Three semiotic layers of spoken communication

John Laver*

Department of Speech and Language Sciences, Queen Margaret University College, Clerwood Terrace, Edinburgh EH12 8TS, Scotland, UK

Received 10 October 2002; accepted 8 April 2003

My background as a phonetician has been more in speech production than perception, but it is never fully possible to deal with one without continually referring to the other. The papers in this volume by Coleman (2003) and Hawkins (2003) fit closely with the related one on phonetics by Local (2003), also in this volume; so I make reference to all the three, as I discuss what I see to be the necessary embedding of the study of the perception of spoken language in the wider study of interactional communication.

I take the metatheoretical view that the constructs of linguistic and phonetic theory which propose analytic categories are convenient fictions, whose convenience for the given purpose temporarily outweighs the evident distortion of the artificiality they introduce. The risk is that one progressively forgets about the distortion associated with these basic categories, and the temporary distortion becomes a habitual conceptual straitjacket. Happily, the disadvantage of this artificiality normally grows, ultimately, to the point where the convenience of the original constructs is realized as no longer profitable. Revision ensues, and a new (temporarily more satisfactory) set of fictions is proposed. Thus science moves forward, and we have probably now reached the point where it is desirable to move beyond local, punctual, short-domain evidence in speech routinely to consider longer-domain and more detailed phenomena for their relevance to both production and perception.

In this commentary, I leave aside the very distributed nature of phonetic information about the identity of linguistic units such as individual phonemes, taking it for granted that Hawkins, Coleman and Local have all demonstrated this beyond challenge. I note in passing, however, that the long anticipation of some types of speech sounds demonstrated in these three excellent papers will bring into re-examination some of the conclusions on self-monitoring for speech errors. The rest of my comments concentrate on the need to set the study of the perception of spoken language in a wider context.

I feel very sympathetic to Local’s (2003) position when he says that “Spoken language is a resource which is systematically deployed in the management of social interaction, its primary site...
of occurrence. The patterns and structures in language are emergent properties ... shaped by the contingencies and demands of social interaction ... Meaning is much more than lexical meaning.” Similarly, I support Coleman’s proposition that “learning phonology requires learning meanings”, as well as Hawkins' insistence that “phonetic information is multi-purpose”.

In sympathy with this view, I would want to say that we speak not just to be understood, but to communicate. Managing social interaction through speech is never limited to spoken language alone. The linguistic strand is only one of the three simultaneous strands of spoken communication, together with paralinguistic and extralinguistic elements. Even if we discount the visible aspects of paralinguistic (or nonverbal) communication, the audible paralinguistic features, especially including tone of voice, help to emphasize key aspects of verbal information, to imply the speaker’s pragmatic intent, to demarcate rhythmic and intonational units of spoken language, to manage the cooperative time-sharing of the role of speaker, to indicate the momentary affective and emotional condition of the speaker, and much else. Extralinguistic aspects communicate information about quasi-permanent physical, social and psychological details of speaker identity. All the three strands are interwoven in every interaction, with each acting potentially as the perceptual ground against which the figures of the other two must be discerned.

The relevance of mentioning the three-stranded nature of spoken communication is that the co-present nature of the three strands presents the listener, especially when faced with an unknown speaker, with a perceptual semiotic dilemma that can only be resolved by longer-term information. Two examples to illustrate this may be helpful, one extreme, one normal. Imagine hearing the previously unknown speaker beginning to talk, in a language in which whispery-voiced stops are contrastive with stops with normal voicing. The vowel after the initial stop in the first syllable is heard to be pronounced with whispery voicing. Should the mode of phonation be perceptually associated with a whispery-voiced linguistic identity of the preceding stop, a paralinguistic one of a confidential tone of voice, or an extralinguistic one of vocal pathology? The decision is contingent on hearing more speech, and on further acquaintance with the speaker.

The second, more everyday example comes from paralinguistic tone of voice. Many attitudinal effects are communicated by manipulating the precise location of given vowels in the vowel space. In English, and perhaps universally, incipient laughter is paralinguistically signaled partly by a medium-term articulatory setting which compresses the vowel space up and back towards the velum and the back wall of the pharynx, phonetically shifting vowels which are normally non-high and non-back slightly up and back in the vowel space. Hearing the first syllable or two of such speech leaves the listener uncertain about the linguistic phonemic values to allocate to the vowels heard, except by paralinguistically recognizing the incipience of the laughter by hearing the following syllables. The biasing effect of the articulatory setting can then be discounted. Similar semiotic difficulties attach to perceiving articulation constrained by another articulatory setting—that of a smile, which can neutralize lip-rounding in languages where this distinguishes two vowels with similar tongue positions. Once again, the classificatory decision is contingent on hearing more speech, and bringing to bear knowledge of lexical structure.

It is also impossible to assess the phonetic or phonemic value of a given vowel pronunciation without knowing the speaker's accent, which will only become obvious after exposure to a longer sample of the speaker's speech. Every speaker distributes their vowels in vowel space in a way which is characteristic both of their sociolinguistic group and idiosyncratically of their own
habitual speech. Hearing a vowel in an isolated syllable from an unfamiliar speaker leaves the listener with inadequate information about any general bias (i.e. any articulatory setting) on average vowel position that the speaker may be characteristically imposing. Without being able to discount this bias—i.e. to normalize the effect of the articulatory setting, the listener cannot allocate the vowel heard to some phonemic value (unless lexical identity constrains the choice) until a longer stretch of speech allows confidence to grow.

The performance of a given vowel by a given speaker on a given occasion is hence a blend of three different phonetic influences: its categorial value as an exponent of some phonological unit, its paralinguistic attitude-communicating function, and its function in signaling aspects of the physical, social or psychological identity of the speaker.

The semiotic dilemmas inherent in distinguishing the linguistic, paralinguistic and extra-linguistic information in spoken communication are not exceptional; they are part of daily experience for all listeners engaged in spoken interaction. It would therefore seem unrealistic to expect that the study of the perception of spoken language can be properly undertaken without a complementary and interlocking study of both paralinguistic and extralinguistic aspects that accompany spoken language in communicative interaction.

References

