has been a growing appreciation for the power of perceptual learning and the context-sensitive nature of auditory processing. Once one begins to study more complex sounds and perceptual behaviors, the distinction between speech

and non-speech processing becomes less clear. So, for example, we now have many examples of non-speech sound categories that demonstrate the characteristics of Categorical Perception (Cutting et al., 1976; Harnad, 1990; Mirman et al., 2004). It also appears that general auditory learning mechanisms are capable of dealing with the lack of invariance problem in formation of categories. Birds can learn speech consonant categories with no obvious acoustic invariant cue (Kluender et al., 1987) and human listeners can readily learn non-speech categories that are similarly structured (Wade and Holt, 2005). Finally, non-speech analogs have been created that result in the same types of context effects earlier witnessed for speech categorization, such as "perceptual compensation for coarticulation" (Lotto and Kluender, 1998; Holt et al., 2000), "speaking rate normalization" (Pisoni et al., 1983; Diehl and Walsh, 1989) and "talker normalization" (Watkins and Makin, 1994; Holt, 2005; Sjerps et al., 2011; Laing et al., 2012).

These findings with non-speech and animal perception of speech sounds (along with many others) call into question the strict dichotomy of speech and general auditory processing (Schouten, 1980). The lack of a clear distinction extends to the famed McGurk effect, which has been successfully modeled using general models of perception (e.g., Massaro, 1998). Stephens and Holt (2010) demonstrated that human adults can learn correlations between features of speech and arbitrary dynamic visual cues that are not related to the gestures of human vocal tracts. Participants in their experiments learned to associate the movements of dials and lighted bars on an animated "robot" display to stimuli varying in vowels and voiced consonant and could use this information to enhance intelligibility in noise. These types of novel mappings demonstrate the effectiveness of perceptual learning even across modalities (though perhaps not leading to as strong of an integration of information as may occur for natural covariations).

THE IMPORTANCE OF RESEARCH INTO MULTISENSORY INTERACTIONS IN SPEECH PERCEPTION

The growth in empirical research into the integration of multisensory information in speech acquisition and perception is a welcome development because it is a recognition that speech is not perceived within a vacuum. Too often, speech perception research has been conducted in an isolated reductionist vein that has made the human accomplishments in speech communication seem almost miraculous. The important realization at the heart of Lindblom's (1990, 1996) Hypo and Hyper Speech Theory is that much of the troubling acoustic variability in speech is actually a result of the changing demands of conversation between two people and the needs for informational precision due to the communication context. When one fails to study speech within a full communication context, this structured variability becomes noise. The isolation of speech research from a communication context has also made it difficult to connect the vast work in phonemic perception with more practical clinical issues in hearing loss and speech pathology. As Weismer and Martin (1992) point out, the concept of intelligibility must include both the speaker and the listener—that is, intelligibility is a measure of the entire communication setting and not just the acoustics of the speaker (see also, Liss, 2007).

The investigation of multisensory integration in speech perception is a step in the direction of attempting to understand the entire communication setting and all of the available information that results in an intelligible message. Some of the well-known findings from an auditory-isolated experiment may in fact be misleading when looked at in this broader context. For example, a highly cited finding is that 9-month-old infants from English-speaking households fail to discriminate a non-native Hindi contrast (Werker and Tees, 1984), which is taken as evidence that they are now perceptually tuned to their native language. However, Yeung and Werker (2009) obtained discrimination for infants in this group when the contrasting sounds were paired consistently with visual novel objects—a situation which mimics more realistically the communication setting of language learning. MacKenzie et al. (2013) in one experiment demonstrated an apparent unwillingness of 12-month-olds to associate novel auditory words with visual objects when the words are not phonotactically acceptable in their native language. However, the infants show far more flexibility in "acceptable" words when the task is preceded by a word-object association game with familiar word-objects. In each of these examples, the presumed perceptual tuning for language becomes less strict once the information available to the infant about the task is expanded. These experiments are stark reminders that speech acquisition and perception occurs in a larger perceptual/cognitive framework. Such results may also extend to adults learning to categorize speech sounds. Lim and Holt (2011) obtained significant increases in categorization performance for Japanese-speaking adults learning the non-native English /l/-/r/ distinction utilizing a video game paradigm. In this game, the categories were associated with different visual creatures that were either "friends" or "enemies" requiring different actions. The implicit mapping of auditory categories to functional dynamic visual objects may account for some of the success of this training.

A CAUTIONARY NOTE

Whereas the section above provides just a few of the many benefits of studying multisensory integration in speech, one must be cautious not to repeat the history of the field by proposing special mechanisms of phenomena for speech perception without thoroughly investigating what processes are available for general perception. The perception of all sound events is almost certainly intrinsically multisensory. Experimental designs that reduce sound event perception to audition run the risk of changing the task demands for the perceiver (as seen above in the examples for speech discrimination in infants).

There are many examples of sound perception being influenced by non-auditory information. Detection of low-intensity sounds is enhanced when paired with a task-irrelevant light stimulus (Lovelace et al., 2003; Odgaard et al., 2004). Saldaña and Rosenblum (1993) reported that when listeners were presented a visual image of a cello either being plucked or bowed, it strongly influenced their auditory judgment of whether the cello was being plucked or bowed. The perceived loudness of tones can be influenced by synchronous tactile information (Schürmann et al., 2004; Gillmeister and Eimer, 2007).

In addition, sensori-motor interactions can be found in music perception (Maes et al., 2013). We should be very cautious in proposing multimodal or sensorimotor interactions that are "special" to speech. It is quite possible that new integrations between senses will be observed using the well-learned complex stimuli of speech sounds (or musical sounds) as opposed to simple noises and tones and unexperienced complex signals. These novel findings should be taken as opportunities to learn general principles of perception, action and cognition as opposed to assigning them special status and missing these opportunities.

Postulating a special speech perception mode or module is a strong theoretical position not to be taken lightly. One must describe how the processes brought to bear in the perception of speech sounds are fundamentally different from those responsible for other forms of complex audition. Speech sounds are "special" in the sense that they are over-learned categories that play a functional role in a larger hierarchical linguistic system. But these attributes on their own do not necessitate the proposal of inherently different processing mechanisms. In the end, speech sounds and the perception/categorization of these sounds is not likely to require special processing. The "specialness" of these sounds comes from being a part of the complex act of communicating. It is the act of communicating that clearly requires integration of the senses and the cooperation of perception and action. We must be wary that speech sound perception ("is this a "ba" or a"da") isolated from the full act of communication is unnatural even when bringing to bear information from other sense modalities. The small and context-specific sensorimotor and multisensory effects we can uncover in this artificial task (Hickok et al., 2009) may not provide much insight into the real act of communication with speech.

REFERENCES

Allen, J. B. (1996). Harvey Fletcher's role in the creation of communication acoustics. *J. Acoust. Soc. Am.* 99, 1825–1839. doi: 10.1121/1.415364

Cutting, J. E., Rosner, B. S., and Foard, C. F. (1976). Perceptual categories for musiclike sounds: implications for theories of speech perception. *Q. J. Exp. Psychol.* 28, 361–378. doi: 10.1080/14640747608400563

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol.* 55, 149–179. doi: 10.1146/annurev.psych.55.090902.142028

Diehl, R. L., and Walsh, M. A. (1989). An auditory basis for the stimulus−length effect in the perception of stops and glides. *J. Acoust. Soc. Am.* 85, 2154–2164. doi: 10.1121/1.397864

Fletcher, H., and Galt, R. H. (1950). The perception of speech and its relation to telephony. *J. Acoust. Soc. Am.* 22, 89–151. doi: 10.1121/1.1906605

Fowler, C., and Deckle, D. (1991). Listening with eye and hand: crossmodal contributions to speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 816–828. doi: 10.1037/0096-1523.17.3.816

Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502–504. doi: 10.1038/nature08572

Gillmeister, H., and Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Res.* 1160, 58–68. doi: 10.1016/j.brainres.2007.03.041

Harnad, S. R. (Ed). (1990). *Categorical Perception: the Groundwork of Cognition*. New York, NY: Cambridge University Press.

Hickok, G., Holt, L. L., and Lotto, A. J. (2009). Response to Wilson: what does motor cortex contribute to speech perception? *Trends Cogn. Sci.* 13, 330–331. doi: 10.1016/j.tics.2009.05.002

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychol. Sci.* 16, 305–312. doi: 10.1111/j.0956-7976.2005.01532.x

Holt, L. L., Lotto, A. J., and Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *J. Acoust. Soc. Am.* 108, 710–722. doi: 10.1121/1.429604

Ito, T., Tiede, M., and Ostry, D. J. (2009). Somatosensory function in speech perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1245–1248. doi: 10.1073/pnas.0810063106

Kluender, K. R., Diehl, R. L., and Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science* 237, 1195–1197. doi: 10.1126/science.3629235

Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *J. Acoust. Soc. Am.* 29, 98–104. doi: 10.1121/1.1908694

Laing, E. J. C., Liu, R., Lotto, A. J., and Holt, L. L. (2012). Tuned with a tune: talker normalization via general auditory processes. *Front. Psychol.* 3:203. doi: 10.3389/fpsyg.2012.00203

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

Liberman, A. M., Delattre, P., and Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am. J. Psychol.* 65, 497–516. doi: 10.2307/1418032

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358. doi: 10.1037/h0044417

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6

Liberman, A. M., and Mattingly, I. G. (1989). A specialization for speech perception. *Science* 243, 489–494. doi: 10.1126/science.2643163

Lim, S.-J., and Holt, L. L. (2011). Learning foreign sounds in an alien world: video game training improves non-native speech categorization. *Cogn. Sci.* 35, 1390–1405. doi: 10.1111/j.1551-6709.2011.01192.x

Lindblom, B. (1990). "Explaining phonetic variation: a sketch of the HandH theory," in *Speech Production and Speech Modelling*, eds W. Hardcastle and M. Kluwer (Netherlands: Springer), 403–439. doi: 10.1007/978-94-009-2037-8_16

Lindblom, B. (1996). Role of articulation in speech perception: clues from production. *J. Acoust. Soc. Am.* 99, 1683–1692. doi: 10.1121/1.414691

Liss, J. M. (2007). "Perception of dysarthric speech," in *Motor Speech Disorders: Essays for Ray Kent*, ed G. Weismer (San Diego, CA: Plural Publishing), 187–219.

Lotto, A. J., and Kluender, K. R. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Percept. Psychophys.* 60, 602–619. doi: 10.3758/BF03206049

Lovelace, C. T., Stein, B. E., and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Res. Cogn. Brain Res.* 17, 447–453. doi: 10.1016/S0926-6410(03)00160-5

MacKenzie, H. K., Graham, S. A., Curtin, S., and Archer, S. L. (2013). The flexibility of 12-month-olds' preferences for phonologically appropriate object labels. *Dev. Psychol.* 50, 422–430. doi: 10.1037/a0033524

Maes, P. J., Leman, M., Palmer, C., and Wanderley, M. M. (2013). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Percept. Psychophys.* 28, 407–412. doi: 10.3758/BF03204884

Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*, Vol. 1. Cambridge, MA: MIT Press.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Miller, J. L., and Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Percept. Psychophys.* 25, 457–465. doi: 10.3758/BF03213823

Mirman, D., Holt, L. L., and McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: differences between steady-state and rapidly-changing acoustic cues. *J. Acoust. Soc. Am.* 116, 1198–1207. doi: 10.1121/1.1766020

Mochida, T., Kimura, T., Hiroya, S., Kitagawa, N., Gomi, H., and Kondo, T. (2013). Speech misperception: speaking and seeing interfere differently with hearing. *PLoS ONE* 8:e68619. doi: 10.1371/journal.pone.0068619

Odgaard, E. C., Arieh, Y., and Marks, L. E. (2004). Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. *Cogn. Affect. Behav. Neurosci.* 4, 127–132. doi: 10.3758/CABN.4.2.127

Pisoni, D. B., Carrell, T. D., and Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Percept. Psychophys.* 34, 314–322. doi: 10.3758/BF03203043

Saldaña, H. M., and Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Percept. Psychophys.* 54, 406–416.

Sato, M., Cavé, C., Ménard, L., and Brasseur, A. (2010). Auditory-tactile speech perception in congenitally blind and sighted adults. *Neuropsychologia* 48, 3683–3686. doi: 10.1016/j.neuropsychologia.2010.08.017

Schouten, M. E. H. (1980). The case against a speech mode of perception. *Acta Psychol.* 44, 71–98. doi: 10.1016/0001-6918(80)90077-3

Schürmann, M., Caetano, G., Jousmäki, V., and Hari, R. (2004). Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. *J. Acoust. Soc. Am.* 115, 830–832. doi: 10.1121/1.1639909

Sjerps, M. J., Mitterer, H., and McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Atten. Percept. Psychophys.* 73, 1195–1215. doi: 10.3758/s13414-011-0096-8

Stekelenburg, J. J., and Vroomen, J. (2012). Electrophysiological evidence for a multisensory speech-specific mode of perception. *Neuropsychologia* 50, 1425–1431. doi: 10.1016/j.neuropsychologia.2012.02.027

Stephens, J. D. W., and Holt, L. L. (2010). Learning novel artificial visual cues for use in speech identification. *J. Acoust. Soc. Am.* 128, 2138–2149. doi: 10.1121/1.3479537

Tomiak, G. R., Mullennix, J. W., and Sawusch, J. R. (1987). Integral processing of phonemes: evidence for a phonetic mode of perception. *J. Acoust. Soc. Am.* 81, 755–764. doi: 10.1121/1.394844

Trout, J. D. (2001). The biological basis of speech: what to infer from talking to the animals. *Psychol. Rev.* 108, 523–549. doi: 10.1037/0033-295X.108.3.523

Vroomen, J., and Baart, M. (2009). Phonetic recalibration only occurs in speech mode. *Cognition* 110, 254–259 doi: 10.1016/j.cognition.2008.10.015

Wade, T., and Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *J. Acoust. Soc. Am.* 118, 2618–2633. doi: 10.1121/1.2011156

Watkins, A. J., and Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral—envelope distortion. *J. Acoust. Soc. Am.* 96, 1263–1282. doi: 10.1121/1.410275

Weismer, G., and Martin, R. (1992). "Acoustic and perceptual approaches to the study of intelligibility," in *Intelligibility in Speech Disorders: Theory, Measurement and Management*, ed R. D. Kent (Amsterdam: John Benjamins), 67–118.

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Yeung, H. H., and Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition* 113, 234–243. doi: 10.1016/j.cognition.2009.08.010

Yeung, H. H., and Werker, J. F. (2013). Lip movements affect infant audiovisual speech perception. *Psychol. Sci.* 24, 603–612. doi: 10.1177/0956797612458802

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 March 2014; paper pending published: 06 April 2014; accepted: 22 April 2014; published online: 03 June 2014.

Citation: Carbonell KM and Lotto AJ (2014) Speech is not special... again. Front. Psychol. ***5****:427. doi: 10.3389/fpsyg.2014.00427*

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