Journal of Phonetics (1992) **20**, 493–496

Perception of overlapping segments: thoughts on Nearey's model

D. H. Whalen

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, U.S.A.

Received 30th September 1991, and in revised form 16th June 1992

Recently, a linear logistic model for speech perception has been proposed in articles by Nearey (*Journal of Phonetics* (1990), **18**, 347–373; *Proceedings 12th International Congress of Phonetic Sciences* (1991) Vol. 1, pp. 40–49). This model is primarily segmental in nature, with allowance for "trans-segmental bias terms". Two considerations need to be emphasized with this theory. First, if segments are to be primary with trans-segments being a bias only, then these two effects should be separable in certain experiments. Second, the acoustic domain of the segment must still be defined, even if the perceptual unit is clearly delimited. The segment encounters difficulties with both of these. Even if trans-segmental terms are different in kind from segmental ones (which is not yet established), it is clear that the primary segmental cues do not display the stability implied by the linear logistic theory. Also, segments influence far more than a segment's worth of acoustic signal, indicating that only overlapping segments are defensible.

1. Introduction

In recent work, Nearey (1990, 1991) argues for the adoption of a linear logistic model of speech perception. Such models allow for bias terms as well as stimulus-tuned terms in equations for a perceptual space. Nearey finds the best model for these results is one which relies chiefly on stimulus-tuned segments augmented by "trans-segmental" bias terms. The model seems to be successful at mapping the majority judgments of subjects in several listening experiments, including two from Whalen (1989). Both the distinction between stimulus and bias terms and the definition of segment present some difficulties, which will be examined the following sections.

2. Bias terms

Bias terms are introduced into the linear logistic models to account for empirical data. For example, one of the studies of Whalen (1989) showed that when fricative judgments (/s/ versus /ʃ/) and vowel judgments (/i/ *vs.* /u/) depended on a single acoustic region that is affected by both fricatives and vowels, the two judgments are dependent on each other. That is, in this ambiguous region, an /s/ judgment occurred most often with an /u/ judgment, since the fricative noise, which was lower

0095-4470/92/040493 + 04 $08.00/0

than usual for an /s/, would make sense in the environment of a low vowel. Conversely, an/ʃ/ judgment usually occurred with an /i/ judgment, since the noise was higher than normal for an /ʃ/, as could be expected in the environment of a high vowel. These dependencies mirror the production dependencies in fricative vowel syllables.

Such a distinction between bias terms and stimulus terms leads to testable predictions. Bias terms should be subject to strategic and other cognitive influences, while stimulus terms should not. It is not clear from Nearey's (1990) discussion whether all the contextual effects due to adjacent phonemes are to be considered bias terms or not, but in any event, there are no direct tests trying to manipulate trans-segmental effects in ways that should affect biases and not perception. Thus this aspect of the theory can only be substantiated by further experimentation. Even within the given analysis, though, only models with the diphone bias terms are tested. Those with stimulus-tuned diphone factors are not tested directly, but only in relation to those models with the bias factor (Nearey, 1990, pp. 359–361). The tests that are performed do not show that such models are less successful, only that they are no more successful than ones with the bias term. Further testing within the log linear model seems warranted.

But where do these bias terms come from? They come from the experimental results themselves. The linear logistic model does not have an existing set of bias terms to implement. It simply incorporates those that are given in the experiment to be described. These bias terms are also unconstrained in the way the interactions are stated. Thus the bias term that shows that /su/ and /ʃi/ are preferred over /ʃu/ and /si/ could just have easily been the reverse. My own analysis, while not providing a numerical prediction for the size of the effect, did contain directionality: if the low frequencies were to be attributed to the vowel, then what remained for the fricative would be higher. Thus /su/ was to be expected for the low vowel classification. Similarly, if the vowel was to be heard as higher, what remained for the fricative had to be lower, thus favoring /ʃi/ judgments. Note that this formulation does not say that syllables are the perceptual unit. It only says that phonemes are perceived in the context of syllables, and all the information available to the listener is taken into account to the extent possible. If part of the signal that could be common to two phonemes is given to one, then less is available to the other. (This partitioning of the acoustic stream is also seen in competition between speech and nonspeech perception; see Whalen & Liberman, 1987; Repp, 1992, Experiments 1 and 2.) Nearey's bias terms do not have this constraint and thus may be, in his own terms, inherently too powerful.

The bias terms are incorporated into the territorial maps of Nearey's model, but so are the stimulus terms. For example, Fig. 5 of Nearey (1990) shows the majority decision for each combination of the two parameters, fricative pole frequency and F_2 frequency. However, the interdependencies in consonant and vowel judgments (Nearey's bias term) could be seen at any single combination of those parameters. That is, if one examines the responses to just one of the stimuli, one sees that the majority of the responses are /su/ or /ʃi/. There is, however, probabalistic overlap across these boundaries. Thus this figure does capture the fact that these two judgments are the main ones that alternate, if one reads the figure correctly. This map, though, represents the output of the entire linear logistic model. It is therefore not possible to determine whether any particular boundary was due primarily to

stimulus-tuned factors or to bias factors. While this does not affect the interpretation of the map for a particular experiment, it does make extrapolations difficult. So, for example, Nearey should predict that the bias terms would react differently from the stimulus-tuned terms to such manipulations as added noise, distractor tasks, contrast and adaptation. No figure can show the entirety of a theory, but these limitations should be kept in mind when reading the decision spaces.

3. Segments

Segments are presumed to be the unit of perception in Nearey's account. If the bias terms, which are primarily trans-segmental, are necessary, then it does not seem warranted to call the theory a segmental one (see also Browman & Goldstein, 1990, p. 419). Even the stimulus-tuned factors take in more than a segment's worth of the acoustic signal, so that, while they are accruing information about a segment, they go beyond the segment to do so. So the theory seems more trans-segmental than segmental. But the problems lie deeper. Nearey's definition of segments depends on several assumptions, namely, that the centers of phonemes are in the order that their percepts are, that their effects are local, that the decisions are made on "bottom-up" information, and that acoustic features extracted near the phoneme centers are "relatively invariant" (Nearey, 1991, p. 48). Although Nearey is allowing his model to modify the acoustics on the way to this relatively invariant representation, his optimism in this regard seems ill-founded. Using the examples of the stimulus changes for a segmental sysnthesis routine in Nearey's (1990) Table 2, more than half list both vowel and consonant influences (and at least two more could have). These influences often extend across the entirety of a neighboring segment's acoustic extent. Nearey attributes such "persistent" effects to extrinsic allophony rather than to coarticulation in production (1990, p. 368), which is basically a diphone model rather than a segmental one (see also Whalen, 1990). In perception, the model "does not need to be any more complex than a diphone-biased segmental model" (1990, p. 368). If the segments genuinely overlap, however, it appears that a stimulus based system will be adequate and, largely, isomorphic between production and perception. Given the benefits of such isomorphism (Liberman & Mattingly, 1985), such a model is to be preferred.

The lack of an explicit account of how a listener gets from the acoustic signal to percepts is problematic. Even if we assumed that phoneme-sized segments of the acoustic are the basis of perception, we do not know how the listener even knows how many such segments there are. Nearey's model is not unique in having this problem. Far from it: Most models do not have an explicit description of how the listener gets from an acoustic signal to a phonemic percept. Those that do, such as the invariance model of Stevens and Blumstein (e.g. 1978), have so far failed to reach human levels of performance. The gestural approach, as seen in Motor Theory (Liberman & Mattingly, 1985), Vector Analysis (Fowler & Smith, 1986), and Articulatory Phonology (Browman & Goldstein, 1989), takes the recovery of gestures as a given, but no functioning computer model has yet been made. However, such articulatory models do maintain that interactions in phonemic judgments, such as those in my experiments as discussed by Nearey, can be based on articulatory principles. If so, then the need for the bias terms in the linear logistic models would disappear because all these effects would be stimulus based. Without

the bias term—that is, with only stimulus effects to be taken care of—any description, even a linear logistic one, would be simpler. As the articulatory models become more computationally explicit, we will be able to judge the success of various approaches.

4. Summary

The linear logistic approach to phonetic description (Nearey, 1990, 1991) has the benefits of being explicit and allowing the incorporation of both stimulus terms and bias terms. However, present evidence does not unambiguously argue that bias terms are necessary at the phonetic level, and the form of the stimulus terms is perhaps too dependent on particular experimental results to be a satisfying theory of perception in general. Similarly, the assumptions adopted in Nearey (1991) to maintain the segment as the unit of perception seem to be inadequate. A great deal of evidence, including some which Nearey cites, leads to the conclusion that, if segments (and not, say, gestures) are to be the perceptual unit, they must be allowed to overlap completely.

The writing of this paper has been supported by NIH grants HD-10995 and DC-00825. An earlier version was presented at the 12th International Congress of Phonetic Sciences, Aix-en-Provence, France, August, 1991. I thank Catherine P. Browman, Terrance M. Nearey, and two anonymous reviewers for helpful comments.

References

Browman, C. P. & Goldstein, L. (1989) Articulatory gestures as phonological units, *Phonology*, **6**, 201–251.

Browman, C. P. & Goldstein, L. (1990) Representation and reality: physical systems and phonological structure, *Journal of Phonetics*, **18**, 411–424.

Fowler, C. A. & Smith, M. (1986) Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In *Invariance and variability of speech processes* (J. S. Perkell & D. H. Klatt (editor). Hillsdale, NJ: LEA.

Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised. *Cognition*, **21**, 1–36.

Nearey, T. M. (1990) The segment as a unit of speech perception. *Journal of Phonetics*, **18**, 347–373.

Nearey, T. M. (1991) Perception: Automatic and cognitive processes. In *Proceedings of the 12th international congress of phonetic sciences*, Vol. 1, pp. 40–49. Aix-en-Provence: Université de Provence.

Repp, B. H. (1992) Perceptual resotration of a "missing" speech sound: auditory induction or illusion? *Perception and Psychophysics*, **51**, 14–32.

Stevens, K. N. & Blumstein, S. E. (1978) Invariant cues for place of articulation in stop consonants, *Journal of the Acoustical Society of America*, **64**, 1358–1368.

Whalen, D. H. (1989) Vowel and consonant judgments are not independent when cued by the same information, *Perception and Psychophysics*, **46**, 284–292.

Whalen, D. H. (1990) Coarticulation is largely planned, *Journal of Phonetics*, **18**, 3–35.

Whalen, D. H. & Liberman, A. M. (1987) Speech perception takes precedence over nonspeech perception, *Science*, **237**, 169–171.