or must be used because of the difficulty, or impossibility, to find appropriate meaningful sequences, thus an even greater care must be taken to monitor prosodic features, since nonsense utterances are less natural for the subject(s) to say than meaningful ones: consequently, a subject is unlikely to be as consistent in his production as in the case of natural utterances.

Wayne A. Wickelgren How do we recognize the order of the phonemes in a word? It will not suffice to recognize only the unordered set of phonemes, else we could not distinguish "cat" from "tack" from "act". Many considerations weigh against the hypothesis that we somehow time label the set of features occurring at times t, $t+\Delta$, $t+2\Delta$, . . . with some ordered set of labels, among which is the evidence of Warren (Warren 1974a, 1974b). Accordingly, I have proposed a different theory of the coding of the serial order of phonemes in words, namely, that the segmental units of words are overlapping phoneme triples or context-sensitive allophones (Wickelgren, 1969a, 1969b, 1972, 1976). For example, "cat" would have as its segmental constituents, $\#\ {}^{k}a,\ k^{a}t,\ a^{t}\ \#$. Without any need to time label these segmental nodes, and regardless of the temporal order of activation of such nodes the necessary order information for distinguishing "cat" from "tack" and "act" is contained in the *unordered set* of these overlapping-triple nodes.

Context-sensitive segmental coding explains many important phenomena in both speech recognition and articulation as I have discussed in the papers referred to. Among these are how we perceive segmental order, the difficulty of segmenting the speech stream into phonemic units within a syllable, the context-conditioned variation in acoustic cues for phonemes, how contextual feedback can be used in recognition without disrupting the encoding of order information, intentional (mentally directed) coarticulation, advance priming of long sequences of segments in articulation while retaining the segmental order information, the functions of accent and syllable juncture, a variety of speech error phenomena, and others.

Context-sensitive coding can be extended to the feature level to achieve Gestalt-like grouping of features occurring in temporal proximity. During any slice of time ($\approx$ 10 msec) assume an encoding of the speech signal by strength values from 0 to 1 on each of f atomic (context-free) feature dimensions, such as the spectral frequencies of a Fourier analysis. Now define the context-sensitive features to be the set of all pairs of such atomic features within the same Δ time window and for i Δ-windows before and after. If there are f atomic feature dimensions, there will be $\frac{1}{2}f(f-1)$ simultaneous unordered-pair feature dimensions (within the same Δ-window) $if f^2$ forward-successive ordered-pair features, and $if f^2$ backward-successive ordered-pair features. If i = 1, then only adjacent time windows are chunked into coding feature transitions (such as frequency transitions). If $i \geq 2$, then a variety of more extensive feature transitions are directly encoded. If i = 10 and f = 1000 (e.g., 1000 different frequency dimensions), then the total number of context-sensitive feature dimensions would be about 2.5×10^7. That's a very modest number in relation to 10^{10} neurons in the brain, so I think we could safely assume at least this number of feature dimensions for the analysis of auditory signals including speech.

A suitable mathematical definition of the strength of a pair feature s(x,y) in terms of the strength of its two atomic features s(x) and s(y) might be the fuzzy-logic multiplicative rule, $s(x,y) = s(x) \cdot s(y)$. This gives a pair feature high strength only if both component features have high strength. A strong, but unproven, conjecture is that one only needs to consider feature pairs, not triples or higher-order combinations and permutations, to discriminatively activate the correct context-sensitive segmental units.