Recognition of Digitized CV Syllables in Multitalker Babble

Sandra Gordon-Salant
University of Maryland, College Park, Md., USA

Abstract. Nonsense syllable recognition by 10 normal-hearing listeners was assessed in quiet and three levels of multitalker babble. Stimuli were 19 consonants paired with /a/, /i/, and /u/ in a CV format. Performance measures included total syllable recognition, consonant feature-class recognition, and consonant feature errors. Results of each analysis were strongly affected by vowel coarticulation and noise level. However, a distinctive pattern of performance in the babble was observed. Comparisons are made between this pattern and those obtained previously in other types of noise backgrounds. Implications for clinical evaluation and auditory training are discussed.

Introduction

During the last decade, multitalker (MT) babble has become popular as a noise competitor in speech discrimination testing. This stems from an interest in simulating a realistic listening environment for testing, so that measures of speech discrimination will approximate a listener's everyday speech reception ability [*Kalikow* et al., 1977]. MT babble also has been used as competition to detect subtle auditory dysfunction, because speech backgrounds are more effective than steady-state noise in demonstrating performance deficits in listeners with mildly impaired auditory systems [*Aniansson,* 1974;

Findlay and Denenberg, 1977]. Investigations incorporating MT babble have shown that monosyllabic word discrimination usually is poorer in its presence than in non speech noise [*Aniansson,* 1974; *Danhauer and Leppler,* 1979]. Further, hearing-impaired listeners consistently perform more poorly than normal-hearing listeners in monosyllabic word tests presented in MT babble backgrounds [*Aniansson,* 1974; *Findlay and Denenberg,* 1977; *Garstecki and Mulac,* 1974; *Miner and Danhauer,* 1976].

MT babble appears to be the competitor of choice for clinical applications. However, the pattern of consonant recognition and

confusions in this type of noise is virtually unknown. The assumption that MT babble produces similar consonant confusion patterns to other types of noises may be erroneous. Previous investigations of consonant recognition in noise have shown that different patterns of performance emerge when different noises are presented. For example, nasality cues are minimally affected in white noise [*Miller and Nicely,* 1955; *Wang and Bilger,* 1973], while nasals are poorly recognized in cafeteria noise [*Dubno and Levitt,* 1981]. Fricatives are also identified with low accuracy in cafeteria noise [*Dubno and Levitt,* 1981] and white noise [*Miller and Nicely,* 1955]. However, in a second study in which white noise was the competitor, frication cues reportedly contributed to recognition performance [*Wang and Bilger,* 1973]. Vowel effects also have been different in the different types of noises. In white noise, consonants followed by /u/ are recognized with higher accuracy than consonants followed by /a/ and /i/, whereas in cafeteria noise, consonants paired with /a/ are recognized with higher accuracy than consonants paired with /i/ or /u/. Finally, place errors predominate in cafeteria noise [*Dubno and Levitt,* 1981] and white noise [*Miller and Nicely,* 1955] although some place cues were preserved in white noise in one study [*Wang and Bilger,* 1973].

It is apparent that different types of noises produce different patterns of recognition performance. Therefore, we cannot assume that specific patterns of performance that emerge in one type of noise will also emerge in the presence of a different noise. Given the apparent clinical utility of MT babble as a background competition, normative data related to its unique interference effects are needed. The purpose of this investigation is to examine recognition performance and error patterns associated with listening in MT babble. These results in turn can be compared to some of the more commonly observed performance patterns obtained in other types of noise competitions. Patterns of performance in a number of signal-to-noise ratio (SNR) conditions will be examined, because error patterns may change as a function of SNR. Vowel context is also varied in the experiment, because overall recognition scores change significantly with changes in vowel coarticulation [*Wang and Bilger,* 1973; *Dubno and Levitt,* 1981].

Method

Ten young adults between the ages of 18 and 28 served as paid subjects in this experiment. Each subject had pure-tone air conduction thresholds no greater than 15 dB HTL at the octave-interval test frequencies between 2.5 and 8 kHz, and excellent (90–100%) monosyllabic word discrimination ability, in each ear. Assessment of middle ear pressure and acoustic reflex thresholds confirmed that each subject had normal middle ear function. All subjects were native speakers of American English who lacked previous experience on nonsense-syllable tasks.

The speech stimuli were selected to represent individual consonant phonemes that occur in the initial position of English words. The stimulus set was composed of 19 consonants, each of which was paired with /a/, /i/, and /u/ in a CV sequence. The consonants included: /b, d, g, p, t, k, m, n, s, z, ʃ, f, v, θ, ð, w, j, r, l/.

All stimuli were recorded on tape by a male speaker of General American dialect who was a trained phonetician. The peak levels of all stimuli were equated to ensure equivalent SNRs across stimuli in the noise conditions. To this end, a computer adjustment procedure was used. The recorded stimuli were low-pass filtered (5 kHz cutoff, 48 dB/octave slope), digitized onto a laboratory computer (11,43 kHz rate), and the RMS energy of each CV was calculated over 20-ms intervals. All stimuli were then scaled in level so that their peak levels in dB SPL were equivalent. Note that this procedure for equating peak levels preserved the natural CV ratio of

each syllable. The 57 peak-equivalent CVs were converted to analog signals (11.43-kHz rate), low-pass filtered at 5 kHz (48 dB/octave slope), and recorded on tape 10 times in randomized order with a 2-second interstimulus interval. Five different tapes were created, each with a unique randomization of the stimuli. A 1-kHz calibration tone, equivalent in dB SPL to the peak level of each stimulus, was recorded at the beginning of each tape.

The MT babble competition was the 12-talker babble of the SPIN test [*Kalikow* et al., 1977]. It consists of two repetitions of 6 adult talkers (3 male and 3 female) reading the same section of a children's storybook. Each spoken passage is similar in phonetic composition to spoken English. The babble is characterized by a relatively flat spectrum below approximately 0.8 kHz, and an attenuation rate of 9–10 dB/octave above 0.8 kHz. The long-term spectrum of this babble is shown in figure 1.

The stimuli and noise were dubbed onto two separate channels of recording tape. During the experiment, the tapes were played back on an Otari MX 5050B tape recorder. The stimuli and noise were separately attenuated, mixed, amplified, and presented monotically to a single TDH-49 earphone. The subject was seated in a double-walled sound-insulated chamber during testing.

Four experimental conditions, corresponding to four background noise levels, were incorporated in this experiment: quiet, +12 dB SNR, +6 dB SNR, and 0 dB SNR. These SNR's were based on pilot data indicating that a range of scores from above chance to below perfect performance would be obtained. Calibration procedures were conducted prior to each experimental session. CVs were adjusted in level so that the 1-kHz calibration tone produced 80 dB SPL at the output of the earphones, as measured in an NBS 9A coupler. The overall level of the MT babble was calibrated to produce 80, 74 and 68 dB SPL at the output of the earphones, corresponding to the three SNRs of 0, +6 and +12 dB. In the quiet listening condition, the output of the background noise channel was disconnected.

During each listening condition, one randomization of the 570 CVs was presented to a listener. The order of listening conditions was randomized across subjects. The subject's task was to identify the consonant he/she perceived from a closed set of 19 choices (appearing in orthographic form in alphabetical order), in a written response. Feedback was not provided. A practice session was conducted in which the subject received instructions and identified one recorded sample of each stimu-
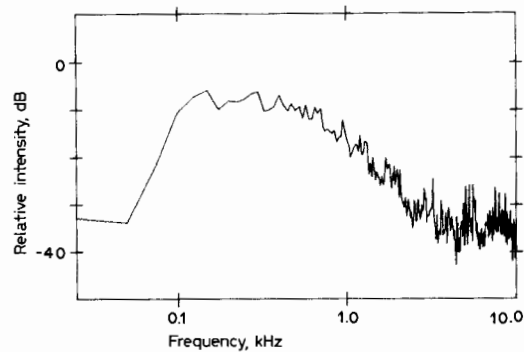

Fig. 1. Long-term averaged spectrum of the 12-talker babble used in the noise conditions.

lus. One complete randomization of 570 CVs was then presented to the subject for identification. Once the subject completed all practice items, the experimental conditions commenced. At the end of the experiment, the quiet condition was presented again to assess performance reliability. The entire experiment consisted of six separate listening sessions which were completed within 6 h.

Results

Overall Intelligibility

The total percent of all consonants correctly identified was calculated for each combination of noise condition (including the quiet condition) and vowel context, for each subject. The percent correct scores of all subjects were then transformed to remove the systematic relationship between treatment means and variances. An arc-sine transformation was used because the data were expressed as proportions [*Kirk,* 1968]. The transformed scores were submitted for analysis of variance (ANOVA) in a randomized block factorial design [*Kirk,* 1968]. The results revealed a significant main effect of noise condition (F = 316.649, p < 0.001, d.f.
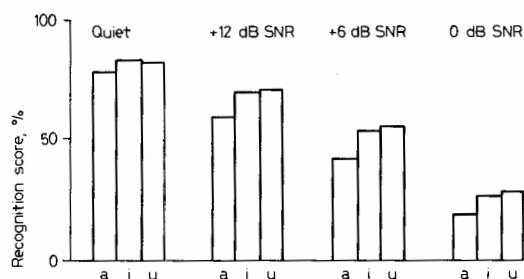
Fig. 2. Percent correct recognition scores, for consonants coarticulated with /a/, /i/, and /u/, in the four listening conditions.

= 3) and a significant main effect of vowel coarticulation (F = 32.107, p < 0.001, d.f. = 2). A significant noise × vowel interaction was not observed.

The mean percent correct scores for each of the four noise conditions across the three vowel contexts are presented in figure 2. As expected, correct identification of consonants decreased with each decrease in SNR. Post hoc multiple-comparison testing [Newman-Keuls, in *Kirk*, 1968] indicated that each of these differences was significant (p < 0.01). Figure 2 also shows that identification of consonants coarticulated with /a/ was consistently poorer than identification of consonants coarticulated with /i/ or /u/ at each noise level. Multiple comparison testing confirmed that this effect also was significant (p < 0.01).

The reliability between performances for the first and second experimental quiet conditions was calculated from the scores of all subjects in the three vowel contexts. The derived correlation coefficients were 0.91 for consonants paired with /a/, 0.86 for consonants paired with /i/, and 0.87 for consonants paired with /u/. These high positive correlations are similar to those reported by

other investigators for nonsense syllable recognition tests [*Dubno and Dirks*, 1982] and suggest that overall recognition scores were highly reliable.

Consonant Recognition

Recognition performance was analyzed according to the consonants' distinctive articulatory features, to determine which attributes of the consonants were degraded by the babble. A classification scheme based on consonant articulatory features, rather than acoustic or perceptual features, was selected to facilitate comparison with results of other investigators. Separate analyses were conducted for the features 'place', 'manner' and 'voicing'. For each analysis, scores were entered for each noise condition and vowel context, because these two variables had a significant effect on the total intelligibility measures. Percent correct performance for a feature category in an individual data matrix was determined by totaling the frequency of correct identifications of the consonants within that feature category and dividing by the frequency of presentations of those consonants.

The stimuli were assigned to one of three place categories, determined by the location of maximum vocal tract constriction during production. The front-place stimuli, /p, b, m, f, v, w/, were produced with maximum constriction at the lips, either bilabially or labiodentally. Stimuli produced interdentally or at the alveolar ridge comprised the mid-place category, and included /θ, ð, t, d, n, l, s, z/. Stimuli assigned to the back place were produced with maximum constriction at the palate or velum, and included /g, k, j, ʃ, r/. The percentage scores reflecting accurate identification of front, mid, and back-place consonants produced in the three vowel con-
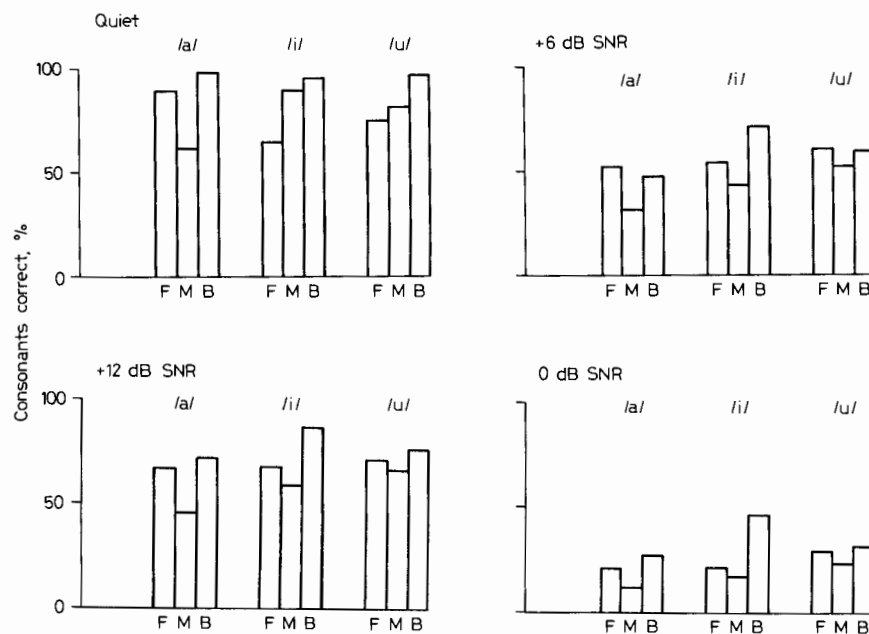
Fig. 3. Consonant recognition, in percent correct, for consonants produced at the front (F), middle (M), and back (B) of the mouth, as a function of vowel context and listening condition.

texts were determined from each subject's performance in the four noise conditions. An ANOVA was conducted on arc-sine transformations of these scores in a three-way design (vowel × noise × place), and a significant triple interaction was observed (F = 9.98, p < 0.001, d.f. = 12). Simple main effects analyses were performed on interactions involving place at each noise × vowel condition. Figure 3 presents the mean place scores in each noise condition across the three vowel contexts. This figure shows that recognition of the consonants according to place varied in each vowel context and noise condition. For consonants followed by /a/, recognition of mid-place consonants was significantly lower than front and back-place consonants, regardless of noise level (p < 0.01). In the /i/ vowel context, conso-

nants produced at the back of the mouth were identified with significantly higher accuracy than front and mid-place consonants, in the presence of MT babble (p< 0.01). However, in quiet, identification of front-place consonants was significantly poorer than identification of either mid or back-place consonants (p < 0.01). A place effect for consonants coarticulated with /u/ was observed only in the quiet condition, where back-place consonants were identified with the highest accuracy (p < 0.01). Thus, most difficulty was encountered in recognizing front and mid-place consonants, although specific patterns of performance were affected by vowel coarticulation and noise level.

Recognition of the consonants according to manner of production was assessed for four manner categories: glides and liquids
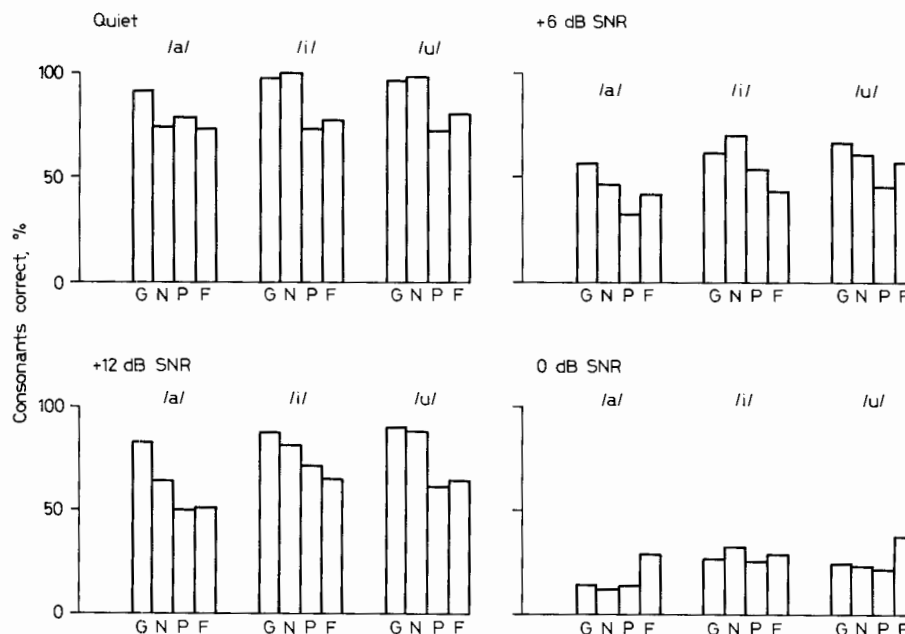
Fig. 4. Consonant recognition, in percent correct, for glides/liquids (G), nasals (N), plosives (P), and fricatives/sibilants (F), as a function of vowel context and listening condition.

/w, j, r, l/, nasals /m, n/, plosives /b, d, g, p, t, k/, and fricatives and sibilants /f, v, θ, ð, s, z, ʃ/. An ANOVA of arc-sine-transformed consonant identification scores for these four categories revealed a significant triple interaction (vowel × noise × manner) (F = 2.33, p < 0.005, d.f. = 18). An inspection of the recognition scores, displayed in figure 4, suggests that perception of manner changed systematically as the level of the MT babble increased. In low noise levels (Q and +12 dB SNR), nasals and glides/liquids usually were identified more accurately than plosives and fricatives/sibilants. These differences were significant for consonants coarticulated with /i/ and /u/ in quiet, and with /u/ at +12 dB SNR (p < 0.01). In the /i/ context at +12 dB SNR, the glides/liquids and nasals retained superior identification to the fricatives/sibi-

lants only (p < 0.01). In the context of /a/, only the glides/liquids were recognized with higher accuracy than the other manner categories, in quiet and +12 dB SNR. This trend was maintained partially at +6 dB SNR, where glides/liquids were recognized with greater precision than all other categories in the /a/ context; and nasals were identified more accurately than fricatives/sibilants in the /i/ context (p < 0.01). No significant differences in identification of manner classes were observed for consonants followed by /u/. At the poorest SNR, a different pattern of performance emerged: recognition was higher for fricatives/sibilants than for the nasals, glides/liquids, and plosives. Multiple comparison tests showed that this effect was significant for consonants followed by /a/ and /u/ (p < 0.01). Thus, as the level of bab-
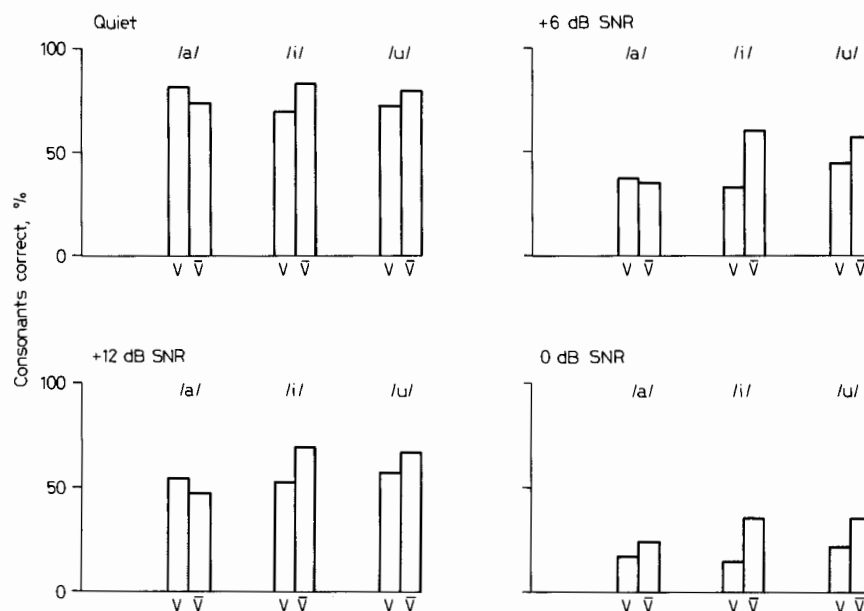
Fig. 5. Consonant recognition, in percent correct, for voiced (V) and voiceless ($\overline{V}$) consonants, as a function of vowel context and listening condition.

ble increased, recognition of the glides/liquids and nasals decreased disproportionately in the /a/ and /u/ contexts.

Stimuli were assigned to two voicing categories based on the presence or absence of vocal fold vibration during consonant production. Only stimuli that were voiced-voiceless cognates were included in the voicing analysis. The voiced stimuli consisted of /b, d, g, v, ð, z/, and the voiceless stimuli consisted of /p, t, k, θ, s, f, ʃ/. Arc sine transformations of the percent-correct voiced and voiceless consonant identifications were submitted to a three-factor ANOVA (vowel × noise × voicing). The results of interest were a significant noise × voicing interaction ($F = 3.023$, $p < 0.01$, d.f. = 3) and a significant vowel × voicing interaction ($F = 26.7$, $p < 0.001$, d.f. = 2). A significant triple inter-

action was not observed. Simple main effects analyses of these interactions revealed a significant voicing effect in the quiet, +6 dB SNR and 0 dB SNR conditions, and in the /i/ and /u/ vowel contexts. Figure 5 presents the mean recognition scores at each noise condition across the three vowel contexts, and indicates that voiceless consonants were identified more accurately than the voiced consonants in each condition where significance was observed.

Consonant Confusions

Another approach to examining the interference effects of MT babble is an analysis of the error patterns that result from its presence. Seven possible confusions can occur among this stimulus set, depending upon the features shared by the stimulus and re-

sponse. Place errors are confusions of stimuli that share voicing and manner characteristics (for example, b/d). Manner errors are those in which the stimulus and response possess the same voicing and place (such as m/b). Voicing errors consist of confusions among voiced-voiceless cognates, sharing the same place and manner of production (e.g., b/p). Errors of place-voicing were among consonants that retain the manner of production in common (b/t); errors of place-manner are confusions of consonants sharing the single feature voicing (p/s); and errors of manner-voicing are among consonants that share the place feature (p/w). Finally, place-manner-voicing errors are confusions of consonants that do not retain any common phonetic features, such as w/k.

Scores for the seven feature errors were determined from each subject's confusion matrices in all listening conditions. A feature error score was calculated by dividing the frequency of that feature error by the total number of errors exhibited by a subject for a particular condition. The feature error scores of all subjects were transformed into arc-sine units and submitted to an ANOVA, in which a three-way interaction (noise × vowel × feature error) was observed ($F = 3.14$, $p < 0.0001$, d.f. = 36). Analyses of simple main effects were conducted to determine how the proportion of each feature error was altered as noise level increased, and to determine which feature error types predominated at each noise level.

The feature error scores in the seven categories, pooled across the 10 subjects at each SNR and vowel context, are presented in figure 6. The effect of varying noise level on the feature error scores can be observed in this figure. Generally, single feature errors (i.e., place, manner, or voicing) are greatest in quiet and decrease as noise level increases, while multiple-feature errors (place-manner, place-voicing, manner-voicing, or place-manner-voicing) are fewest in quiet and increase as noise level increases. However, the proportion of certain feature errors is independent of noise level, including manner errors, place errors for consonants paired with /u/, voicing errors for consonants paired with /a/, place-voicing errors for consonants paired with /i/, and manner-voicing errors for consonants paired with /i/ and /u/.

The feature error patterns at each noise level also are shown in figure 6. In quiet, errors of place, manner, and voicing are significantly higher than all multiple feature errors for consonants produced with the vowels /i/ and /u/ ($p < 0.01$). However, consonants coarticulated with /a/ demonstrate a significantly higher proportion of place errors than any other error type ($p < 0.01$). As the level of the babble increases, voicing errors decrease and place-manner errors emerge. At +12 dB SNR, place and manner errors exceed all other error types, except for place-manner errors, in the /a/ and /u/ vowel contexts ($p < 0.01$). For consonants paired with /i/, only place and manner errors are significantly more frequent than place-voicing and place-voicing-manner errors ($p < 0.01$). Errors of place, manner, and place-manner are significantly greater than all other confusions at +6 dB SNR, regardless of vowel coarticulation ($p < 0.01$). In the most degraded condition, place-manner errors account for approximately one third of all errors, and are more common than all other error types, independent of vowel coarticulation ($p < 0.01$). The single-feature errors of manner and place are also frequent. Errors of place-voicing and voicing are rare in all vowel contexts at 0 dB SNR.
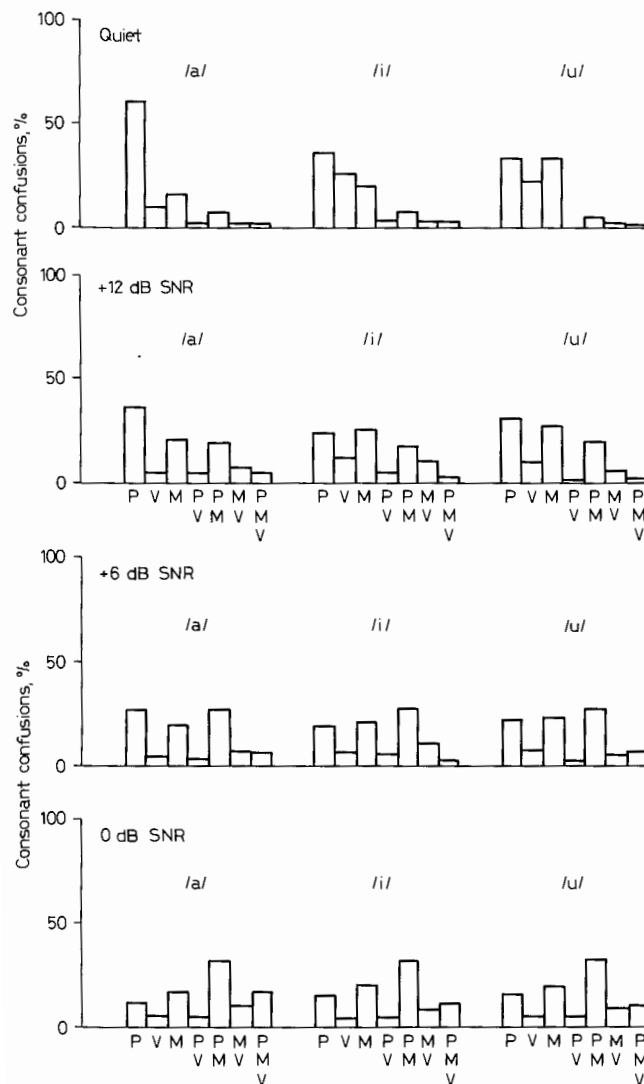
Fig. 6. Consonant confusions, in percent error, for seven feature error types [place (P), voicing (V), manner (M), place-voicing (PV), place-manner (PM), manner-voicing (MV), place-manner-voicing (PMV)], as a function of vowel context and listening condition.

Discussion

MT babble produces certain systematic effects on normal-hearing listeners' recognition of nonsense syllables. These effects often are not consistent with those reported by other investigators for recognition of conso-nants in the presence of other types of noise backgrounds. The current findings suggest that MT babble imposes a unique interference pattern on consonant recognition.

Vowel coarticulation influences performance in MT babble, with consonants followed by /a/ being more difficult to identify

than consonants followed by /i/ or /u/. *Wang and Bilger* [1973] also reported that consonant recognition was affected by vowel context, for CVs presented in white noise. Poorest recognition was reported for consonants followed by /i/, while best recognition was for consonants followed by /u/. For CVs presented in cafeteria noise (with intelligible conversation edited out), *Dubno and Levitt* [1981] found that consonants accompanied by /a/ were easier to identify than those accompanied by /i/ or /u/. Differences in the effects of the accompanying vowel may result either from differences in the noise spectra or differences in stimulus level adjustment procedures. In the present study and in that of *Wang and Bilger* [1973], the peak levels of the CVs were adjusted to be equivalent. This procedure was not applied in the *Dubno and Levitt* study. Variations in the stimulus levels may have contributed to the comparatively high recognition of consonants paired with /a/, because the amplitude of /a/ in natural speech usually is higher than that of other vowels [*Peterson and Barney,* 1952].

The effects of MT babble on the recognition of consonant articulatory features also contrasted with those in other reports. Consonants produced with maximum constriction at the middle of the mouth were recognized more poorly than front and back-place productions in the context of /a/, while recognition of back-place consonants was superior to recognition of front and mid-place consonants in the context of /i/. *Wang and Bilger* [1973] reported the opposite finding: the consonant features 'high-anterior' and 'anterior' (incorporating front and middle places of production) contributed to listeners' identifications of CV syllables in white noise. Differences in the noise spectra may

account for these discrepancies. In cafeteria noise, which has a similar spectrum to the babble, identification of front consonants was poorer than middle and back consonants in all vowel contexts [*Dubno and Levitt,* 1981]. Thus, the current investigation and that of *Dubno and Levitt* are different in the comparative recognizability of mid-place consonants. That back-place consonants were well perceived in the babble is consistent with earlier findings that velar stimuli in place continua are highly perceived in MT babble and speech spectrum noise [*Gordon-Salant and Wightman,* 1983].

Manner classes of speech sounds are perceived differently in babble compared to other noises. In white noise, it consistently has been reported that nasality is well-perceived [*Miller and Nicely,* 1955; *Wang and Bilger,* 1973; *Horii* et al., 1970], as is sibilance [*Wang and Bilger,* 1973; *Mitchell and Singh,* 1974]. In severe levels of white noise, recognition of fricatives reportedly is reduced [*Miller and Nicely,* 1955; *Wang and Bilger,* 1973; *Mitchell and Singh,* 1974; *Horii* et al., 1970]. When babble is present at low levels, glides/liquids and nasals are perceived better than plosives and fricatives/sibilants; while in high levels of babble, fricatives/sibilants are perceived with comparatively high accuracy in the /a/ and /u/ contexts. The reduction in recognition of glides/liquids and nasals in the presence of babble is attributed to direct masking of the low-frequency cues in these consonants by the low-frequency energy of the babble. To determine the source of the comparatively high recogniton of the fricative/sibilant category, an analysis of individual phoneme accuracy was undertaken. This analysis revealed that the voiceless sibilant (ʃ) was re-

cognized with near-perfect accuracy by all subjects in severe levels of babble. This phoneme is characterized by a long duration of (mid-to-high frequency) aperiodic noise, which may be an important cue to its intelligibility in the babble. Further, although the babble itself contains numerous fricative and sibilant sounds, its average spectrum exhibits weak energy in the high frequencies. Thus, the phoneme /ʃ/ does not appear to be masked directly by the high-frequency energy in the babble. Surprisingly, *Dubno and Levitt* [1981] reported that glides and affricates are recognized with higher accuracy than plosives, fricatives (which include sibilants) and nasals in the presence of cafeteria noise. A shift in this pattern may have been observed if performance was assessed at lower SNRs.

Voiceless consonants were perceived better than voiced consonants in the MT babble background. This finding is in direct contrast with the reported voicing effect observed in cafeteria noise [*Dubno and Levitt,* 1981]. Procedural variations in data analysis may have contributed to this discrepancy. Recall that only voiced-voiceless cognate pairs were included in the voicing analysis of the current study. That is, the nasals and glides/liquids were not included in the analysis because they have no voiceless counterparts. However, the glides/liquids and nasals were included in the voicing analysis of the *Dubno and Levitt* study. The high identification of the glides and liquids previously reported may have contributed to the comparatively high recognition of the voiced consonants reported in that study.

The pattern of consonant confusions observed in MT babble was similar to previous reports. Errors of place, manner, and place-manner were frequent in MT babble, while voicing and place-voicing errors were rare. Other investigators have reported a predominance of place errors in white noise [*Miller and Nicely,* 1955; *Wang and Bilger,* 1973] and in cafeteria noise [*Dubno and Levitt,* 1981]; and a low proportion of voicing errors in white noise [*Miller and Nicely,* 1955; *Wang and Bilger,* 1973]. The low incidence of voicing errors in MT babble was contrary to our expectation that the low-frequency spectrum of the babble would directly mask the low-frequency acoustic cues for voicing, such as the relative onset of the fundamental frequency [*Haggard* et al., 1970], and the first formant transition [*Cooper* et al., 1952]. Apparently, the listeners were able to resolve some of the redundant cues to voicing which are inherent in the higher frequencies. Specifically, the relative onsets of broadband aperiodic energy and periodic resonant energy (e.g., the formants) may have served as a cue for voicing in the presence of the babble [*Soli and Arabie,* 1979].

A unique finding among the confusion data was that place-manner errors were most prevalent in severe levels of the babble. In reviewing the data, it was noted that the frequency of place and manner errors remained high in the babble, but the total number of place-manner errors increased sufficiently to cause a shift in the relative proportion of these three error types. Obviously, as the babble increased in level, listeners were unable to detect the acoustic cues for place and manner either separately or in combination. Interestingly, *Dubno and Levitt* [1981] reported that place errors were more common in cafeteria noise than either manner errors or place-manner errors, a finding somewhat different from our own. That place-manner errors were not more common in the *Dubno and Levitt* [1981] study

may be attributed to their use of only one noise condition.

In summary, MT babble produced effects different from those reported earlier for white noise and cafeteria noise on recognition of consonant place, manner, and voicing. Identification of consonants paired with different vowels in MT babble also contrasted with coarticulation effects observed in other noises. Although consonant confusions in low levels of MT babble were comparable to those reported in white noise and cafeteria noise, a different pattern emerged in several levels of the babble. These findings support the notion that consonant recognition and confusion patterns observed in the presence of white noise and cafeteria noise are not generalizable to other types of noise backgrounds. Spectral differences between noises as well as the relative levels between signal and noise are likely sources of variability between performance patterns that emerge in the presence of different competitors.

MT babble is useful in clinical evaluation of word recognition because it demonstrates performance deficits in hearing-impaired listeners and represents realistic listening conditions. The current study has detailed some of the interference patterns obtained with this type of noise background on consonant recognition by normal-hearing subjects. This data base can serve as a normative reference with which to compare possible performance deficits in word and consonant recognition by hearing-impaired listeners. In addition, the current findings may serve as a guide in the development of programs to improve speech recognition in everyday listening environments for hearing-impaired listeners. Specifically, phonetic-level auditory training in speech backgrounds can be struc-

tured in difficulty according to the current findings. For example, recognition training of back-place consonants would precede that of front and mid-place consonants; recognition training of voiceless consonants would precede voiced consonants, etc. The finding that interference effects vary with vowel context suggests that training in one vowel context will not necessarily transfer to other vowel contexts. The same rationale applies to training at different SNRs. Finally, the observed patterns of consonant confusions readily suggest response alternatives in closed-set auditory training tasks that can be ordered in difficulty.

Reconnaissance de syllabes consonne/voyelle synthétiques sur fond de bavardage

La reconnaissance de syllabes sans signification par 10 personnes à l'ouïe normale a été évaluée dans le silence et sur fond de bavardages à trois niveaux différents. Les stimulus ont été dix-neuf consonnes, mises par paires avec /a/, /i/, et /u/ dans un format consonne-voyelle. On a inclus dans les mesures de performance la reconnaissance totale des syllabes, la reconnaissance des caractéristiques et de la classe des consonnes, et les erreurs dans les caractéristiques des consonnes. Les résultats de chaque analyse ont été fortement affectés par la coarticulation des voyelles et par le niveau de bruit. Néanmoins, un modèle distinctif de performance sur fond de bavardage a été observé. Des comparaisons de ce modèle et des modèles obtenus auparavant avec d'autres genres de bruits de fond sont faites. Les implications pour des évaluations cliniques et pour l'entraînement auditif sont discutées.

References

Aniansson, G.: Methods for assessing high frequency hearing loss in every-day listening situations. Acta oto-lar., suppl. 320, pp. 1–50 (1974).

Cooper, F.S.; Delattre, P.C.; Liberman, A.M.;

Borst, J.M.; Gerstman, L.J.: Some experients on the perception of synthetic speech sounds. J. acoust. Soc. Am. *24:* 597–606 (1952).

Danhauer, J.L.; Leppler, J.G.: Effects of four noise competitors on the California Consonant Test. J. Speech and Hearing Disorders *44:* 354–361 (1979).

Dubno, J.R.; Dirks, D.D.: Evaluation of hearing-impaired listeners using a nonsense-syllable test. I. Test reliability. J. Speech and Hearing Res. *25:* 135–140 (1982).

Dubno, J.R.; Levitt, H.: Predicting consonant confusions from acoustic analysis. J. acoust. Soc. Am. *69:* 249–261 (1981).

Findlay, R.C.; Denenberg, L.J.: Effects of subtle mid-frequency auditory dysfunction upon speech discrimination in noise. Audiology *16:* 252–259 (1977).

Garstecki, D.C.; Mulac, A.: Effects of test material and competing message on speech discrimination. J. audit. Res. *3:* 171–178 (1974).

Gordon-Salant, S.M.; Wightman, F.L.: Speech competition effects on synthetic stop-vowel perception by normal and hearing-impaired listeners. J. acoust. Soc. Am. *73:* 1756–1765 (1983).

Haggard, M.; Ambler, S.; Callow, M.: Pitch as a voicing cue. J. acoust. Soc. Am. *47:* 613–617 (1970).

Horii, Y.; House, A.S.; Hughes, G.W.: A masking noise with speech-envelope characteristics for studying intelligibility. J. acoust. Soc. Am. *49:* 1849–1856 (1970).

Kalikow, D.N.; Stevens, K.N.; Elliott, L.L.: Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. acoust. Soc. Am. *61:* 1337–1351 (1977).

Kirk, R.E.: Experimental design: procedures for the behavioral sciences (Brooks/Cole, Belmont 1968).

Miller, G.A.; Nicely, P.E.: An analysis of perceptual confusions among some English consonants. J. acoust. Soc. Am. *27:* 338–352 (1955).

Miner, R.; Danhauer, J.L.: Modified rhyme test and synthetic sentence identification test scores of normal and hearing-impaired subjects listening in multitalker noise. J. Am. audiol. Soc. *2:* 61–67 (1976).

Mitchell, L.M.; Singh, S.: Perceptual structure of 16 prevocalic English consonant sententially embedded. J. acoust. Soc. Am. *55:* 1355–1357 (1974).

Peterson, G.E.; Barney, H.L.: Control methods used in a study of the vowels. J. acoust. Soc. Am. *24:* 175–184 (1952).

Soli, S.D.; Arabie, P.: Auditory versus phonetic accounts of observed confusions between consonants. J. acoust. Soc. Am. *66:* 46–59 (1979).

Wang, M.D.; Bilger, R.C.: Consonant confusions in noise: a study of perceptual features. J. acoust. Soc. Am. *54:* 1248–1266 (1973).

Sandra Gordon-Salant,
University of Maryland,
College Park, MD 20742 (USA)