TRADING RELATIONS AMONG ACOUSTIC CUES IN SPEECH PERCEPTION: SPEECH-SPECIFIC BUT NOT SPECIAL*

Bruno H. Repp

The perception of most, if not all, phonetic distinctions is sensitive to multiple acoustic cues. That is, there are several distinct aspects of the acoustic speech signal that enable listeners to distinguish between, for example, a voiced and a voiceless stop consonant, or between a fricative and an affricate. Although some cues are more important than others for a given distinction, listeners can usually be shown to be sensitive to even the less important cues when the primary cues are removed or set at ambiguous values. All cues that are relevant to a given phonetic contrast seem to carry information for listeners.

The relevance of a cue can be predicted from comparisons of typical utterances exemplifying the phonetic contrast of interest. Any acoustic property that systematically covaries with a phonetic distinction may be considered a relevant cue for that distinction and may be expected to have a perceptual effect when the conditions are appropriate.

In many recent speech perception experiments several acoustic cue dimensions have been varied simultaneously. Provided the cues are adjusted so that each has an opportunity to influence the perception of the relevant phonetic distinction, it can easily be demonstrated that a little more of one cue can be traded against a little less of another cue, without changing the phonetic percept. This is called a phonetic trading relation.

The perceptual equivalence of acoustically different stimuli obtained by trading two cue dimensions goes beyond the mere equivalence of response distributions. As several recent studies have shown, these stimuli are very difficult to tell apart in a discrimination task. Thus the trade-off among the cue dimensions takes place entirely without the listener's awareness, and only extensive auditory discrimination training might reveal the differences that exist at the auditory level.

Phonetic trading relations are a ubiquitous phenomenon. Whenever two acoustic cues contribute to the same phonetic distinction, they can also be traded against each other, within a certain range. Thus, these trading relations are a manifestation of a more general perceptual principle of cue integration, by which I mean the assumption that, in phonetic perception, the

*This position paper was presented at the Tenth International Congress of Phonetic Sciences in Utrecht, August 1983.

Acknowledgment. Preparation of this paper was supported by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]
information conveyed by a variety of acoustic cues is integrated and combined into a unitary perceptual experience that can be described in terms of linguistic categories. But what causes the various cues to be integrated and to trade with each other?

One current school of thought holds that the integration of cues and the ensuing trading relations are due to auditory interactions of one sort or another. Proponents of this hypothesis, while ready to admit that the psychoacoustics of complex speech signals are not yet well understood, nevertheless believe that phenomena known from research with nonspeech stimuli, such as auditory adaptation, masking, or integration, can account for trading relations in speech. Opponents of this hypothesis, on the other hand, like to point out the great acoustic diversity of the cues involved and their distribution over considerable temporal intervals. Obviously, and especially as far as specific trading relations are concerned, this dispute can only be settled by empirical research. A number of recent experiments have addressed this issue, employing several different techniques, but which are alluded to in my abstract. I do not have the time to summarize the results here; suffice it to say that the available evidence suggests that many phonetic trading relations occur only when listeners engage in the phonetic classification of speech signals, and not when they identify analogous nonspeech stimuli or discriminate auditory properties of speech. Thus these trading relations seem to be a product of phonetic categorization, not of interactions in the auditory system. This is not to say that auditory interactions do not occur in speech signals, although it is possible that, due to the intimate familiarity of listeners with speech, such interactions have less of a perceptual impact than in less familiar nonspeech stimuli. Certain effects of irrelevant signal properties on phonetic perception do seem to require a psychoacoustic explanation. And indeed, some of the many trading relations that now appear to be phonetic in origin may eventually be proven to rest on an auditory interaction. It seems extremely unlikely, however, that all of them will be so explained.

The reason for this prediction of mine is that psychoacoustic approaches to speech perception often seem to ignore a crucial fact—that phonetic classification takes place with reference to norms established through past experience with a language. Although this experience has been filtered and transformed by the constraints and nonlinearities in the auditory system through which it had to pass, the current input undergoes precisely the same transformations, so that the topological relationship between it and the internal representation of past experience remains essentially unchanged. It is this relationship that determines the phonetic percept by a principle of proximity: The input is perceived as whatever it resembles most in the past experience of the individual. There is, of course, much more to be learned about the perceptual metric that relates speech stimuli and the representations of phonetic categories in the listener's mind, and auditory nonlinearities may indeed influence that metric. The essential point, however, is that the perception of phonetic categories derives from a relationship, and not from any properties of the acoustic signal per se. Neither the relevance nor the perceptual importance of acoustic cues can be predicted from an inspection of the input alone. Rather, the integration and weighting of the cues is a perceptual function based on a relationship of input to knowledge within the speech domain. Phonetic trading relations are, therefore, necessarily a speech-specific phenomenon, even if certain individual trading relations could potentially (or do in fact) arise from auditory interactions. As we learn
more about the peripheral auditory transformations of speech signals, we may eventually be able to redefine the perceptual cues in a way that makes the trading relations among them exclusively phonetic.

Having argued for the speech-specificity of phonetic trading relations, I would now like to address the question of whether the perceptual integration of phonetically relevant cues is achieved by some special machinery or process, or whether it reflects a general principle of perception. In the past, it has often been argued that speech perception makes reference to speech production, and that perceptual processes actually make use of some of the neural networks engaged in articulation. This certainly remains an interesting and important hypothesis at the neurophysiological level. To the perceptual theorist, however, it really should be a truism: Since speech perception occurs with reference to internal criteria based on language experience, and since language is produced in a systematic manner by human vocal tracts, the listener's internal representation of past experience with his or her language necessarily embodies articulatory constraints as well as language-specific characteristics. In other words, I would like to argue that speech perception must reflect the way speech is produced because the criteria for perceptual classification are the production norms of the language. To say, therefore, that speech perception refers to speech production is merely to state the obvious.

A more specific hypothesis regarding phonetic trading relations might be proposed, however. It might be argued that many individual cues that trade in perception also trade in production, in the sense that there is a continuous covariation of the two acoustic cues, due to some articulatory reciprocity, even within phonetic categories. If it were the case that perceptual trading relations are obtained only for cues that show such continuous covariation in production, then it might be argued that speech perception makes use of specific knowledge of patterns of articulatory variability, and since the brain presumably cannot store an infinity of variants, it might be inferred that reference is made to an internal representation of the articulatory mechanism that enables listeners to generate specific cue relationships. Although this hypothesis needs to be explored in greater depth, it seems to me that the continuous covariation of cues in production should not be a necessary condition for perceptual trading relations to occur. All that is required is that typical instances of two different phonetic categories differ along two or more acoustic dimensions. It is much more plausible and parsimonious to assume that the listener's brain retains a record of typical instances of utterances, that is of the central tendencies in the variability encountered, rather than of the variability itself. While this system of phonetic category prototypes must be adjustable to the changing characteristics of ongoing speech, at any given point in time it provides the stable reference points that guide speech perception.

From this broad vantage point, phonetic classification is a form of pattern recognition. Speech signals may be thought of as points or traces in a multidimensional auditory space that also harbors the appropriately tuned category prototypes, and phonetic categories are selected on the basis of some distance metric. Trading relations among the various acoustic dimensions of this auditory-phonetic space are an obvious consequence. What makes speech special, in this view, is not the processes or mechanisms employed in its
Repp: Trading Relations Among Acoustic Cues

perception but the unique structure of the patterns that are to be recognized, which reflect in turn the special properties of the production apparatus and the language-specific conventions according to which it is operated.

In summary, then, I have argued that, on the one hand, phonetic trading relations are speech-specific but, on the other hand, they are not special as a phenomenon. They are speech-specific because their specific form can only be understood by examining the typical patterns of a language. They are not special because, once the prototypical patterns are known in any perceptual domain, trading relations among the stimulus dimensions follow as the inevitable product of a general pattern matching operation. Thus, speech perception is the application of general perceptual principles to very special patterns.