Signed and spoken utterances have at least two aspects that are of interest to a perceiver. First of all, they have a physical aspect, the significance of which is given in the lawful relations among utterances, the information-bearing media structured by them, and the perceptual systems of observers and listeners. Secondly, they have a linguistic aspect, the significance of which is given in the conventional or ruleful relations between forms and meaning. In part because our time was limited, and in part because so little work has been done on the conventional significance of events (as opposed to the intrinsic significance [cf. Gibson, 1966]), our work group chose to focus on the physical aspect. Nevertheless, it will be seen that we did have a speculative word or two to say about the origins of some linguistic conventions, and we would draw attention to the report of the Event/Cognition group, as well as to Verbrugge's remarks (discussant for the address by Studdert-Kennedy), for more elaborate treatments of this important topic.

Roughly, our daily discussions centered around five topic areas: (1) useful descriptions of signed and spoken events; (2) natural constraints on linguistic form; (3) the origins of some linguistic conventions; (4) the ecology of conversation; and (5) conducting language research from an event perspective. Our review of these topics will highlight what seemed to us to be the obvious applications of the event approach and also its apparent limitations.

USEFUL DESCRIPTIONS OF SIGNED AND SPOKEN EVENTS

We considered the minimal linguistic event to be an utterance, and identified as such anything that a talker (signer) might choose to say (sign). Obviously, this definition is unsatisfactory on a number of grounds; however, it does identify the minimal event of interest as being articulatory (gestural) in origin, and rejects as irrelevant those properties of articulation
(gesture) that are not intended to have linguistic significance. We first attempted to verify that utterances have the "nested" character of other ecological events and that the nestings are perceived; next we considered how to discover the most useful characterization of utterances for the investigators' purposes of studying them as perceived events.

**Signing and Speaking as Nested Events**

Natural events are nested in the sense that relatively slower, longer-term or more global events are composed of relatively faster, shorter-term or more local ones. For example, a football game is a longer-term event composed of shorter-term plays. It is clear from research—particularly Johansson's (e.g., 1973, 1975) on the perception of form and motion in point-light displays—that viewers are sensitive to the nested structure of events. In his address to this conference, Johansson described an example of light points placed on a rolling wheel. When a single point is affixed to the rim, a viewer who sees only that point gets no sense of the wheel's motion; instead, the percept is of a light moving in a cycloid pattern. However, when a second light is attached, now to the hub of the wheel, the viewer perceives rolling instead of the cycloid motion. Thus, two appropriately placed lights provide sufficient optical information to specify the distal event of rolling.

In geometric terms, rolling involves two kinds of motion: translatory and rotary. These are temporally nested; a series of rotations occurs as the wheel translates over the ground plane. The translatory component affects the behavior of both light points (since both are attached to the translating wheel), but only the point on the rim is affected by the rotary component as well (since it rotates about the point on the hub). Apparently, perceptual sensitivity to the translation (as specified by the correlated activity of the two lights) forms a sort of "backdrop" for detection of the rotation; in essence, the translational component is "factored out" of the cycloid movement of the rim light, thereby revealing its rotational component.

Now let us consider whether these observations apply to signing and its perception. In American Sign Language (ASL), signs are specified by three properties: the shape of the hand or hands, the place of articulation of the sign within a signing space, and the movement of the hand or hands. Signs can be inflected by modulating the movement. For example, a 'distributional' inflection indicating that all of the individuals under discussion are affected by some act is produced by sweeping the arm through the central body plane. By signing, say, GIVE while making such an arm sweep the signer communicates GIVE TO ALL OF THEM. Likewise, a 'temporal' inflection, one indicating the repeated occurrence of an act, is produced by rotating the wrist about a body-centralized point; with this gesture, GIVE is modified to mean GIVE AGAIN AND AGAIN.

Finally, and most importantly for the current discussion, several inflections can be superimposed. Carrying our previous example a step further, it proves possible to sign the complexly inflected verb GIVE TO ALL OF THEM AGAIN AND AGAIN. This is accomplished by rotating the wrist while the arm sweeps through its arc. Notice that when this is done, the optical information for the 'temporal' inflection undergoes a radical transformation; the wrist no
longer rotates about a single point fixed at the center of the body, but rather about a point moving with the sweeping arm. It appears that observers treat the sweeping motion (common to all points of the hand, wrist, and arm) as both specifying one signed event (the 'distributional' inflection), and as providing a moving frame of reference for the interpretation of the nested 'temporal' inflection.

Spoken language, with its syntactic units—phonological segments, morphemes, words, and syntactic phrases—and its metrical units—syllables, feet, phonological phrases (see Selkirk, 1980)—lends itself readily to the characterization "nested." We will take an example of nested articulatory and perceived events from a relatively low-level phenomenon, coarticulation. In fluent speech, the productions of successive phonetic segments overlap such that the articulatory gestures often satisfy requirements for two or more segments at the same time. Typically, for example, unstressed vowels coarticulate with the stressed vowels of adjacent syllables. It is therefore tempting to think of the production of the unstressed vowels as being nested within that of their stressed counterparts, and to think of unstressed vowels as being perceived relative to their stressed-vowel context. This way of thinking is promoted by findings (Fowler, 1981) that under some conditions listeners behave as if they have "factored out" the articulatory/acoustic contributions of the context when judging the quality of unstressed vowels—more or less as Johansson's subjects seem to have factored out common and relative motions in an optical display.

In trisyllabic nonsense words with medial /ə/, the medial vowel coarticulates with both of its flanking stressed vowels such that the F2 of /ə/ in, for instance, /ibəbi/ is higher than it is in /ubəbu/. (Compatibly, F2 is high for /i/ and low for /u/.) When extracted from their contexts, the medial /be/ syllables do sound quite different, but when presented in context they sound alike—more alike, in fact, than do two acoustically identical /bə/ syllables presented in different contexts.

A nested-events account of these data would hold that when the /be/ syllables are extracted from the context in which they had been produced, the perceiver has no way to detect (factor out) the contribution that the stressed vowels have made to that portion of the acoustic signal in which /ə/ correlates predominate over the correlates of other segments—no more than Johansson's subjects can separate the rotary from the translatory components of movements when they see just the one light on the rim of a wheel. Presentation in the context of flanking vowels, on the other hand, allows the perceiver to factor out components in common with those vowels, and to recognize the quality of what is left. This leads to the perceived identity of the acoustically "different" /bə/ syllables (in the /ibəbi/ and /ubəbu/ contexts), and to the perceived difference of the acoustically "identical" syllables (in the different trisyllable contexts).

Identifying Speech Events: The Problem of Description

Several theories of speech perception—including Gibson's (1966) and one more familiar to speech investigators, the motor theory (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967)—adopt a view consistent with an event perspective: namely, that the perceived categories of speech are
articulatory in origin. Gibson's view is distinguished from the other by its working assumption that the perceived articulatory categories are fully reflected (however complexly) in the acoustic signal and hence need not be reconstructed by articulatory simulations. What are the reasons for this major disagreement among theorists who agree on the question of what is perceived? One reason may be that they differ in terms of how they describe the acoustic signal or even the articulatory event.

In speech, articulatory activities and their acoustic correlates are both richly structured, and consequently can be described in a great many different ways. Each of the various descriptions may be most appropriate for certain purposes, but none is privileged for all purposes, and just one or a few are privileged for the purposes of understanding what a talker is doing and why a listener perceives what he or she perceives. A theorist who is convinced that the acoustic support for perceptual categories is inadequate may be correct; but, alternatively, she or he may have selected a description of articulatory events and their acoustic correlates that fails to reveal the support.

There are many reasons why a particular description might be inappropriate for aiding our understanding of speech perception and production. It could specify excessive detail (as when, in Putnam's [1973] example, information about the positions and velocities of the elementary particles of a peg and pegboard are invoked to explain why a square peg won't fit in a round hole). Or, for any level of detail, it could be inappropriate because it classifies components in ways that fail to capture the talker's organization of them or the listener's perceived organization. Appropriate descriptions of vocal activity during speech, then, must capture the organization imposed by the talker; those of the acoustic signal must capture those acoustic reflections of the articulatory organization that are responsible for the listener's perception of it.

Appropriate descriptions of perceived articulatory categories. In some time frame, a talker might be said to have raised his larynx (thereby decreasing the volume of the oral cavity), abducted the vocal folds, increased their stiffness, closed the lips, and raised the body of the tongue toward the palate. This description lists a set of apparently separate articulatory acts. In fact, however, the first three of them have the joint effect of achieving voicelessness; these and the next, lip closure, are the principal components of /p/ articulation; and all five acts together are essential to the production of the syllable /pi/. Thus, the aggregate of occurrences in this time frame have a coordinated structure of relations something like the following: \{[(larynx raising, vocal cord abduction, vocal cord stiffening)(lip closure)](tongue-body gesture)\}.

If an investigator settles for the first description—a list of the activities of individual articulators—then, from his perspective, information about the phonetic segments of an utterance is already absent and he cannot expect to find any evidence of it in the acoustic signal. Consequently, when a perceiver recovers segments in speech, the recovery must be considered reconstructive. Before settling for this conclusion, however, the investigator can try standing back a little from his first perspective on the vocal tract activity and looking for organizations among gestures that were not initially apparent. These organizations will only be revealed from a temporal
perspective broad enough that coupled changes among the coordinated structures can be observed. Certainly if there are coordinative articulatory relations among gestures and if the relations have acoustic reflections, then the listener is likely to be sensitive to the coordinated structure, rather than to the unstructured list of gestures from which it is built, for by detecting the structure of the relations among these gestures, the listener detects the talker's structure—here the featural and phonetic segmental structure of the utterance—which is what she or he must do if the utterance is to be understood.

In support of this general approach to phonetic perception, there is some evidence that listeners do perceive aggregates of articulatory acts as if those acts were coordinated segmental structures. One example of this involves the perception of voicing. Following the release of a voiceless stop consonant, the fundamental frequency \( f_0 \) of the voice is relatively high and falls (Halle & Stevens, 1971; Hombert, 1978; Ohala, 1979). Following a voiced stop, \( f_0 \) is low and rises. Although the reasons for this differential patterning of \( f_0 \) are not fully understood (Hombert, 1978; Hombert, Ohala, & Ewan, 1979), it is generally agreed that it results from the timing of certain laryngeal adjustments and from certain aerodynamic conditions that the talker establishes in maintaining voicelessness or voicing during the production of the consonant (cf. Abramson & Lisker, 1965). That is, the talker does not plan to produce a high falling \( f_0 \) contour following release of a \(/p/\). Instead, he plans to maintain voicelessness of the consonant and an unintended consequence of that effort is a pitch perturbation following release. Compatibly, listeners do not normally hear this pitch difference as such (that is, they do not notice a higher pitched vowel following \(/p/\) than \(/b/\). Instead, in the context of a preceding stop, a high falling \( f_0 \) contour in a vowel may serve as information for voicelessness of a preceding consonant (Haggard, Ambler, & Callow, 1970; Fujimura, 1971), even though, when removed from the consonantal context, the \( f_0 \) contours are perceived as pitch changes (Hombert, 1978; Hombert et al., 1979).

Also suggestive of the perceptual extraction of coordinated articulatory structures are occasions when the perceiver seems to be misled. Ohala (1974, in press) believes that certain historical sound changes can be explained as results of listeners' having failed to recognize some unplanned articulatory consequence as unplanned. An example related to the first one is the development of distinctive tones in certain languages. These languages evolved from earlier versions without tone systems, but with distinctions in voicing between pairs of consonants. Over time, the \( f_0 \) difference just described between syllables differing in initial stop voicing became exaggerated and the voicing distinction was lost. Ohala's interpretation of the source of the change is that in these languages listeners tended to hear the \( f_0 \) differences on the post-consonantal vowels as if pitch had been a controlled articulatory variable, rather than an uncontrolled consequence of adjustments related to voicing. Therefore, when these individuals produced the vowels, they generated controlled (and larger) differences in \( f_0 \) of voiced and voiceless stop-initial syllables. Eventually, because the \( f_0 \) differences had become highly distinctive, the now redundant voicing distinction was lost and the words that formerly had differed in voicing of the initial consonant now differed in tone. According to Ohala, this process occurred during the separation of Punjabi from Hindi.
Appropriate descriptions of the acoustic signal. Because very little is known about how a talker organizes articulation, descriptions of the acoustic signal useful for purposes of understanding perception cannot be guided strongly by information about articulatory categories. However, we do know enough to recognize that the usual method of partitioning the acoustic signal into segments or into "cues" can be improved on. Such partitioning often obscures the existence of information for the phonetic segmental structure of speech because the structure of measured acoustic segments is not coextensive with the phonetic structure of the utterance. For one thing, phonetic segments as produced have a time course that measured acoustic segments do not reflect. The component articulatory gestures of a phonetic segment gradually increase in relative prominence over the residual gestures for a preceding segment and consequently the acoustic signal gradually comes to reflect the articulatory character of the new segment more strongly than that of the old one. Thus, phonetic segments are not discrete on the time axis, although they can be identified as mutually separate and serially ordered by tracking the waxing and waning of their predominance in the acoustic signal (cf. Fant, 1960).

Acoustic segments, on the other hand, are discrete. (Such segments are stretches of the acoustic signal bounded by abrupt changes in spectral composition.) An individual acoustic segment spans far less than all of the acoustic correlates of a phonetic segment and, in general, it reflects the overlapping production of several phonetic segments (cf. Fant, 1960). Looking at the signal as a series of discrete acoustic segments, then, obscures another way of looking at it: as a reflection of a series of overlapping phonetic segments successively increasing and declining in prominence.

Partitioning acoustic signals into acoustic segments also promotes assigning separate status to different acoustic "cues" for a phonetic feature, even though such an assignment tends to violate the articulatory fact that many of these cues, no matter how distinct their acoustic properties may be, are inseparable acoustic products of the gestures for a single phonetic segment (Lisker & Abramson, 1964; Abramson & Lisker, 1965). The findings of "trading relations" among acoustically distinctive parts of the speech signal indicate that these cues are not separable for perceivers any more than they can be for talkers. For example, certain pairs of syllables differing on two distinct acoustic dimensions—the duration of a silent interval following frication noise and the presence or absence of formant transitions into the following vocalic segment—are indistinguishable by listeners (Fitch, Halwes, Erickson, & Liberman, 1980). Within limits, a syllable with a long silent interval and no transitions sounds the same as one with a short silent interval and transitions. It is as if the transitions in the second syllable are indistinguishable from the extra silence in the first. A perceptual theory in which this observation is natural and expected is difficult to imagine—unless the theory recognizes that detecting acoustic segments per se is not all there is to perceiving speech. We would argue that the cues in these stimuli are indistinguishable to the degree that they provide information about the same articulatory event. Thus, 24 msec of silence "trades" with the formant transitions because both cues specify production of /p/. It is our view that source-free descriptions of acoustics will never succeed in capturing what a speech event sounds like to a perceiver, because it is information carried in the signal, not the signal itself, that sounds like something.
NATURAL CONSTRAINTS ON LANGUAGE FORM

Shifting perspectives from ongoing articulation and its reflections in proximal stimulation, we considered how, over the long term, properties of the articulators in speech or of the limbs in sign may have shaped linguistic forms. Similarly, we considered how perceptual systems and acoustic or optical media, with their differential tendencies to be structured by various properties of distal events, may have shaped the forms of sign and speech.

Sign has several regular properties suggestive of natural constraints on manual-language forms. One (Battison, 1974; cited in Siple, 1978, and Klima & Bellugi, 1979) takes the form of a symmetry constraint on two-handed signs: if both hands move in the production of a sign, the shapes and movements of the two hands must be the same and symmetrical. This constraint is compatible with anecdotal evidence (from novice piano players, for example), and more recently with experimental evidence (Kelso, Southard, & Goodman, 1979; Kelso, Holt, Rubin, & Kugler, in press) that it is difficult to engage in different activities with the two hands. One reason for this may be a tendency for actors to reduce the number of independently controlled degrees of freedom in complex tasks by organizing structures coordinatively (e.g., Turvey, 1977). Kelso's experiments suggest that the two arms and hands tend to be organized coordinatively even when such an organization would seem unnecessary or even undesirable (Kelso et al., 1979; Kelso et al., in press); when subjects were required to engage in different activities with the two hands or arms, the "different" movements tended to retain similar properties.

A second constraint, called the "Dominance" constraint by Battison, may have a similar origin in general constraints on movement organization. For signs in which just one hand moves and the other hand serves as a base for the movements (a place of articulation), the base hand must either have the same configuration as the moving hand or one of a very limited set of other configurations.

An example of a constraint in spoken languages may be the tendency for syllable structures to respect a "sonority hierarchy" (e.g., Kiparsky, 1979) whereby sonority (roughly, vowel-likeness) increases inward toward the vowel from both syllable edges. Hence, for example, /tr/, a sequence in which sonority increases from left to right, is an acceptable prevocalic sequence, but postvocically the order must be /rt/.

As for language features owing to properties of perceptual systems and stimulating media, Lindblom's proposed constraints on the evolution of vowel systems provide an example in spoken languages (1980; see also, Bladon & Lindblom, 1981). Lindblom has proposed that vowel systems maximize the perceptual distances among member vowels. Based on estimates of distances among vowels in perceptual space, he succeeds in predicting which vowels will tend to occur across languages in vowel systems of various sizes. This implies a constraint on phonological inventories that perceivers be able to recover distinct phonetic segments when distinct ones are intended. Talkers cannot elect to realize distinct phonetic segments by using articulatory gestures (however distinct they may be themselves) that fail to leave
distinguishing traces in the acoustic medium or in the neural medium of perceptual systems. (Analogous articulatory constraints also operate to shape vowel systems. Thus, the relatively densely populated front vowel space and the sparsely populated back vowel space doubtless reflect the relatively greater agility and precision of movement of the tongue tip and blade compared with the tongue body.)

Lane proposed that similar perceptual and articulatory constraints may shape the evolution of sign inventories. Facial expressions provide information in ASL and perceivers tend to focus on a signer's face. This creates a gradient of acuity peaking at the face. According to Siple (1978), signs made well away from the face tend to be less similar one to the other than signs made in its vicinity; in addition, two handed signs made in the periphery are subject to the Symmetry and Dominance constraints just described, which provide redundancy for the viewer who may not see them as clearly as signs produced near the face. Lane suggested that the relative frequency of signs in various locations in signing space might be predicted jointly by the acuity gradient favoring signs located near the face and a work-minimizing constraint favoring signs closer to waist level.

THE ORIGINS OF SOME LINGUISTIC CONVENTIONS

As we noted earlier, the conventional rather than necessary relationship between linguistic forms and their message function is central to the nature of language, freeing linguistic messages from having to refer to the here and now, and thereby allowing past, future, fictional and hypothetical events all to be discussed. For Gibson, this property of language removes it from the class of things that can be directly perceived:3

[Perceptual cognition] is a direct response to things based on stimulus information; [symbolic cognition] is an indirect response to things based on stimulus sources produced by another human individual. The information in the latter is coded; in the former case it cannot properly be called that (1966, p. 91).

The study group did not discuss language comprehension in relation to event theory, perhaps because event theory currently offers little guidance on that subject. However, there was discussion of the origins of some linguistic conventions. Several examples suggest an origin of certain conventional relations as elaborations of intrinsic ones. The example of tonogenesis given earlier illustrates this idea. Ohala proposes that in some languages distinctive tones originated as controlled exaggerations of the pitch perturbations on vowels caused by the voicing or voicelessness of a preceding consonant.

A second example is so-called "compensatory lengthening" (e.g., Grundt, 1976; Ingria, 1979)—a historical change whereby languages concurrently lost a final consonant in some words and gained a phonological distinction of vowel length, with the words that formerly had ended in a consonant now ending in a phonologically long vowel. In spoken languages, the measured length of vowels shortens when they are spoken before consonants (e.g. Lindblom, Lyberg, & Holmgren, 1981). Of course, since vowels coarticulate with final consonants, this measured shortening may not reflect "true" shortening; presumably, acoustic evidence of their coarticulating edges is obscured by acoustic
correlates of the overlaid consonant. In any case, the loss of a final consonant leads to measured lengthening of the vowel. If that unintended lengthening was perceived as controlled lengthening (just as, hypothetically, uncontrolled pitch perturbations were perceived as controlled pitch contours), and was subsequently produced as a controlled lengthening, it could serve as the basis for a phonological distinction in vowel length.

A final example in speech apparently has an analogue in sign. Some speech production investigators have proposed that vowels and consonants are produced by relatively separate articulatory organizations in the vocal tract, and that vowel production may go on essentially continuously during speech production, uninterrupted by concurrently produced consonants (e.g., Ohman, 1966; Perkell, 1969; Fowler, 1980). These proposals are based on observations that vowel-to-vowel gestures that occur during consonant production (Ohman, 1966; Perkell, 1969) sometimes look very similar to vowel-to-vowel gestures in VV sequences (Kent & Moll, 1972). Also, a relatively separate organization of vowel and consonant production with continuous production of vowels may promote such linguistic conventions as vowel infixing in consonantal roots in Arabic languages (McCarthy, 1981) and vowel harmony in languages including Turkish (and in infant babbling [e.g., Menn, 1980]).

Vowel infixing will provide an illustration. In Arabic languages, verb roots are triconsonantal. For example, the root 'ktb' means "write." Verb voice and aspect (e.g., active/passive, perfective/imperfective) are indicated by morphemes consisting entirely of vowels. In McCarthy's recent analysis (1981), the consonantal roots and vowel morphemes are interleaved according to specifications of a limited number of word templates and a small number of principles for assigning the component segments to the templates. Some derivationally related words in Arabic are: katab, ktabab, kutib, and kuutib. The consonantal root in each case is 'ktb'; the vowel morphemes are 'a' (perfective, active) and 'ui' (perfective, passive); and the relevant word templates are CVCVC, CCVCVC, CVVCVC (where C is a consonant and V is a vowel). The general rules for assigning roots and morphemes to templates are (1) to assign the component segments left to right in the template, and (2) if there are more C slots than consonants or more V slots than vowels, to spread the last consonant or last vowel over the remaining C or V slots. The only exception to this generalization is 'i' in 'ui', which is always assigned to the right-most V in the template. Below are two illustrations of verb formation according to this analysis:

\[
\begin{align*}
&\text{CCVCVC} \rightarrow k\text{t}a\text{b}a\text{b} \\
&k\text{t} \rightarrow k\text{t}b
\end{align*}
\]

Kegl discussed an analogous system in ASL. A particular root morpheme can be associated with different sign templates to express derivationally or inflectionally related versions of the morpheme. The templates have slots for locations (L) and movements (M), where the former specify person and number and the latter specify aspect. To take an example, the template that underlies I GIVE TO HIM is (LM + ML). Movements and locations are assigned to it as in McCarthy's analysis:
A template can include several L's and M's—more, in fact, than there are distinct movements in a root morpheme. In this case, the movements of the root morpheme are assigned left to right in the template until they are exhausted, and then the right-most movement spreads to fill the empty M slots. In I GIVE TO X, Y, AND Z the template and assignments of root morpheme movements are as follows:

Analyzed this way, the meshing of movements and locations is similar to the meshing of vowels and consonants in languages with infixing and vowel harmony systems. This leads to the question of whether the system is favored as a linguistic device, and, if so, whether it is favored by virtue of the signer's motor organization for producing it. It might be favored, for example, if the motor organization underlying sign production readily produced cyclic repetitions of a movement (as those underlying stepping, breathing, chewing and perhaps vowel production do), and if minimal adjustments to the organization would enable shifts in location without changing the form of the movement.

THE ECOLOGY OF CONVERSATION

A scan of the various conference addresses shows the close ties between the event approach and Gibson's ecological theory of perception. Indeed, Gibson's radical rethinking of classic perceptual problems includes the notion that a perceiver does not operate in a series of "frozen moments," but rather in an ongoing stream of events. We therefore thought it useful to examine the ecology of the speech event, and in doing so we were reminded that both the speaker and the listener (the signer and the observer) have a stake in the success of a communicative episode. This is a rather unique circumstance; it invites both a familiar analysis of the perceiver as an active seeker of information (cf. Gibson, 1966), and a less familiar analysis of the producer as an active provider of informational support.

As to the perceiver's active role, we first of all see behavior intended to enhance signal detection: the head can be rotated to an optimal orientation, the source can be approached, and so on. Beyond this there can be direct communicative intervention; that is, the perceiver can make requests for repetition or clarification. On the producer's part, there are the well-known redundancies of language; in essence, more than enough information is provided to ensure the accuracy of communication. Also, perhaps to avoid syntactic ambiguities, the talker may provide careful prosodic marking for clause boundaries and the like (e.g., Cooper & Paccia-Cooper, 1980). And
finally, a talker will enunciate more clearly (and a signer gesture more distinctly) when there is a great distance to the perceiver or when the message context makes a particular word unpredictable.

CONDUCTING LANGUAGE RESEARCH FROM AN EVENT PERSPECTIVE

If there is a theme to the event conference, it is surely that psychologists have paid too little attention to the systematic (and potentially informative) nature of change. With respect to speech, this can be seen in the common practice of decomposing the speech stream into a succession of discrete acoustic segments (e.g., release bursts, aspiration, formant transitions, and the like). A whole literature speaks, in turn, of the difficulty in bringing these acoustic segments into some correspondence with linguistic segments. In the case of sign, the perceptual significance of change was overlooked in early attempts to devise sign glossaries: investigators were preoccupied with cataloguing the featural properties of hand shapes and failed at first to recognize the importance of the gestures being made with the hands (Klima & Bellugi, 1979, chapter 12 and passim; Bellugi & Studdert-Kennedy, 1980).

The members of our group were agreed that a shift of emphasis is needed: investigators of both speech and sign should give greater consideration to the time-varying properties of those events. To begin with, this will involve focusing on the dynamics of the source events themselves. These investigations of the source can suggest compatible and appropriate perceptual analyses. Some recent work using Johansson's point-light techniques to study the coordinated activities of the signer, and the perception of lexical movements and inflections (e.g., Poizner, Bellugi, & Lutes-Driscoll, 1981), seems to offer promising beginnings for such an approach.

Alternatively, analyses of time-varying properties of the signal may provide guidance in understanding the ways in which talkers and signers structure articulatory activity (cf. Fowler, 1979; Tuller & Fowler, 1980). On this issue, our group spent a good deal of time considering the recent work of Remez, Rubin, Pisoni, and Carrell (1981; Remez, Rubin, & Carrell, 1981). They have shown that the phonetic message of an utterance can be preserved in sinewave approximations that reproduce only the center frequencies of its first three formants. These stimuli have no short-time acoustic constituents that vocal tracts can produce and consequently lack many acoustic elements heretofore identified by investigators as speech cues. Presumably the stimuli are intelligible because information is provided by relations among the three sinusoids, information that the sinusoidal variations are compatible with a vocal origin.

These findings are important not because they show short-time acoustic cues to be unimportant to speech perception. After all, naive listeners did not spontaneously hear the sinewaves as phonetic events. Instead, the findings are important in showing that time-varying properties of the signal can provide sufficient information for word and segment identification in speech. In this respect, as Remez and Rubin point out (Note 1), their demonstration is closely analogous to Johansson's demonstrations with point-light displays of moving figures. In both demonstrations, change provides essential information for form.
The conclusion we draw from all of the examples considered here is that students of language should not be misled by the timeless quality of linguistic forms. Signing and speaking are coherent activities and natural classes of events. It is only reasonable to expect that the signatures of these events will be written in time as well as space.

REFERENCE NOTE


REFERENCES


Hombert, J.-M., Ohala, J., & Ewan, W. Phonetic explanation for the develop-


Ohala, J. The listener as a source of sound change. In M. F. Miller (Ed.), *Papers from the parasession on language and behavior*. Chicago: Chicago Linguistic Society, in press.


Society of America, 1981, 69, S114. (Abstract)


FOOTNOTES

1 We do not intend to suggest by the word conventional that the linguistic aspects of utterances have been established by popular acclaim. We intend only to distinguish the linguistic aspects from the physical aspects in terms of their "relative arbitrariness." Let's consider a physical example first: the articulatory and acoustic differences between the versions of /d/ in /di/ and /du/ are necessary and lawful, given the nature of vocal tracts. This contrasts with the aspiration difference between the versions of /p/ in "pie" and "spy," the production of which is required of English speakers only by convention or rule. We know this to be the case since speakers of other languages (e.g., French) make no such distinction.

2 In Gibson's view:

The relation of a perceptual stimulus to its causal source in the environment is of one sort; the relation of a symbol to its referent is of another sort. The former depends on the laws of physics and biology. The latter depends on a linguistic community, which is a unique invention of the human species. The relation of perceptual stimuli to their sources is an intrinsic relation such as one of projection, but the relation of symbols to their referents is an extrinsic one of social agreement. The conventions of symbolic speech must be learned, but the child can just about as easily learn one language as another. The connections between stimuli and their sources may well be learned in part, but they make only one language, or better, they do not make a language at all. The language code is cultural, traditional and arbitrary; the connection between stimuli and sources is not (p. 91).

3 It is interesting in this regard that theories of perception developed within the information-processing framework have relied almost exclusively on verbal materials as stimuli and propose that perception is indirect.