Perceptual Processing and Ecological Validity

Willy Serniclaes

Brussels' Free University

ABSTRACT

According to Nearey [1], direct effects of stimulus properties would be limited to the lowest level stage of speech perception. This raises the question to know whether the experimental stimuli display the relevant properties for being processed at higher stages of perceptual processing. The interest of ecological validity for explaining syllabic and sentence context effects is illustrated.

INTRODUCTION

In the keynote address by T.M. Nearey [1], two main topics emerge. The first is the proposal to leave aside the controversy between motor and auditory theories of speech perception in favour of a better understanding of the relationship between the two domains, via language-specific regularities in the speech wave. The second topic concerns the broad version of the "segmental filter" model, which states that the direct effects of stimulus properties are limited to the low-level stage of phone-sized segments identification and do not interfere with lexical factors in the course of top-down processing. Although I am in complete agreement with Nearey's paradigm, the strong version of the segmental filter model is, in my opinion, quite premature.

Much of the work in the field of speech communication is devoted either to the evaluation of perceptual models or to the collection of acoustic data for testing hypotheses on speech production, whereas little has been done for relating the distribution of acoustic cues in the speech wave to their perceptual processing. Yet, distributional characteristics have important perceptual implications, not only for explaining cross-linguistic differences, but also for understanding the specificity of the perceptual use of speech cues. In the ecological view developed by Brunswick [2], the description of semiregularities of the environment, or "ecological validities", is a prerequisite for testing the adaptative response of organism. Before concluding that acoustic properties cannot affect some stage of processing, we should have the insurance that the experimental stimuli display the relevant properties for being processed at that stage. In this framework, the hypothesis which states that stimulus properties do not directly affect the syllabic stage of processing can only be accepted insofar the cues under study indeed provide relevant information for syllable-sized units. In the same way, it is only when the relevant information for being processed at the lexical level is present in the stimuli that the absence of interference between sentence context and sensory input can provide an argument in favor of a modular approach of speech comprehension [3].

SYLLABIC CONTEXT EFFECTS

One should be very cautious in interpreting the failure of the stimulus-tuned diphone model to improve the fit of Whalen's data [1, 4]. While log-linear analyses clearly suggest that the acoustic cues under study do not directly affect the syllabic stage of processing of isolated CV stimuli, an important question is to know whether the cues really provide relevant information to feed the syllabic decoder. Let us first look at the acoustic cues in Wahlen's first experiment. The fact that duration and F1 frequency each depend on the identity of the vowel and of the following consonant provides a clear motivation for their use as perceptual cues both for vowel and consonant distinctions. But this does not necessarily mean that a further analysis of these cues for the classification into syllabic units is also motivated. This depends on whether the cues also convey some specific information on the syllable, irreductible to the one they give on vowels and consonants. In quantitative terms, a stimulus-based syllabic decoding is required only insofar vowel and consonant have interactive effects on the production of the acoustic cues.

In the measurements of vocalic durations in CVC syllables presented by Peterson and Lehiste ([5]; Tab.II, p.702), 28 values corresponding to 7 vowel duration and voicing contrasts (such as bead-beat-bit-bid) have been taken here for testing the additivity of vowel and consonant effects. For the logarithm of the durations, which is more appropriate for perceptual modelling, the vowel and consonant effects are largely significant (vowel: $F1,24 = 46.86$; $P < .001$; consonant: $F1,24 = 50.25$; $P < .001$) whereas the interaction is by far below the theshold of significance ($F1,24 = .962$; $P = .336$). The additivity of vowel and consonant effects is also apparent in the vocalic duration measurements in connected speech reported by Crystal and House ([6], Tab.X, p.1560). The magnitude of the effect of consonantal voicing on vocalic duration is reported here in Tab.1 for vowels preceding prepausal word-final stops, the only

not the case. The ratio between F2 frequencies , taken at 10 different intervals during the vocalic segment, can be derived from Soli's data (Figs. 3, 4, 7 8) and are presented here in Tab.2.

Table 2. Ratio between F2 frequencies (adapted from [9])

Voicing included (n=60*10)	
si/su	1.48
chi/chu	1.55
zi/zu	1.37
ji/ju	1.31
Voicing excluded (n=120*10)	
si, zi/ su, zu	1.43
chi, ji/ chu, ju	1.43

As can be seen, the magnitude of the vowel effect (i versus u) does not change systematically as a function of consonant place of articulation (s versus ch or z versus j). Although variances are not available, this suggests that vowel and consonant do not have interactive effects on F2 frequency and hence that the magnitude of the vowel effect on F2 does not depend on the consonant, and vice-versa. Just like the three other cues used by Nearey for testing the relevance of stimulus information at the syllabic stage of processing, F2 frequency, as a joint cue for vowel and fricative identity, does not convey adequate information for being processed at this stage. Taking account of the absence of ecological validation, the fact that these cues do not interfere with syllabic identification in the course of perceptual processing [1] does not allow to conclude that stimulus properties in general cannot reach the syllabic level. Other cues, if any, might exhibit an interactive relationship with two or more segments and could then provide the adequate stimuli for testing the segmental filter model.

SENTENCE CONTEXT EFFECTS

The phonetic structure of the stimulus also has implications for the general debate on the modularity of speech processing stages. Non-acoustic top-down processes are known to aid and bias speech perception. The phoneme restauration effect [10] shows that context can control the perception of phonemes. Many researchers have examined the influence of some kind of non-acoustic information on phonetic categorisation. Sentence context can bias phonetic categorisation but only when the available phonetic information is ambiguous [11, 12, 13]. The effect of lexical information on speech perception has also been investigated. It has been demonstrated that ambiguous stop consonants tend to be perceived so that the whole stimulus is a meaningful word than a non-word [14, 15]. The problem, however, is whether lexical information either biases phonetic categorisation or directly affects the mechanisms of cue integration before feature categorisation. Conine and Clifton [16] showed that prestored lexical information may be used directly in perception, contrary to semantic

position for which the effect is clearly present.

Table 1. Ratio between mean vocalic durations before voiced or voiceless consonants for long and short vowel categories (adapted from Crystal & House [6] Tab.X p.1560). N indicates the number of tokens for voiced and voiceless categories.

Vowels:	Long	Short
Obstruents	1.23	1.16
N	23 & 42	33 & 48
Stops	1.21	1.22
N	18 & 33	27 & 24
Fricatives	1.24	0.73
N	5 & 9	3 & 2

The increment of vocalic duration before voiced consonants is fairly stable, around 20 %, at the exception of short vowels before fricatives, for which the effect is reversed, probably as a consequence of the reduced sample size.

Given the absence of interaction between the vowel and consonant effects on the production of vocalic duration, the information conveyed by this cue can be entirely extracted at the low-segmental level and does therefore not require a further analysis at the syllabic level. Vocalic duration is thus not a good candidate for testing the relevance of acoustic information at a syllabic stage of processing. The other cue under study in Whalen's first experiment is the rate of F1 transition which covaries with the F1 stable frequency. Given the lack of acoustic data, we do not know whether the effects of vowel and consonant on these cues are additive or not. Notice however that vowel and consonant identification do not depend on the same aspect of F1 contour. As far as I know, voicing identification only depends on F1 transition rate (or at least on F1 initial frequency: [7, 8]) whereas vowel identification of course depends on F1 characteristics but not specifically on transition rate. F1 frequency does thus not provide the same cue for vowel and consonant and is therefore also not a good candidate for testing the relevance of acoustic information at the syllabic stage of processing. Finally, in the second example taken by Nearey and which deals with the effects of fricative pole and F2 frequency on the si-su-chi-chu distinctions (Whalen's [4] third experiment), only the latter cue clearly has a significant effect on both vowel and consonant identification. The fact that F2 frequency does not affect the syllabic stage of perceptual processing could therefore again suggest that stimulus information cannot reach this level. The question, again, is to know whether this cue conveys relevant information for being processed in a syllabic frame. The acoustic measurements of F2 in English initial fricatives presented by Soli [9] seem to indicate that this is

context which is used postperceptually. No firm conclusion can however be drawn in view of the versatility of the lexical effect on phonetic categorisation. Among the factors which contribute to the effect, stimulus structure has a special interest for the present discussion. The relevance of the phonetic structure of the stimulus is evidenced by the the fact that the lexical effect disappears when stimulus variations more nearly approximate the multiple acoustic differences between phonetic categories in natural speech [17]. This does not mean that the lexical effect is a laboratory phenomenon. Strong sentential effects have indeed be obtained for voicing perception in excerpts from spontaneous French speech [18]. This requires some further explanations. Natural voiceless French stops usually exhibit a silent interval at the vicinity of closure release whereas voiced stops almost always display continuous periodic vibrations. As other voicing cues are by far less reliable, the presence vs. absence of silence provides a major perceptual cue [19]. However, in spontaneous speech, voiceless stops may exhibit continuous periodic vibrations and are then identified as voiced when excerpted from the sentence, although they are identified as voiceless in the sentence frame. This effect, as evidenced by other aspects of the data, is clearly due to semantic factors and shows that the influence of top-down processing is not restricted to ambiguous stimuli. Even when the major acoustic cue is completely non-ambiguous, sentential context can completely modify the phonetic decision. The question which is raised is whether such a large top-down effect arises from a bias in phonetic categorisation or, more conceivably, from a direct effect of lexical information into the process of acoustic cue integration.

More importantly for my purpose here, is the fact that the magnitude of the lexical effect is seemingly due to the presence of conflicting cues in the stimulus, the major cue supporting a voiced percept whereas secondary cues support a voiceless percept. Such an internal conflict within the acoustic correlate could generate a strong appeal for extraneous evidence from lexical processes, which would in turn explain the exceptional magnitude of the lexical effect. Ecological validity, which in this case depends on the presence of contradictions within the acoustic correlate, would then again have decisive importance for testing perceptual models.

REFERENCES

[1] NEAREY, T.M. (1991) "Perception: Automatic and Cognitive Processes" Keynote address in this volume.

[2] BRUNSWICK,E. (1957) Scope and aspects of the cognitive problem. In Contemporary Approaches to Cognition. Cambridge, MA: Harvard U. Press, 5-40.

[3] FODOR, J.A. (1983) "The modularity of mind" Cambridge, MA: MIT Press.

[4] WHALEN, D. (1989) "Vowel and consonant judgments are not independent when cued by the same information" Percep. and Psychophys., 46, 284-292.

[5] PETERSON,G.E. and LEHISTE,I. (1960) "Duration of syllable nuclei in English" J.Acoust.Soc.Am. 32, 693-703.

[6] CRYSTAL,T.H. and HOUSE,A.S. (1988) "The duration of American English vowels: an overview" J. of Phonetics, 16, 263-284.

[7] LISKER,L. (1975) "Is it VOT or a first formant transition detector?" J.Acoust.Soc.Am. 57, 1547-1551.

[8] SUMMERFIELD,Q., and HAGGARD,M. (1977) "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants" J.Acoust.Soc.Am. 62, 435-448.

[9] SOLI,S.D. (1981) "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation" J.Acoust.Soc.Am., 70, 976-984.

[10] WARREN,R.M. (1970) "Perceptual restauration of missing speech sounds" Science 167, 392-393.

[11] GARNES,S. and BOND,Z.S. (1976) "The relationship between semantic expectation and acoustic information" Phonologica, 3, 285-293.

[12] MILLER,J.L., GREEN,K., and SCHERMER,T. (1984) "A distinction between effects of sentential speaking rate and semantic congruity on word identification" Percept. and Psychophys., 36, 329-337.

[13] CONNINE,C. (1987) "Constraints on interactive processes in auditory word recognition: The role of sentence context" J. of Memory and Language, 26, 527-538.

[16] GANONG, W.F. (1980) "Phonetic categorization in auditory word perception" J. of Exp. Psychology: Human Perc. & Perform., 6, 110-125.

[15] FOX,R.A. (1984) "Effect of lexical status on phonetic categorization" J. of Exp. Psychology: Human Perc. & Perform., 10, 526-540.

[16] CONNINE,C., and CLIFTON,C.(1987) "Interactive use of lexical information in speech perception" J. of Exp. Psychology: Human Perc. & Perform., 13, 291-299.

[17] BURTON,M.W., BAUM,S.R., and BLUMSTEIN,S.E. (1989) "Lexical effects on the phonetic categorization of speech: The role of acoustic structure" J. of Exp. Psychology: Human Perc. & Perform., 15, 567-575.

[18] SAERENS, M., SERNICLAES, W., and BEECKMANS, R. (1989). "Acoustic versus contextual factors in stop voicing perception in spontaneous speech" Language and Speech 32, 1989, 291-314.

[19] SERNICLAES, W. (1987) Etude expérimentale de la perception du trait de voisement des occlusives du français. Thèse de doctorat en Sciences psychologiques, ULB, Octobre 1987.