

A right-ear advantage for perception of graveness in monaural processed speech

HOWARD J. KALLMAN, DENISE DAVIDSON, and EILEEN JOYCE
State University of New York at Albany, Albany, New York

(Gordon Gallup, *Sponsor*)

The Diagnostic Rhyme Test (DRT) is designed to assess the degree to which various distinctive features of speech degraded by processing are perceived by listeners. The DRT involves the presentation of lists of verbally presented words. As each word is presented, two response alternatives that differ in terms of a distinctive feature of speech are displayed visually. In the present experiment, eight highly practiced listeners were tested on the DRT, using speech that had been degraded by processing; right- and left-ear performance was contrasted. A right-ear advantage (REA) in identifying the distinctive feature of graveness was found, but no other distinctive feature yielded an REA. The implications of these results are discussed.

When speech sounds are presented dichotically, recognition is generally more accurate for those sounds presented to the right than for those presented to the left ear (Berlin & McNeil, 1976; Bradshaw & Nettleson, 1981; Geffen & Quinn, 1984). Despite substantial research on this phenomenon, questions pertaining to the right-ear advantage (REA) remain unanswered. For example, although it is generally acknowledged that REAs are less likely to occur with monaural than with dichotic presentation, there are varying explanations for this. At one time, the consensus was that REAs occurred only with dichotic presentations because the REA depended on the occlusion of activity in the ipsilateral ear-to-cortex pathways by activity in the contralateral pathways (Rosenzweig, 1951). According to this view, input to an ear is transmitted directly to the contralateral hemisphere given dichotic presentation of stimuli. Therefore, speech that is presented to the right ear is preferred, due to its direct transmission to the hemisphere that is specialized for language processing in most individuals. Given that occlusion of ipsilateral activity by contralateral activity would not be a factor with monaural presentations, monaural input would be transmitted to both hemispheres, via the contralateral and ipsilateral pathways. Although the contralateral pathways are stronger than the ipsilateral, this was thought to be insufficient reason to predict REAs with monaural stimulation.

Although this explanation of why ear differences were found with dichotic but not with monaural stimulation received a wide measure of acceptance, recent evidence suggests that the increased task difficulty brought about by dichotic stimulation may be at least as important.

This research was supported by a grant from the General Electric Research and Development Center, Niskayuna, NY. We thank GE for its support. Requests for reprints may be sent to Howard J. Kallman, Psychology Department, State University of New York at Albany, 1400 Washington Ave., Albany, NY 12222.

Robust REAs with monaural stimulation have been demonstrated when relatively difficult tasks are used (Bradshaw, Farrelly, & Taylor, 1981; Young & Ellis, 1980). Furthermore, if sensitive dependent measures, such as reaction time, are used instead of less sensitive accuracy measures, ear differences with monaural stimulation are often found (Fry, 1974; Kallman, 1977; Morais, 1976; Morais & Darwin, 1974; Rastatter & Gallaher, 1982). One advantage of using monaural stimuli to test ear differences is that such stimuli are easier to produce than are dichotic stimuli. Accordingly, one of the challenges for researchers is to develop tasks of sufficient difficulty, or that use sufficiently sensitive measures, to assess ear asymmetries with monaural stimuli.

Another issue of interest concerns the specific features of the speech stimulus that yield REAs. There have been attempts to address this using dichotic listening procedures. For example, Hayden, Kirstein, and Singh (1979) presented dichotic pairs of consonant-vowel (CV) syllables and had subjects report the items from one or the other ear. All possible syllable-initial consonants were used except /h/. Although an REA that was greatest with stop consonants was found, the authors concluded that "the analysis and interpretation of dichotic data are complicated because of the sheer number of perceptual phenomena and their interaction as well as the multiplicity of stimulus variables" (p. 1045). Indeed, one of the problems with the use of dichotic tasks is that performance becomes a function not only of the speech stimulus presented to a given ear, but also of its interaction with the stimulus presented to the other ear. This problem can be overcome through the use of monaural rather than dichotic stimulus presentations (Fry, 1974; Morais, 1976).

In the present experiment, the Diagnostic Rhyme Test (DRT) was used to evaluate ear advantages for specific distinctive features of monaurally presented speech. The DRT has previously been used as a tool to evaluate the

degree to which various distinctive features (Jakobson, Fant, & Halle, 1963; Miller & Nicely, 1955) are intelligible in processed speech (Allen, Strong, & Palmer, 1981; Sandy, 1982; Smith, 1969; Voiers, 1977, 1983). During administration of the DRT, a single word (the target) is presented per trial. At the same time, the listener sees two words, the target and a lure. The target and lure rhyme and differ by a distinctive feature. To the extent that the target is correctly identified, evidence for the intelligibility of the tested distinctive feature is provided. The specific distinctive features tested by the DRT are voicing, nasality, sustention, sibilation, graveness, and compactness. Because performance on the DRT depends on perception of single distinctive features embedded in a wide variety of single-syllable words, the task is a relatively difficult one in comparison with other speech perception tasks.

We administered the DRT to practiced listeners and compared left- and right-ear performance on each of the DRT indices. We decided to use highly trained listeners because they tend to yield the most stable scores on the DRT indices. The use of practiced subjects was of some theoretical import as well. Some studies have suggested that ear advantages may disappear when subjects are repeatedly tested with the same stimuli and task, presumably because processing becomes more automatic and less dependent on specialized processors (Kallman & Corballis, 1975). Kallman and Corballis, using musical stimuli, found a left-ear advantage that disappeared within an hour of testing. Although similar practice effects have also been demonstrated with lateralized visual stimuli (Hellige, 1976), it was not clear whether speech REAs would also disappear given repeated testing of the same speech stimuli. Although the present study did not provide a direct measure of the effect of practice on speech REAs, the demonstration of REAs with our listeners, who had extensive practice with the DRT stimuli, would suggest that speech REAs persist despite repeated exposure to the same speech stimuli.

METHOD

Subjects

Eight listeners, each of whom had approximately 20 h of previous experience as listeners with the DRT, were paid for their participation. The subjects were initially recruited from the State University of New York at Albany community, and each reported being right-handed.

Stimuli, Apparatus, and Procedure

Three lists of DRT words were recorded for use in the present experiment; each list was read by a different male speaker. These lists are referred to as Lists A, B, and C. A list consisted of 232 words that were presented at the rate of approximately 1 every 1.3 sec. The lists were generated using the guidelines provided by Voiers (1977). Within each list, there were 32 trials per each of the six distinctive features plus 40 "filler" trials. The words on the filler trials were not chosen to test distinctive features and were included in the present study only because they are standardly used in DRT tests.

The six distinctive features tested were voicing, nasality, sustention, sibilation, graveness, and compactness. An example of a trial that was designed to test the voicing feature was one in which the word *bond*

was presented and the subject had to choose between *bond* and *pond*. Within each word list, the features tested alternated from trial to trial such that the six distinctive features were all tested once before any was tested a second time, and so forth. Although within each list there were 32 trials per distinctive feature, only 16 pairs of words were used per distinctive feature. During the first half of the list, one of the words served as target, and the other word served as target during the second half of the list.

The three word lists were recorded on tape and then processed by a Codec and then a sub-band coder. The Codec had a sampling rate of 5,800 Hz and the high-frequency filtering was such that the output was attenuated by 3 dB at 2,500 Hz and by about 40 dB at 2,900 Hz. The output from the Codec was processed by a sub-band coder, which had four bands spanning the frequency ranges (1) 181¼ to 362½, (2) 362½ to 725, (3) 725 to 1,450, and (4) 1,450 to 2,900. The sampling rate of each band was twice the bandwidth. The output from the sub-band coder was recorded onto audiotape. The primary effect of processing the speech was to limit the high frequencies and add quantization noise.¹

During the experiment itself, the tape containing the processed speech was played back over an Akai ADF-770 tape deck and amplified by Coulbourn Instruments Model S82-24 mixer-amplifiers. Subjects heard the stimuli over TDH-50 matched headphones at an average intensity of 72 dB(A) as measured by a Bruel & Kjaer Model 2203 sound-level meter and Type 4152 artificial ear.

Subjects listened to List A, List B, and then List C. The three lists were then repeated in the same order. During each list presentation, all of the words were presented to either the right (R) or left (L) ear. From list to list, the ear of presentation was alternated. Half of the subjects heard the first list at the left ear, while the other subjects heard it at the right.

Subjects were tested individually in a sound-attenuated room. A Commodore 64 computer system was used to present visually the response alternatives and to record the listeners' responses, which were made by pressing one of two buttons to indicate which of two words was thought to be heard. A third response key allowed for changing the response as long as a response to the next word had not yet been made. No feedback about the correct response was given. The programs used to control administration of the DRT are described elsewhere (Calistri & Kallman, 1986).

RESULTS

DRT scores for each listener were computed for each distinctive feature at each ear; these scores, averaged across lists, are presented in Table 1. Each DRT score was computed according to the formula:

$$DRT = [100 \times (\# \text{ correct} - \# \text{ wrong})] / \text{total } \# \text{ of trials,}$$

which incorporates a correction for guessing (Voiers, 1977).

An analysis of variance (BMDP2V) was performed on the DRT scores. Ear (2 levels) and distinctive feature (6 levels) served as within-subjects variables. The effect of distinctive feature was significant [$F(5,35) = 47.01, p < .01$]. Performance was best for voicing, worst for graveness, and intermediate for the other features. A significant REA was found [$F(1,7) = 9.22, p < .05$], but the effect of ear interacted with feature [$F(5,35) = 3.01, p < .05$].

To determine which of the features yielded reliable REAs, six *t* tests, each comparing left- and right-ear performance for a particular distinctive feature, were computed. The *t* values appear in Table 1. The only distinc-

Table 1
DRT Scores for Distinctive Features Presented to the Left and Right Ears

	Voicing	Nasality	Sustention	Sibilant	Graveness	Compactness
Left Ear	91.4	97.7	86.4	86.7	68.5	92.1
Right Ear	89.4	97.9	88.7	86.3	74.2	92.3
<i>t</i> Value	1.23	0.17	1.59	0.36	4.08	0.13

The *t* values (7 *df*) for each left/right ear comparison are also presented. (The *t* value for the graveness feature was the only one significant at $p < .05$.)

tive feature that yielded a significant REA was graveness. This conclusion followed, regardless of whether the alpha level was set at .05 for each of the six tests or whether the more conservative procedure of dividing the alpha level equally among the six comparisons was followed.

DISCUSSION

A monaural speech REA was found for the distinctive feature of graveness, but not for any of the other tested features. The present experiment thus showed limited success in demonstrating an REA with monaural stimuli. The REA was obtained despite the use of an accuracy, rather than a more sensitive reaction-time, measure. Furthermore, the REA was obtained using highly practiced listeners who had been exposed to the same stimuli and basic task for, on an average, over 20 h. Finally, the REA was found in a situation in which subjects always knew in advance to which ear the stimulus would be presented, a situation that rules out an explanation of the REA in terms of an attentional bias and suggests that uncertainty about presentation ear is not a necessary condition for obtaining monaural REAs (see Geffen & Quinn, 1984).

It is of interest that the REA that was obtained was for the graveness feature only. Grave sounds are characterized by predominance of energy in the lower half of the acoustic spectrum, whereas the contrasting acute sounds are characterized by energy in the upper half. The processing of the speech stimuli by the Codec and sub-band coder had the effect of removing much of the high-frequency information. This type of processing, then, generally makes the grave-acute contrast particularly difficult to perceive. It could be that the reason the REA was found for the distinctive feature of graveness, but not for the other distinctive features, was the increased difficulty of perceiving the graveness feature brought about by the speech processing. However, such an explanation does not detract from the conclusion that the left hemisphere is specialized for processing those cues that are used to perceive the grave-acute distinction; when those cues are obscured by speech processing, perception of the feature becomes difficult and an REA obtains.

The DRT appears to be a useful device for exploring the features of speech that yield ear advantages because it provides a convenient method for focusing on the perception of specific speech features. Future research should focus on the effect of various types of speech processing on the ear advantages obtained for the various distinctive features of speech.

REFERENCES

- ALLEN, D. R., STRONG, W. J., & PALMER, E. P. (1981). Experiments on the intelligibility of low-frequency speech. *Journal of the Acoustical Society of America*, *70*, 1248-1255.
- BERLIN, C. I., & MCNEIL, M. R. (1976). Dichotic listening. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 327-387). New York: Academic Press.
- BRADSHAW, J., FARRELLY, J., & TAYLOR, M. (1981). Synonym and antonym pairs in the detection of dichotically and monaurally presented targets: Competing monaural stimulation can generate a substantial right ear advantage. *Acta Psychologica*, *47*, 189-205.
- BRADSHAW, J. L., & NETTLESON, N. C. (1981). The nature of hemispheric specialization in man. *The Behavioral & Brain Sciences*, *4*, 51-91.
- CALISTRI, R., & KALLMAN, H. J. (1986). DRT/***: Programs to administer and score the Diagnostic Rhyme Test. *Behavior Research Methods, Instruments, & Computers*, *18*, 57-58.
- FRY, D. B. (1974). Right ear advantage for speech presented monaurally. *Language & Speech*, *17*, 142-151.
- GEFFEN, G., & QUINN, K. (1984). Hemispheric specialization and ear advantages in processing speech. *Psychological Bulletin*, *96*, 273-291.
- HAYDEN, M. E., KIRSTEIN, E., & SINGH, S. (1979). Role of distinctive features in dichotic perception of 21 English consonants. *Journal of the Acoustical Society of America*, *65*, 1039-1046.
- HELLIGE, J. B. (1976). Changes in same-different laterality patterns as a function of practice and stimulus quality. *Perception & Psychophysics*, *20*, 267-273.
- JAKOBSON, R., FANT, C. G. M., & HALLE, M. (1963). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.
- KALLMAN, H. J. (1977). Ear asymmetries with monaurally-presented sounds. *Neuropsychologia*, *29*, 833-835.
- KALLMAN, H. J., & CORBALLIS, M. C. (1975). Ear asymmetry in reaction time to musical sounds. *Perception & Psychophysics*, *17*, 368-370.
- MILLER, G. A., & NICELY, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, *27*, 338-352.
- MORAIS, J. (1976). Monaural ear differences for reaction times to speech with a many-to-one mapping paradigm. *Perception & Psychophysics*, *19*, 144-148.
- MORAIS, J., & DARWIN, C. J. (1974). Ear differences for same-different reaction times to monaurally presented speech. *Brain & Language*, *1*, 383-390.
- RASTATTER, M. P., & GALLAHER, A. J. (1982). Reaction-times of normal subjects to monaurally presented verbal and tonal stimuli. *Neuropsychologia*, *20*, 465-473.
- ROSENZWEIG, M. R. (1951). Representation of the two ears at the auditory cortex. *American Journal of Physiology*, *167*, 147-158.
- SANDY, G. F. (1982). "Digital Voice Processor Consortium Interim Report," Appendix A, MTR-81W0159-02, Mitre Corp., McLean, VA.
- SMITH, C. P. (1969). Perception of vocoder speech processed by pattern matching. *Journal of the Acoustical Society of America*, *46*, 1562-1571.
- VOIERS, W. D. (1977). Diagnostic evaluation of speech intelligibility. In M. E. Hawley (Ed.), *Speech intelligibility & recognition* (pp. 374-387). Stroudsburg, PA: Dowden, Hutchinson and Ross.
- VOIERS, W. D. (1983). Evaluating processed speech using the Diagnostic Rhyme Test. *Speech Technology*, *3*, 30-39.
- YOUNG, A. W., & ELLIS, H. D. (1980). Ear asymmetry for the perception of monaurally presented words accompanied by binaural white noise. *Neuropsychologia*, *18*, 107-110.

NOTE

1. The reason for using processed speech was to increase the difficulty of the perceptual task; the choice of a sub-band coder for processing was based solely on its availability.