Are the objects of speech perception articulatory gestures or auditory qualities?

10 April 2009

Perceptual compensation for coarticulation: Implications for theories of speech perception

Navin Viswanathan
Perception, Action & Cognition
University of Connecticut

Accounts of speech perception differ on whether vocal tract gestures (Fowler, 1986, e.g.) or the acoustic signal itself (Diehl, Lotto, and Holt, 2004) constitute the proper objects of speech perception. One extensively studied phenomenon, that has been used to distinguish these two kinds of accounts, is perceptual compensation for coarticulation. However, little agreement exists regarding the implications of evidence on perceptual compensation. In this talk, I attempt to lay out the competing viewpoints of this phenomenon and present a series of experiments with a view to dissociate them. I then consider the implications of this evidence for the accounts under consideration, as well as for speech perception in general. Finally, I outline the new questions that arise from these findings, and briefly describe the directions in which I hope to proceed with my research program.

Contrast or Compensation? Results from a masking study

John Kingston
Linguistics Department
University of Massachusetts, Amherst

The perception of a speech sound is often affected by what sound occurs next to it. For example, listeners categorize more of a series of stimuli varying incrementally from [da] to [ga] as "ga" after [al] than [ar] (Mann, 1980). Two kinds of explanations for this finding presently compete with one another. On the one hand, auditorists (Lotto and Holt, 2006) contend that the target sound contrasts auditorily with its context. In the example, [ga] differs from [da] in concentrating energy lower in its spectrum, and [al] differs from [ar] in concentrating energy higher in its spectrum. Auditory contrast causes a syllable that is intermediate between high [da] and low [ga] to sound lower and thus more like [ga] after high [al] than low [ar], much like a gray square looks darker on top of a white square than on top of a black one. On the other hand, direct realists (Fowler, 2006) propose that listeners compensate for coarticulation between the target and its context. [ga] is pronounced further back in the mouth than [da], and similarly [ar] is pronounced further back in the mouth than [al]. Listeners will perceive an intermediate syllable as being pulled forward by coarticulation when it follows [al], undo that fronting, and perceive it as the back alternative [ga] in that context. Lotto and Kluender (1998) showed that non-speech pure tone contexts that differed spectrally in the same way as [al] and [ar] could shift listeners' categorization of the [da]-[ga] series in the same way as the syllables; they responded "ga" more often after a high than a low pure tone. As listeners would not perceive the syllable as coarticulating with a preceding non-speech sound, Lotto and Kluender's finding appears to support the auditorist explanation over the direct realist one. However, Fowler, Brown, and Mann (2000) argued that the non-speech contexts were instead masking the following syllable because their rms amplitudes were as great as those of the original syllables. As support for this interpretation of Lotto and Kluender's results, they showed that the effect of a non-speech context was much weaker when the level was increased of the relevant part of the spectrum at the beginning of the following syllable. In the talk, we will argue that this masking account is implausible on a number of grounds, but the bulk of it will be a report of a recent experiment we have carried out that tests this account directly. We orthogonally manipulated three properties of the stimuli that should have affected masking: whether energy in the non-speech context – the potential masker – occupied the same parts of the spectrum as energy in the target syllable, how intense the context was relative to

the target, and how long the gap lasted between the context and target. If a non-speech context masks the following speech target, then the effect of the difference between the spectrally high and low contexts should be greater when the context has energy in the same parts of the spectrum as the target's energy, when the context is louder, and when the gap between them is shorter. We instead found that the effect is greater when the context has energy in different parts of the spectrum than the target and when the gap between context and target was longer. Only the effect of context level conformed to the expectations raised by the masking account, but we will argue that the effect of level can also be explained by the auditory contrast account.

References

Diehl, R. L., Lotto, A. J., Holt, L. L., 2004. Speech perception. Annual Review of Psychology 55, 149–179.

Fowler, C., 1986. An event approach to the study of speech perception from a direct realist perspective. Journal of Phonetics 14, 3–28.

Fowler, C., 2006. Compensation for coarticulation reflects gesture perception, not spectral contrast. Perception and Psychophysics 68, 161–177.

Fowler, C., Brown, J., Mann, V., 2000. Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. Journal of Experimental Psychology: Human Perception and Performance 26, 877–888.

Lotto, A. J., Holt, L. L., 2006. Putting phonetic context effects into context: A commentary on fowler (2006). Perception and Psychophysics 68, 178–183.

Lotto, A. J., Kluender, K. R., 1998. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. Perception and Psycholinguistics 60, 602–619.

Mann, V., 1980. Influence of preceding liquid on stop-cosonant perception. Perception and Psychophysics 28, 407–412.