591

Reprint Series
10 July 1987, Volume 237, pp. 169–171

SCIENCE

Speech Perception Takes Precedence over Nonspeech Perception

D. H. WHALEN AND ALVIN M. LIBERMAN

Speech Perception Takes Precedence over Nonspeech Perception

D. H. Whalen and Alvin M. Liberman

Some components of a speech signal, when made more intense, are heard simultaneously as speech and nonspeech—a form of duplex perception. At lower intensities, the speech alone is heard. Such intensity-dependent duplexity implies the existence of a phonetic mode of perception that takes precedence over auditory modes.

One theory of speech perception holds that there is a biologically distinct system, or module, specialized for extracting phonetic elements (especially consonants and vowels) from the sounds that convey them (*1*). The percepts produced by this module are immediately phonetic in character; accordingly, they stand apart from auditory percepts that are composed of standard dimensions such as pitch, loudness, and timbre. There is, then, no first-stage auditory percept, as most other theories of speech suppose (*2*), and hence no need for a subsequent stage in which the auditory tokens are matched to phonetic prototypes and thereby made appropriate for further processing as language. Indeed, as the experiments reported here show, it is the phonetic module that has priority, as if its processes occurred before, not after, those that yield the standard dimensions of auditory perception.

Consistent with the existence of a distinct phonetic mode is the observation that a particular piece of sound can evoke radically different percepts, depending on whether or not it engages the phonetic module. Consider, for example, acoustic patterns sufficient for synthesizing on a computer the syllables "da" and "ga" (Fig. 1, top). The three formants represent resonances of the vocal tract and have, at their onsets, frequency sweeps called transitions. These transitions last approximately 50 msec and reflect the way in which the resonances change as the tongue and jaw move from the consonant to the vowel. Normally, the perceived distinction between "da" and "ga" depends on many acoustic variables; as seen in Fig. 1, however, it can be made to depend only on differences in the transition of the third formant. Thus, in the context of the syllable, these transitions become crucial to the phonetic percept. In isolation (Fig. 1, bottom right), however, they are heard as the glissandi or differently pitched "chirps" that would be expected on the basis of psychoacoustic considerations. These two ways of perceiving the formant transitions—one phonetic, the other auditory—are strikingly different: there is no hint of chirpiness in the "da" or "ga," and no "da"-ness or "ga"-ness in the chirps. Moreover, the transitions are discriminated differently depending on the mode in which they are perceived (*3*).

Under special circumstances, the transitions can simultaneously evoke the phonetic and auditory percepts. This effect, called duplex perception, occurs when the third-formant transition is presented by itself to one ear, while the remainder of the pattern, called the base (Fig. 1, bottom left), is presented to the other. Listeners then simultaneously hear a chirp (in the ear to which the transition is presented) and the syllable "da" or "ga" (in the other ear), as determined by the transition. These simultaneous percepts, and the different discrimination functions they yield, are nearly the same as those produced separately by the isolated transitions and the whole syllable (*4*).

Since duplex perception occurs in response to a fixed acoustic pattern and results in two simultaneous percepts, it cannot be attributed to auditory interactions arising from changes in acoustic context or to a shifting of attention between two forms of an ambiguous stimulus. Further, that the "da" or "ga" is perceived to be entirely in one ear, even though the critical transition is presented only to the other, indicates that the incorporation of the transition into the base is an integration at the perceptual level, not a "cognitive" afterthought that deliberately combines what had initially been perceived as separate. Thus the phenomenon of duplex perception provides support for the view that there are distinct phonetic and auditory ways of perceiving the same (speech) signal. At the same time, however, it raises the question of why, in the normal case, the components of speech are not perceived in a duplex fashion; that is, why is the "da" or "ga" not normally accompanied by the chirp?

Relying on considerations of plausibility and simplicity, Mattingly and Liberman (*5*) proposed that the phonetic module preempts the phonetically relevant parts of the signal before making the remainder available to auditory processing. This proposal seemed plausible because, in contrast to the indefinitely large set of acoustic events that occur, phonetic events form a natural class that is defined by its correspondence to the acoustic results of specialized movements of the articulatory organs. The proposal was simple because the very processes of phonetic perception remove from the signal all evidence of those phonetic events and thus preclude such (parallel) processing as would cause them to be perceived yet again as chirps. This preemptiveness is similar to the precedence described above, which we have here demonstrated directly with a new and somewhat simpler version of a duplex phenomenon (*6*).

Our procedure differs from that used previously in that the two parts of the signal are not divided between the ears but are presented equally to both. Duplexity is produced (in both ears at once) by changing the intensity of the transition relative to the base. At relatively low intensities, the transitions serve only their expected phonetic function. At higher intensities, however, the transitions continue to make their phonetic contribution but simultaneously evoke nonspeech chirps. On the basis of these observations, which we made initially in pilot experiments, we tested the following generalizations.

1) In isolation, neither transition sounds like "da" or "ga."

2) In syllabic context, the transitions will, at some intensity, evoke nonspeech chirps, establishing a duplexity threshold.

3) Above the duplexity threshold, the chirps can be matched to those evoked by the transitions in isolation.

4) Both below and above the duplexity threshold, the transitions appropriately determine whether the syllable is heard as "da" or "ga."

The stimuli were the same as those represented in Fig. 1, except that the third-formant transitions were not frequency bands excited by a fundamental (as were the formants of the base) but rather time-vary-

D. H. Whalen, Haskins Laboratories, New Haven, CT 06511.
A. M. Liberman, Haskins Laboratories, New Haven, CT 06511; Department of Linguistics, Yale University, New Haven, CT 06511; and Department of Psychology, University of Connecticut, Storrs, CT 06268.

ing sinusoids that follow the center frequencies. Such sinusoidal transitions combine with the formant-synthesized base to make coherent phonetic percepts, in this case "da" and "ga." The sinusoids have the advantage, for our purposes, that in isolation they produce whistles, which were more easily discriminated than the chirps and also less speechlike.

The base syllable was created with a software formant synthesizer; the sinusoids were created with another software synthesizer designed for the generation of pure tones. From a set of input parameter values representing frequencies and amplitudes, each synthesizer calculated a digital waveform that was then turned into sound through a digital-to-analog converter. The base was synthesized in one computer file and the two sinusoidal transitions (one modeled after "d" and one after "g") in two other files. The base and one transition could then be output through synchronized digital-to-analog channels, separately attenuated, and electronically combined for presentation through headphones as a single sound to subjects. The base was presented at a fixed intensity of 72-dB sound-pressure level.

Eleven young adult speakers of English (six female and five male) with no reported hearing problems were tested in separate sessions. None knew anything about the composition of the stimuli or the purpose of the experiment. One subject did not perceive in a duplex fashion at the intensity levels available and therefore was excluded from all analyses.

Initially, subjects were asked to identify the sinusoidal transitions as "da" or "ga." Twenty repetitions of each were presented in random order. The subjects' responses are shown in Table 1, task 1. (For all tests, there was no significant difference between the responses to the "d" and "g" stimuli, so that only the combined percentages are reported.) Most subjects identified one whistle or the other as "da" and held to that consistently. Some happened to identify the correct one; others were just as consistently wrong. One (subject 9) simply called all the whistles "da." Overall, identification accuracy did not differ significantly from chance [$t(9) = 1.22$].

To find the intensity at which the sinusoids in syllabic context evoked nonspeech whistles in addition to "da" or "ga" (the duplexity threshold), we had the subjects adjust the attenuator that controlled the intensity of the sinusoid until the whistle was just audible. This was done three times for each sinusoid. The mean duplexity thresholds for all subjects, expressed in relation to the steady state of the third formant, were −6.4 dB (SD, 5.0 dB) for the "da" sinusoid and 0.0 dB (SD, 4.9 dB) for the "ga" sinusoid. This difference in duplexity thresholds, which was found for all ten subjects, is consistent with the fact that, in isolation, the "da" sinusoid (the one with the lower duplexity threshold) was the louder of the two.

To ensure that the whistle component of the duplex percept was comparable to the whistle of the sinusoid in isolation, we performed a matching test. On each trial, three stimuli were presented: first one sinusoid in isolation, then either of the two sinusoids in syllabic context, and finally the other sinusoid in isolation. Each sinusoid occurred with the syllable 20 times, matching the first sinusoid or the last an equal number of times. The sinusoid in the syllable was presented at 6 dB above the duplexity threshold for "ga." Subjects judged whether the duplexly perceived whistle was more like the isolated whistle that preceded or followed it. Subjects were able to make this judgment accurately well above the level of chance [$t(9) = 5.50$, $P < 0.001$; Table 1, task 2] (*7*).

To test whether the sinusoids reliably determined how the syllable was perceived below the duplexity threshold, we set them 4 dB below the "da" duplexity threshold and presented 20 repetitions of each in random order. Subjects were to identify the consonant as "d" or "g." Again, they did so at a level well above chance [$t(9) = 8.88$, $P < 0.001$; Table 1, task 3].

It remained, then, to determine whether the sinusoids continue to provide phonetic information even when they also evoke whistles. For that purpose, we set the sinusoids at 6 dB above the higher ("ga") duplexity threshold and performed another identification test. Subjects' identifications were no less accurate above the duplexity threshold than below it [$t(9) = 32.60$, $P < 0.001$; Table 1, task 4].

Table 1. Correct performance (in percent) on the four main tasks (results from 40 trials per subject). Task 1, identification of isolated sinusoids as "d" or "g"; task 2, match of duplex to isolated sinusoids; task 3, identification of syllables as "da" or "ga" below duplexity threshold; task 4, identification of syllables as "da" or "ga" above duplexity threshold.

Subject	Task 1	Task 2	Task 3	Task 4
1	72.5	92.5	100.0	100.0
2	100.0	65.0	100.0	97.5
3	15.0	97.5	100.0	100.0
4	95.0	97.5	100.0	100.0
5	30.0	85.0	97.5	100.0
6	95.0	72.5	92.5	85.0
7	100.0	87.5	82.5	97.5
8	0.0	95.0	52.5	100.0
9	50.0	47.5	100.0	97.5
10	90.0	65.0	100.0	100.0
Mean	64.8	79.5	92.5	97.8
SEM	±12.1	±5.4	±4.8	±1.5

Thus, at lower levels of intensity, the sinusoids provide the basis for the perceived distinction between "da" and "ga"; at higher levels, they serve this same phonetic purpose but also evoke nonspeech whistles. As we found from our own listening, the phonetic information is provided over a range of approximately 20 dB below the duplexity threshold (*8*); the whistles, which are of course barely audible at the duplexity threshold, become louder as the intensity of the sinusoid is increased. These results show that processing of the sinusoid as speech has priority, thereby defining what we mean by precedence of the phonetic module.

Unlike the earlier version of a duplex

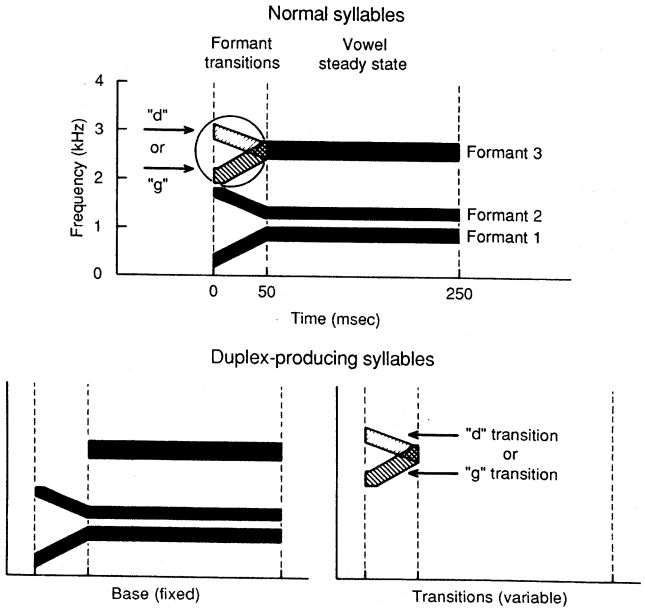

Fig. 1. Schematic representation of the syllables "da" and "ga."

phenomenon (5), which required that the transitions and the remainder of the pattern be presented to different ears, the one reported here puts all parts of the pattern equally into both ears. It thereby avoids such complications of interpretation as may arise with dichotic stimulation and so makes more straightforward the inference that duplex perception reflects distinct auditory and phonetic ways of perceiving the same stimulus. Beyond that, the results obtained with the new form of the duplex phenomenon support the hypothesis that the phonetic mode takes precedence in processing the transitions, using them for its special linguistic purposes until, having appropriated its share, it passes on the remainder to be perceived by the nonspeech system as auditory whistles. Such precedence reflects the profound biological significance of speech.

REFERENCES AND NOTES

1. A. M. Liberman and I. G. Mattingly, *Cognition* **21**, 1 (1985).
2. R. A. Cole and B. Scott, *Psychol. Rev.* **81**, 348 (1974); G. C. Oden and D. W. Massaro, *ibid.* **85**, 172 (1978); K. N. Stevens, in *Auditory Analysis and Perception of Speech*, G. Fant and M. A. Tatham, Eds. (Academic Press, New York, 1975), pp. 303–330.
3. I. G. Mattingly, A. M. Liberman, A. K. Syrdal, T. Halwes, *Cogn. Psychol.* **2**, 131 (1971).
4. V. A. Mann and A. M. Liberman, *Cognition* **14**, 211 (1983).
5. I. G. Mattingly and A. M. Liberman, in *Functions of the Auditory System*, G. M. Edelman, W. E. Gall, W. M. Cowan, Eds. (Wiley, New York, in press).
6. See page 206 of C. S. Darwin and N. S. Sutherland [*Q. J. Exp. Psychol.* **36A**, 193 (1984)] for a related observation.
7. Below the duplexity threshold, such matching would presumably be at the level of chance. It is possible, however, that forced matching is a more sensitive measure than the one we used to obtain the threshold itself. We therefore applied the matching procedure at 4 dB below the lower ("d") threshold, using eight highly practiced subjects. As expected, the responses [45.3% correct, $t(7) = -1.28$, $P > 0.2$] were at the level of chance.
8. S. Bentin and V. A. Mann [*Haskins Laboratories Status Report on Speech Research SR-76* (1983)] found a similar range in a dichotic task, although they interpreted it as a difference in sensitivity, not as preemption.

9. Supported by National Institute of Child Health and Human Development grant HD-01994 to Haskins Laboratories. We thank C. A. Fowler, I. Mattingly, B. Repp, L. Rosenblum, P. Rubin, and M. Studdert-Kennedy for helpful comments.

23 December 1986; accepted 22 April 1987