Department of Veterans Affairs

Journal of Rehabilitation Research and Development Vol. 27 No. 3, 1990
Pages 247–254

Audio-visual consonant recognition with the 3M/House cochlear implant

Jeffrey L. Danhauer, PhD; Fatema B. Ghadialy, MA; Douglas L. Beck, MA; Lisa E. Lucks, PhD; Edward A. Cudahy, PhD
Department of Speech and Hearing Sciences, University of California at Santa Barbara, Santa Barbara, CA 93106; House Ear Institute, Los Angeles, CA 90057

Abstract—Eight experienced 3M/House cochlear implant users' consonant recognition was evaluated with videotaped vowel-consonant vowel lists presented in auditory implant only (A), visual (V), and auditory-visual (AV) conditions. All subjects' scores were better than chance. Results revealed that the AV scores were significantly better than the V scores, which were better than the A scores. Sequential Information Analysis of the consonant errors revealed that different features were transmitted better in each condition. Sonorant and voicing features were transmitted well for the A condition, but features related to high-frequency and place cues were not. Place features were transmitted best in the V condition, but acoustic features were not. Both place and acoustic features were transmitted in the AV condition, but they were influenced most by visual cues.

Key words: *auditory-visual, cochlear implants, consonant recognition, 3M/House.*

INTRODUCTION

The single-channel 3M/House cochlear implant (6,12) has been used in over 800 profoundly deaf patients. Although several studies have examined speech perception by cochlear implant users, only a few have reported consonant recognition data for the 3M/House device in auditory implant only (A), visual (V), and auditory-visual (AV) conditions (5,8,10,15). Studies that have evaluated the 3M/House implant have generally found that the subjects' speech recognition abilities are poor. However, they have usually used only a few subjects tested in the auditory-only condition with various stimuli and contexts, and assumed that most 3M/House users perform poorly with their implants.

This study does not argue the merits of the 3M/House device, but provides additional information to the limited existing database for consonant recognition with this implant. Indeed, few studies have actually looked at A, V, and AV consonant recognition in a controlled fashion (i.e., using videotaped stimuli instead of live-voice face-to-face presentations which vary too much to provide a constant stimulus) with any of the cochlear implant devices that are available. Even fewer studies have analyzed the subjects' consonant errors and presented information about perceptual features obtained from Sequential Information Analyses (SINFA). This study evaluated closed-set consonant recognition by experienced users of the 3M/House device.

METHOD

Subjects were eight postlingually, profoundly deaf persons (7 women and 1 man) between 23 and 60 (M = 42) years of age who had used their 3M/House implants for at least two years. They represented a fair cross section of experienced users of the 3M/House implant, were in good general health, had normal or normally corrected

Address all correspondence and reprint requests to: Jeffrey L. Danhauer, PhD, Department of Speech and Hearing Sciences, 1614 Snidecor Hall, University of California at Santa Barbara, Santa Barbara, CA 93106.

Fatema B. Ghadialy is now at Beltone Electronics Corporation, Chicago, IL; Douglas L. Beck is now at St. Louis Medical Center, St. Louis, MO; Lisa E. Lucks is now at the University of Mississippi, University, MS; Edward A. Cudahy is now at the Lexington Hearing and Speech Center, Jackson Heights, NY.

DOI: 10.1682/JRRD.1990.07.0247

visual acuity (20/40 using a Snellen vision chart), and intelligible speech.

The subjects were volunteers who were available, agreed to participate in testing from a group of about 25 patients meeting the selection criteria, and who were within driving distance to Los Angeles, CA. They were selected from the files of the Otologic Medical Group in Los Angeles and also are to be part of a larger study to be conducted later. They had not previously received the particular stimuli used in this study. Unfortunately, few preoperative speech perception data (and no auditory-only speech recognition scores) were available in the subjects' files. Generally, the subjects met the conditions for implantation with this device (i.e., no speech recognition with the use of traditional hearing aids prior to implantation). Although performance on the stimuli used in this study was not assessed preimplant, based on our experience with other users of this device, these stimuli would probably have been too difficult for them preimplant. The subjects' files revealed that their performance on the House Ear Institute Environmental Sounds tests was 75 percent or better.

Three videotape-recorded General American male-talker 60-item vowel-consonant-vowel (VCV) lists consisting of five random repetitions of each of 12 consonants (/p,b,k,g,ʃ,s,f,v,r,l,m,n/) in an /ʌCʌ/ context were presented via a videotape playback in A, V, and AV conditions. All items were produced in a natural fashion without facial exaggeration. The visual stimuli were close-up face and neck shots. All stimuli were judged to be equally intelligible by three normal-hearing listeners in the A and AV conditions. Stimuli were routed from the videotape playback through a mixer, amplifier, and attenuator, and directly connected to the microphone input of each subject's cochlear implant processor. This procedure bypassed the frequency characteristics of the microphone, eliminated body-baffle effects, and may have provided the subjects with an unfamiliar signal spectrum, but was used to control as much of the variability associated with the signal input as possible. This procedure is similar to the direct-connect condition used by Rosen, Walliker, Brimacombe, and Edgerton (13). Stimuli in the V and AV conditions were directed to 48.5 cm diagonal color video monitor.

Each subject was tested individually in a quiet, well-lighted room and received one stimulus list per presentation mode. Order of stimulus and presentation modes was randomized for each subject. The signal level at the input of the signal processor was adjusted to 3 mV peak-to-peak (~79.5 dB SPL peak). Each subject adjusted the processor volume to a most comfortable level for connected discourse. Subjects received oral and written instructions, and a practice list of the 12 consonants at the beginning of each session. The list of the 12 possible consonants (closed-set) was available to the subjects throughout testing. Subjects' responses were phonetically transcribed and confirmed by the subject before progressing to the next item; guessing was encouraged and no feedback was provided.

RESULTS AND DISCUSSION

Figure 1 shows individual subject data for all conditions. Confidence intervals were calculated using the normal approximation to the binomial distribution to determine a chance score for the 60 items (5 repetitions × 12 consonants) of each list on this test; all scores were better than the 18 percent level required to be significantly above chance at the upper 99 percent confidence limit.

Wilcoxon matched-pairs signed-ranks tests for small samples (14) revealed that the subjects' VCV scores in the AV condition (M = 91 percent) were significantly better than in either the A (M = 47 percent) or V (M = 60 percent) conditions (both $T = 0$, $p \leq 0.01$). Performance in the AV condition was good, ranging from 72 to 100 percent correct as compared to 35 to 60 percent and 48 to 93 percent for the A and V conditions, respectively. All subjects' scores improved in the AV condition, even those for Subject #7 who has exceptionally good speechreading skills. The subjects' scores in the V condition were significantly better than those in the A condition ($T = 1$, $p \leq 0.01$).

The subjects' consonant errors were pooled and converted to confusion matrices for the A, V, and AV conditions as shown in **Table 1, Table 2**, and **Table 3**, respectively. The matrix for each condition was submitted to a Sequential Information Analysis (SINFA) (18), to determine if different types and amounts of information were transmitted by selected *a priori* features. The features specified for this analysis are provided in **Table 4**.

The SINFA program partials out the effects of one feature on another by first estimating the unconditional transmitted information for each feature in the feature system according to the percentage of information transmitted. SINFA then proceeds through a series of iterations in which the feature with the highest percentage of information transmitted is revealed for iteration number one (No. 1). In iteration No. 2, the feature identified in iteration No. 1 is held constant or partialed out, and the conditional transmitted information for the remaining features is determined (i.e., this transmitted information is independent of that for the feature identified in iteration No. 1). The feature having the highest percentage of transmitted

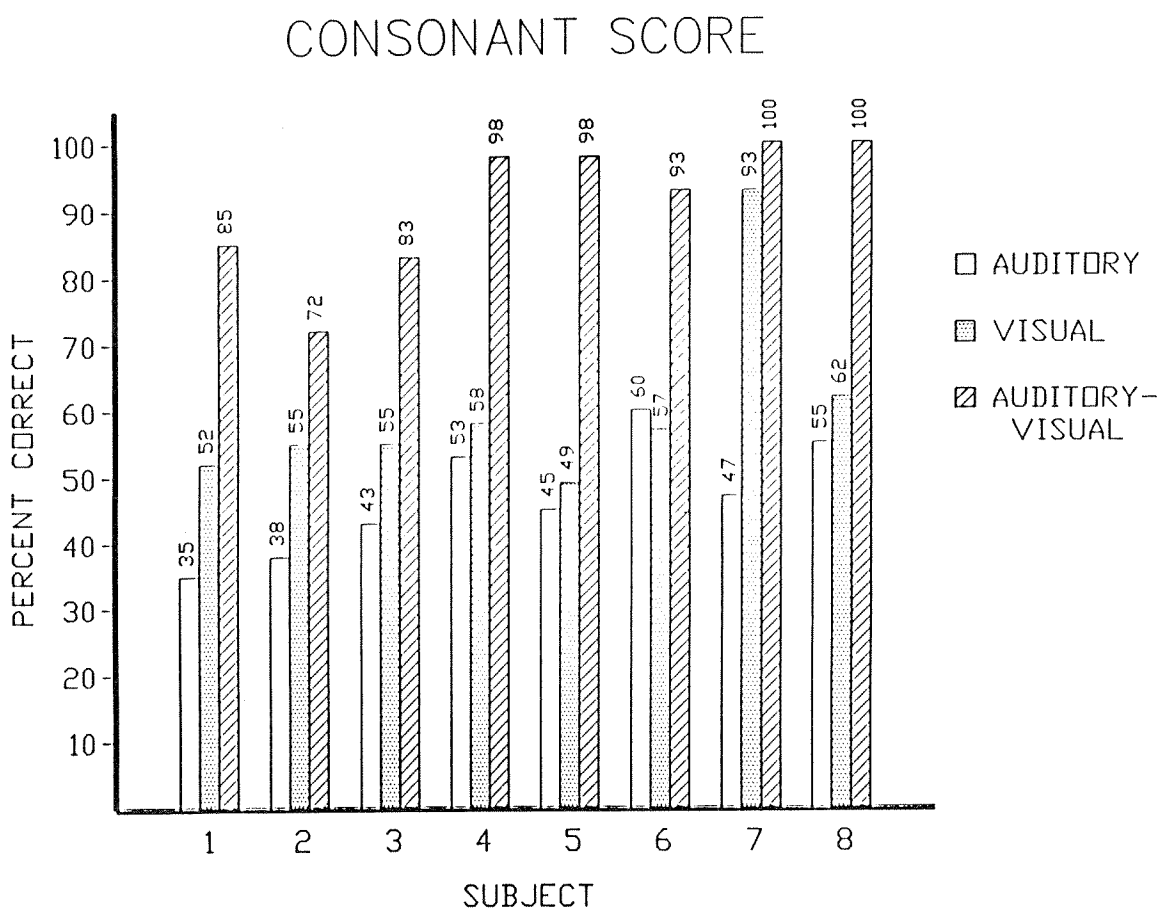

Figure 1.
Individual percent correct consonant recognition scores for the eight cochlear implant users in the auditory, visual, and auditory-visual modes (chance at 99 percent confidence interval > 18 percent).

information in iteration No. 2 is then identified and is held constant for iteration No. 3, and so on.

Table 5 summarizes the SINFA results for the A condition. The stimuli feature information computed for each feature in the system is presented in the first column. The conditional transmitted information in bits for each stimulus for the features is in the second column. The percentage of conditional information transmitted is in column three. SINFA also indicates when redundant information is associated with the features, and determines when remaining features in the system become equivalent. Features depicted in the later iterations are probably of negligible perceptual importance (18).

Table 6 summarizes the SINFA results for the three conditions and shows the features identified and the percent of (conditional) transmitted information for each iteration. The iteration number is included in parentheses. Features without an entry in certain columns were those having negligible contributions to the total transmitted information according to SINFA. Iterations No. 5 and 6 for the V, and No. 3, 6, and 7 for the AV conditions, were redundant as indicated by the same percentages of transmitted information.

Table 6 shows that with the exception of palatal, features transmitted well in the A condition related to sonorancy and voicing. A high degree of redundancy is inherent in these features, as all sonorants are voiced, and the nasals and liquids can also be classified as sonorants. These features involve low-frequency information that should be available through the 3M/House implant. However, high-frequency features (e.g., sibilant) and those related to place of articulation (e.g., front/back, bilabial, dental, alveolar, and velar) were not transmitted well by the implant. The subjects almost always recognized the palatals, /r/ and /ʃ/, correctly in the A condition. They may have been responding primarily to intensity and/or

Journal of Rehabilitation Research and Development Vol. 27 No. 3 Summer 1990

Table 1.
Pooled confusion matrix for auditory condition.

Stimulus \ Response	p	b	k	g	ʃ	s	f	v	r	l	m	n
p	19	1	8	3	2	2	5					
b	2	27		4	1		2	3		1		
k	18		12	1		5	3	1				
g	4	18	2	7		1	4	1	1	2		
ʃ			2		32	6						
s	2	1	2	4	2	15	14					
f	2	2	1	4		15	14	1		1		
v	1	8	1	5		1	4	19				1
r								1	36	3		
l									14	19	3	4
m		1				5		1	2	10	10	11
n						2			2	9	13	14

duration cues for /ʃ/, and to low frequency (e.g., F1 transition) cues for /r/. These two sounds are almost always identified correctly by these subjects in open-set contexts, in and out of test situations. As expected, features transmitted best in the V condition related to place of articulation (e.g., bilabial, dental, etc.), whereas acoustic features (e.g., voicing, sonorant, nasal, etc.) were not transmitted well. Sibilant was identified on iteration No. 5 for the V condition; however, the facial contortions involved in producing the sibilant sounds frequently make these (otherwise acoustic phonemes) highly visible and distinguishable from other phonemes (4).

Interestingly, the features identified in the AV condition resulted in a combination of both auditory and visual features. Sibilant was the first feature identified, but as just discussed, this may have been due more to visual than to acoustic cues. Other acoustic features identified in the A condition (i.e., voicing, sonorant, and nasal) were not identified until iteration No. 6. Thus, the feature information transmitted for these cochlear implant users seemed to be driven more by visual cues available through speechreading than by acoustic cues provided by their implants.

These results are in agreement with those from many other studies that have evaluated consonant perception in

Table 2.
Pooled confusion matrix for visual condition.

Stimulus \ Response	p	b	k	g	ʃ	s	f	v	r	l	m	n
p	16	17					1				6	
b	12	17									11	
k			15	22						1		2
g			12	18					1	3		6
ʃ					35	3						2
s					2	38						
f							27	13				
v							26	14				
r									40			
l				1						39		
m	13	18									9	
n			2	6	1	9				2		20

Table 3.
Pooled consonant confusion matrix for auditory-visual condition.

Stimulus \ Response	p	b	k	g	ʃ	s	f	v	r	l	m	n
p	39	1										
b	2	37						1				
k			37	3								
g			3	36								1
ʃ					37	3						
s					1	39						
f							32	8				
v				1			3	36				
r									40			
l										39		1
m	2	3									35	
n	1			2		2				1		34

Table 4.
Features specified for the sequential information analysis.

Phoneme	VOIC	FR/B	SIBL	SONR	NASL	LIQD	PLOS	FRIC	BILB	DENT	ALVR	PALT	VELR
p	0	0	0	0	0	0	1	0	1	0	0	0	0
b	1	0	0	0	0	0	1	0	1	0	0	0	0
k	0	1	0	0	0	0	1	0	0	0	0	0	1
g	1	1	0	0	0	0	1	0	0	0	0	0	1
ʃ	0	1	1	0	0	0	0	1	0	0	0	1	0
s	0	0	1	0	0	0	0	1	0	0	1	0	0
f	0	0	0	0	0	0	0	1	0	1	0	0	0
v	1	0	0	0	0	0	0	1	0	1	0	0	0
r	1	1	0	1	0	1	0	0	0	0	0	1	0
l	1	0	0	1	0	1	0	0	0	0	1	0	0
m	1	0	0	1	1	0	0	0	1	0	0	0	0
n	1	0	0	1	1	0	0	0	0	0	1	0	0

Table 5.
Sequential information analysis of consonant confusions in the auditory condition.

Iteration No. 1 Feature	Feature Inf.	Trans. Inf.	% Trans. Inf.
VOIC	0.980	0.514	0.525
FR/B	0.918	0.123	0.134
SIBL	0.650	0.193	0.297
SONR	0.918	0.708	0.771
NASL	0.650	0.240	0.369
LIQD	0.650	0.355	0.546
PLOS	0.918	0.335	0.365
FRIC	0.918	0.306	0.333
BILB	0.811	0.057	0.070
DENT	0.650	0.092	0.141
ALVR	0.811	0.078	0.096
PALT	0.650	0.330	0.508
VELR	0.650	0.029	0.044

(**Table 5** continues on the following two pages)

Journal of Rehabilitation Research and Development Vol. 27 No. 3 Summer 1990

Table 5. (continued)

Iteration No. 2
Feature Held Constant: SONR

Feature	Cond. Feat. Inf.	Cond. Trans. Inf.	% Cond. Trans. Inf.
VOIC	0.628	0.257	0.409
FR/B	0.904	0.161	0.178
SIBL	0.536	0.141	0.263
NASL	0.322	0.091	0.282
LIQD	0.322	0.091	0.282
PLOS	0.666	0.173	0.260
FRIC	0.666	0.173	0.260
BILB	0.796	0.064	0.080
DENT	0.541	0.055	0.102
ALVR	0.687	0.041	0.060
PALT	0.628	0.322	0.513
VELR	0.538	0.011	0.021

Iteration No. 3
Features Held Constant: SONR, PALT

Feature	Cond. Feat. Inf.	Cond. Trans. Inf.	% Cond. Trans. Inf.
VOIC	0.566	0.212	0.374
FR/B	0.494	0.007	0.014
SIBL	0.339	0.026	0.078
NASL	0.194	0.023	0.121
LIQD	0.194	0.023	0.121
PLOS	0.573	0.124	0.217
FRIC	0.573	0.124	0.217
BILB	0.710	0.044	0.062
DENT	0.497	0.040	0.080
ALVR	0.551	0.029	0.052
VELR	0.494	0.007	0.014

A, V, and AV conditions by normal-hearing listeners, hearing-impaired hearing aid wearers, and single-channel cochlear implant users (1,3,7,9,10,11,13,16,17). Although direct predictions to conversational speech cannot be drawn from these results, the improvements in consonant scores in the AV condition suggest that the implant should help these subjects perform well in everyday speechreading situations having more content and contextual cues. The subjects' scores in the V condition generally reflected their average-to-excellent speechreading ability. Their performance in the AV condition was probably due to their ability to combine the auditory cues provided by the cochlear implant with visual cues and not just due to speechreading alone.

In summary, this study adds to the database on consonant recognition with cochlear implants and provides

Table 5. (continued)

Iteration No. 4
Features Held Constant: SONR, PALT, VOIC

Feature	Cond. Feat. Inf.	Cond. Trans. Inf.	% Cond. Trans. Inf.
FR/B	0.481	0.016	0.032
SIBL	0.268	0.011	0.039
NASL	0.194	0.023	0.121
LIQD	0.194	0.023	0.121
PLOS	0.554	0.156	0.281
FRIC	0.554	0.156	0.281
BILB	0.697	0.039	0.056
DENT	0.487	0.057	0.116
ALVR	0.477	0.011	0.023
VELR	0.481	0.016	0.032

Iteration No. 5
Features Held Constant: SONR, PALT, VOIC, PLOS/FRIC

Feature	Cond. Feat. Inf.	Cond. Trans. Inf.	% Cond. Trans. Inf.
FR/B	0.311	0.007	0.022
SIBL	0.158	0.000	0.000
NASL	0.194	0.023	0.121
LIQD	0.194	0.023	0.121
BILB	0.521	0.007	0.014
DENT	0.158	0.000	0.000
ALVR	0.367	0.001	0.002
VELR	0.311	0.007	0.022

consonant recognition data for A, V, and AV conditions using the 3M/House implant. Further, it provides findings about how much information was transmitted by each perceptual feature in each condition. Primary features for the A condition were voicing, nasality, and sonorant; those for the V condition related to place of articulation; and those for the AV condition were a combination of those for the A and V conditions plus sibilancy. The study shows which features were and were not transmitted well by this device. Visual features were transmitted better than auditory-only features, but subjects performed very well when both auditory and visual cues were available, even with the 3M/House device that is admittedly simpler than some other implants currently available. Assuming that the other devices are capable of providing more acoustic cues, patients using those devices may be expected to perform even better than patients using the 3M/House device.

ACKNOWLEDGMENTS

This research was supported by the House Ear Institute, a University of California Santa Barbara (UCSB) General Research Grant (#8-587529-19900-7), and a Biomedical Research NIH Grant (#RR07099-18-8-448790-24328) administered by the UCSB Social Processes Research Institute.

Appreciation is expressed to J. Phil Mobely, Michael J. Danley, and Lars B. Nielsen for their assistance with the instrumentation in this study and comments on the manuscript.

Table 6.
Summary of percentage of transferred information from sequential information analysis for auditory, visual, and auditory-visual conditions.

Feature	Auditory	Visual	Auditory-Visual
VOIC	(3) 0.374		(8) 0.605
FR/B		(6) 0.282	(6) 0.806
SIBL		(5) 0.714	(1) 0.957
SONR	(1) 0.771		(7) 0.751
NASL	(5) 0.121		(7) 0.751
LIQD	(5) 0.121	(3) 0.818	(4) 0.916
PLOS	(4) 0.281		(7) 0.751
FRIC	(4) 0.281	(5) 0.714	(3) 0.962
BILB		(1) 0.975	(2) 0.949
DENT		(2) 1.000	(3) 0.962
ALVR		(6) 0.282	(6) 0.806
PALT	(2) 0.513	(4) 0.787	(5) 0.860
VELR		(6) 0.282	(6) 0.806

Special appreciation is expressed to Robert V. Shannon for his assistance with the SINFA analysis of the data.

REFERENCES

1. **Binnie CA, Montgomery AA, Jackson PL:** Auditory and visual contributions to the perception of consonants. *J Speech Hear Res* 17:619-630, 1974.
2. **Brimacombe JA, Edgerton BJ, Doyle KJ, Erratt JD, Danhauer JL:** Auditory capabilities of patients implanted with the House single-channel cochlear implant. *Acta Otolaryngol* S411:204-216, 1984.
3. **Danhauer JL:** Consonant perception by normals in conditions of filtering. *J Am Aud Soc* 4:117-121, 1978.
4. **Danhauer JL, Appel MA:** INDSCAL analysis of perceptual judgments for 24 consonants via visual, tactile, and visual-tactile sensory inputs. *J Speech Hear Res* 19:68-77, 1976.
5. **Danhauer JL, Erratt JD, Edgerton BJ:** Auditory-visual speech perception by cochlear implant users and hearing aid wearers. *Amer J Otol* 7:354-360, 1986.
6. **Danley MJ, Fretz RJ:** Design and function of the single-electrode cochlear implant. *Ann Otol Rhinol Laryngol* 91:21-26, 1982.
7. **Doyle KJ, Danhauer JL, Edgerton BJ:** Features from normal and sensorineural listeners' Nonsense Syllable Test errors. *Ear Hear* 2:117-121, 1981.
8. **Doyle KJ, Danhauer JL, Edgerton BJ:** Vowel perception: Experiments with a single-electrode cochlear implant. *J Speech Hear Res* 29:179-192, 1986.
9. **Edgerton BJ:** Rehabilitation and training of postlingually deaf adult cochlear implant patients. *Semin Hear: Cochlear Implants* 6:65-89, 1985.
10. **Edgerton BJ, Brimacombe JA:** Effects of signal processing by the House/3M cochlear implant on consonant perception. *Acta Otolaryngol* 411:115-123, 1984.
11. **Farran CL, Danhauer JL:** Perceptual features of speech-reading. *Acta Symbolica* 7(9):73-83, 1978.
12. **Fretz RJ, Fravel RP:** Design and function: A physical and electrical description of the 3M/House cochlear implant system. *Ear Hear* 6:14S-19S, 1985.
13. **Rosen S, Walliker J, Brimacombe JA, Edgerton BJ:** Prosodic and segmental aspects of speech perception with the House/3M single-channel implant. *J Speech Hear Res* 32:93-111, 1989.
14. **Siegel S:** *Non-parametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill, 1956.
15. **Tyler RS, Gantz BJ, McCabe BF, et al.:** Audiological results with two single channel cochlear implants. *Ann Otol Rhinol Laryngol* 94:133-139, 1985.
16. **Walden BE, Montgomery AA:** Dimensions of consonant perception in normal and hearing-impaired listeners. *J Speech Hear Res* 18:444- 455, 1975.
17. **Walden BE, Prosek RA, Worthington DW:** Auditory and audiovisual feature transmission in hearing-impaired adults. *J Speech Hear Res* 18:272-280, 1975.
18. **Wang MD, Bilger RC:** Consonant confusions in noise: A study of perceptual features. *J Acoust Soc Am* 54:1248-1266, 1973.