Children's recognition of American English consonants in noise^{a)}

Kanae Nishi,^{b)} Dawna E. Lewis, Brenda M. Hoover, Sangsook Choi, and Patricia G. Stelmachowicz

Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 29 December 2008; revised 9 March 2010; accepted 10 March 2010)

In contrast to the availability of consonant confusion studies with adults, to date, no investigators have compared children's consonant confusion patterns in noise to those of adults in a single study. To examine whether children's error patterns are similar to those of adults, three groups of children (24 each in 4–5, 6–7, and 8–9 yrs. old) and 24 adult native speakers of American English (AE) performed a recognition task for 15 AE consonants in /a/-consonant-/a/ nonsense syllables presented in a background of speech-shaped noise. Three signal-to-noise ratios (SNR: 0, +5, and +10 dB) were used. Although the performance improved as a function of age, the overall consonant recognition accuracy as a function of SNR improved at a similar rate for all groups. Detailed analyses using phonetic features (manner, place, and voicing) revealed that stop consonants were the most problematic for all groups. In addition, for the younger children, front consonants presented in the 0 dB SNR condition were more error prone than others. These results suggested that children's use of phonetic cues do not develop at the same rate for all phonetic features.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3377080]

PACS number(s): 43.71.Es, 43.71.Ft [JES]

Pages: 3177–3188

I. INTRODUCTION

Following the seminal study by Miller and Nicely (1955), extensive research has been conducted regarding consonant recognition in noise for adult listeners [e.g., elderly normal-hearing listeners (Gelfand *et al.*, 1985, 1986), listeners with hearing loss (Wang and Bilger, 1973; Bilger and Wang, 1976; Wang *et al.*, 1978; Dubno *et al.*, 1982), cochlear implant users (Donaldson and Kreft, 2006), and non-native listeners (Singh and Black, 1966; Cutler *et al.*, 2004, 2008; Garcia Lecumberri and Cooke, 2006)]. In contrast, only a few studies have explored children's consonant recognition (Neuman and Hochberg, 1983; Danhauer *et al.*, 1986; Johnson, 2000). Furthermore, no study with children reported detailed confusion patterns as studies with adults did. To date, consonant confusion patterns exhibited by adults and children have not been compared in a single study. While it is possible to perform a meta-analysis using results from previous studies, it is ideal to compare adults and children in a single study using the same stimulus materials and methodology because confusion patterns may be influenced by the number of target consonants and the number of response alternatives (Bell *et al.*, 1989). The present study was designed to compare consonant confusions in noise for both normal-hearing children and adults. In Secs. I A and I B, a brief summary of previous studies of children's consonant confusion, and the design and the goals of the present study are described.

A. Consonant perception by children

Among the few studies on children's consonant perception, four studies (Graham and House, 1971; Neuman and

Hochberg, 1983; Danhauer *et al.*, 1986; Johnson, 2000) are relevant to the present study. To accommodate the cognitive and linguistic skills of young children, Graham and House (1971) used a discrimination task instead of an identification task. Using live-voice stimuli, they examined consonant confusions by 3–4 year olds using 16 consonants embedded in the medial position of a trisyllabic nonsense form /hə'Caðə/. They concluded that “the perceptual behavior of the children studied is similar to that of adults, except that the children produced more errors than an adult is expected to make in a comparable task (p. 565).” Although a confusion matrix was presented, the lack of adult data and the fact that a discrimination task was used for children make comparisons with other studies of consonant confusions difficult.

Neuman and Hochberg (1983) investigated children's consonant recognition in a manner similar to that of Miller and Nicely (1955). They presented 19 American English consonants embedded in vowel-consonant-vowel (VCV) nonsense syllables in reverberant monaural and binaural conditions for 5–13-yr. old children and young adults. A male talker recorded stimuli in a carrier phrase. Performance in quiet was similar across all groups. All groups performed better in binaural condition than monaural condition. The largest binaural listening benefit was found for the youngest group. The results for the binaural reverberant condition revealed that adult-like performance was not achieved until age 13. Specifics of the confusion patterns as a function of age were not reported.

Danhauer *et al.* (1986) used Edgerton–Danhauer nonsense syllable test list A (NST-A; Auditec, St. Louis, MO). NST-A consisted of 19 English consonants and ten vowels in various combinations. Danhauer *et al.* compared the phonetic features used by normal-hearing and hearing-impaired children between six and 12 yrs. of age when perceiving consonants in initial and medial positions at four sensation levels ranging from 25 to 55 dB in quiet. Listeners' responses were

^{a)} Portions of this research were presented at the 154th meeting of the Acoustical Society of America, New Orleans, LA, Nov. 2007.

^{b)} Author to whom correspondence should be addressed. Electronic mail: kanae.nishi@boystown.org

summarized in confusion matrices and were analyzed using a multidimensional scaling technique. Results showed that the confusions made by hearing-impaired children were similar to those of normal-hearing children, but occurred more frequently. They also reported that both hearing-impaired and normal-hearing children used perceptual features (place, voicing, nasality, sonorancy, and sibilancy) similar to the adults. No developmental changes were examined. However, each phoneme did not appear an equal number of times in the word list, nor did they report whether responses were scored correct based on the entire word or only the consonants. Even if only the consonants were scored, it is still not clear how they handled the unequal number of responses per consonant in their analysis.

Using the same materials as [Danhauer et al. \(1986\)](#), [Johnson \(2000\)](#) investigated phoneme recognition in four listening conditions (quiet, noise-only, reverberation-only, and reverberation+noise) by children (6–15 yrs.) and adults. Similar to [Neuman and Hochberg \(1983\)](#), children did not exhibit adult-like performance in noise-only [+13 dB signal-to-noise ratio (SNR)] or reverberation-only conditions until the early teenage years. As with the study of [Neuman and Hochberg \(1983\)](#), this investigation did not report confusion patterns as a function of age.

Collectively, these studies showed that children perform less accurately than adults, and that there are developmental changes in consonant perception in noise and/or reverberation. However, these previous studies do not provide information regarding changes in confusion patterns with age. Although young children's use of acoustic cues have been reported to differ from those of adults depending on the segment to be perceived ([Sussman, 2001](#); [Mayo and Turk, 2004, 2005](#)), developmental changes in consonant confusion patterns have not been previously reported.

B. The present study

The present study was designed to examine adult-child differences in consonant confusion patterns using the same experimental procedure for both groups. Based on the results of previous studies, it was hypothesized that the overall performance of young children would be poorer than that of adults across all SNR conditions, but that overall performance would improve with age. If such adult-child differences were observed, analyses of confusion patterns should reveal whether the differences between adults and children were due to the increase in all types of confusions or an increase in specific confusions.

To test these hypotheses, three groups of children (4–5, 6–7, and 8–9 yrs. old) and adults were asked to identify 15 English consonants embedded in VCV nonsense syllables presented in three SNR conditions. The influence of listener age and SNR was examined for overall performance, confusion patterns, and phonetic features (place, manner, and voicing).

II. METHOD

A. Subjects

Three groups of children (24 each in 4–5, 6–7, and 8–9 yrs. old) and 24 adults (19–41 yrs. old, mean=27.8 yrs.) with normal hearing served as listeners. This age range was chosen to capture broader developmental changes than those reported in previous studies while minimizing the influence of early articulatory development. Participants' gender was counterbalanced within each group. All listeners had thresholds ≤ 20 dB hearing level (HL) for octave frequencies from 250 through 8000 Hz. Although [Johnson \(2000\)](#) reported that 10–11-yr. olds' consonant identification scores in the +13 dB SNR noise-only condition were significantly lower than adults, the speech materials used in her study (CVCVs with multiple vowels) were more complex than in the present study (VCVs in a single vowel context). Thus, it was anticipated that the oldest group in the present study (8–9-yr. olds) would exhibit adult-like performance.

For the children only, two additional screening tests were administered prior to data collection. The Bankson–Bernthal Quick Screen Test of Phonology ([Bankson and Bernthal, 1990](#)) was used to identify children with articulation errors that would influence scoring. In addition, the receptive vocabulary of each child was estimated using the Peabody Picture Vocabulary Test (PPVT-III, form B; [Dunn and Dunn, 1997](#)). All children's articulation was intelligible, and only children with PPVT-III scores within normal limits for age were included.

B. Stimuli

Test materials were VCV nonsense syllables. Each VCV nonsense syllable included one of the 15 English consonants /p/, /t/, /k/, /f/, /s/, /ʃ/, /b/, /d/, /g/, /z/, /v/, /l/, /r/, /m/, and /n/ in an /a/ vowel context. These consonants were chosen because at least 50% of children between 4 and 9 yrs. old are known to produce them reliably ([Sander, 1972](#)). Based on this criterion, the two interdental fricatives, /θ/ and /ð/, were excluded. VCVs rather than CVs or VCs were chosen to avoid omission responses as well as to preserve relative duration cues among the consonants. An adult female speaker of general American English served as the talker in this study. Recordings took place in a double-walled sound booth. A condenser microphone (AKG Acoustics C535 EB, Vienna, Austria) was placed approximately 30 cm from the speaker's mouth and routed to a preamplifier (Shure M267, Niles, IL). Test stimuli were digitally recorded (CARDDELUXE) at a sampling rate of 44.1 kHz (16 bits, monaural). The talker read a randomized list of VCV nonsense syllables several times, and responses were saved as a single wavefile. Care was taken to ensure that alveolar stops were not flapped, and the distinction between /ada/ and /ata/ tokens was maintained. One of the authors listened to the individual VCV stimuli and identified the best token (no unwanted noise, clearest enunciation without exaggeration, and highest amplitude) from this wavefile. Tokens were equated in overall root-mean-square (rms) amplitude and mixed with speech-shaped noise to yield stimulus files at three SNRs (0, +5, and +10 dB). A quiet condition was not included be-

cause, in quiet, the of 5-year olds' perception of naturally produced VCVs has been reported to be equivalent to that of adults (Neuman and Hochberg, 1983). Although negative SNRs have been used in previous studies with adult listeners, these were not included in the present study because previous studies have shown that while adults can understand familiar spoken materials at 0 dB SNR, young children require a more favorable SNR than adults (see Nelson and Soli, 2000 for review). Specifically, when grade 1 students were tested in various classrooms, their mean score on the Word Intelligibility by Picture Identification test (WIPI; Ross and Lerman, 1971) at -5 dB SNR was 56.5% (Bradley and Sato, 2008). Finitzo-Hieber and Tillman (1978) also reported relatively poor performance (60%) for CVC monosyllabic real words at 0 dB SNR for older children (8–12 yr. olds). Similarly, Stuart *et al.* (2006) reported chance performance for 4–5-yr. old children at -10 dB SNR using stimuli from the Northwestern University-Children's Perception of Speech test (NU-CHIPS, Auditec, St. Louis, MO). Considering that the materials used in the present study were nonsense syllables, the youngest children were expected to have considerable difficulty at negative SNRs.

C. Procedures

The experimental procedures employed were similar to those of Neuman and Hochberg (1983), in which children repeated VCV nonsense syllables as heard. Nonsense syllables were used in lieu of real words or meaningful sentences to minimize the effects of lexical knowledge. Neuman and Hochberg (1983) used nonsense syllables with the same structure with children as young as 5 years old. However, unlike their study (19 consonants $\times$ 3 vowel contexts $\times$ 2 repetitions in four listening conditions), in the present study, 15 consonants were embedded in an /a/ vowel context only, and each nonsense syllable was presented only once in each of the three SNR conditions. The lack of multiple presentations was necessary to accommodate the youngest children's attention span.¹

Listeners were individually tested in a sound booth. Stimuli were presented binaurally via earphones (Sennheiser 25D, Old Lyme, CT) at a fixed overall rms level of 50 dB SPL. Stimulus presentation and response acquisition was controlled by a computer program developed at Boys Town National Research Hospital [Behavioral Auditory Research Tests (BART)].

The same 16-alternative response system (15 target consonants and "other") was used for children and adults. Children were instructed to repeat each VCV as heard. An experimenter seated beside each child listened to their oral response and voted for the children using the 16-item closed-set displayed on a touch-screen monitor. Some of the older children with literacy responded using the touch screen or mouse, but they were also asked to repeat the VCVs before voting so that the experimenter could verify their response and note any discrepancies. Visual reinforcement (pictures of cute animals, etc.) on a computer monitor was given immediately after each response. This reinforcement was used only to maintain the children's interest in the task and was

FIG. 1. Overall percent-correct consonant recognition scores for 4–5, 6–7, and 8–9 yr. olds, and adults in 0 dB, +5, and +10 dB SNR conditions. Error bars represent ± 1 SD from the mean.

not contingent upon correct responses. Each child completed a block of 45 trials (15 consonants $\times$ 3 SNRs). All stimuli were randomized within a block. Due to the limited attention span of the youngest group and the length of the session for the parent project, each child heard each exemplar only once. The procedures for the adult listeners were the same as for the children; except that they were not asked to repeat the VCVs, but used a mouse to indicate their responses while the experimenter monitored the progress of testing from outside the sound booth.

III. RESULTS

A. Overall performance

The first analysis compared overall percent-correct consonant recognition for the three groups of children and the adults for the three SNR conditions (Fig. 1, also see Appendix for confusion matrices). Each age group is represented by different symbols and error bars show ± 1 standard deviation (SD) from the mean.

As can be seen, overall performance improved as a function of both age and SNR. In addition, differences among the groups appeared to increase with decreasing SNR.

To statistically evaluate the above observations, all percent-correct scores were converted to rationalized arcsine units (RAUs; Studebaker, 1985) and were subjected to a group $\times$ SNR mixed-design analysis of variance (ANOVA). Prior to analysis, the assumption of sphericity was confirmed by non-significant Mauchley's $W[W(2)=0.99, p < 0.52]$. Results of ANOVA revealed that only the main effects were significant (group [$F(3, 92)=13.57, p < 0.001, \eta_p^2=0.31$]; SNR [$F(2, 184)=74.01, p < 0.001, \eta_p^2=0.45$]; group $\times$ SNR interaction [$F(6, 184)=0.53, p < 0.78, \eta_p^2=0.02$]). Tukey's Honestly Significant Difference (HSD) test revealed that performance for 4–5 and 6–7 yr. olds was significantly poorer than that of adults ($p < 0.001$ and $p < 0.002$, respectively). A significant difference was also found between 4–5 and 8–9 yr. olds ($p < 0.001$). Furthermore, across all groups, performance improved significantly as the SNR increased ($p < 0.001$ for all comparisons).

TABLE I. Phonetic feature coding.

Feature	Values	Phonemes
Manner	Stop	/p/, /t/, /k/, /b/, /d/, /g/
	Fricative	/f/, /s/, /ʃ/, /v/, /z/
	Liquid	/l/, /r/
	Nasal	/m/, /n/
Place	Front	/p/, /f/, /b/, /v/, /m/
	Mid	/s/, /t/, /z/, /d/, /l/, /n/
	Back	/k/, /g/, /ʃ/, /r/
Voicing	Voiced	/b/, /d/, /g/, /v/, /z/, /l/, /m/, /n/, /r/
	Voiceless	/p/, /t/, /k/, /f/, /s/, /ʃ/

B. Types of confusions

To explore the confusion patterns in more detail and to examine potential developmental trends in the use of acoustic-phonetic cues, a set of three phonetic feature values (manner, place, and voicing) were assigned to each of the 15 consonants (see Table I). Note that this coding system generally followed that of [Cutler et al. \(2004\)](#) except that the place feature had only three levels (front, mid, and back). The front place included labial consonants (/p/, /f/, /b/, /v/, and /m/), the mid place included alveolar consonants (/s/, /t/, /z/, /d/, /l/, and /n/), and the back place included velar (/k/ and /g/) and palatal (/ʃ/ and /r/) consonants.

Based on the coding shown in Table I, all confusions were analyzed using a method similar to that of [Dubno and Levitt \(1981\)](#). First, the 210 possible confusions (15 target consonants $\times$ 14 incorrect response alternatives) in the confusion matrix were classified into one of the seven types: place, manner, voicing, place+manner (pl/mn), place+voicing (pl/vc), manner+voicing (mn/vc), and place+manner+voicing (pl/mn/vc). For example, both /d/ responses for the target /b/ and /b/ responses for the target /d/ (/b/-/d/ confusions) were classified as “place confusions,” /b/-/v/ confusions as “manner confusions,” /b/-/p/ confusions as “voicing confusions,” /b/-/s/ confusions as “place/manner/voicing confusions,” and so on. No type was assigned to the “other” responses because the frequency of occurrence was low for all groups (1.7% for children and 0.3% for adult). Only the responses that exceeded chance performance (1/16=6.25%) were tallied for this analysis. These results are summarized in Fig. 2.

Individual panels in Fig. 2 show the results for the three SNR conditions. Age groups are represented by different bars, and confusion types are shown along the horizontal axis. A missing bar indicates a count of zero.

There are several notable trends. First, all types of confusions decreased as SNR improved. Manner confusions decreased with SNR improvement but not as much as the other confusion types. On average, the adults made fewer manner confusions than any of the child groups, even though manner confusions were the only type of errors that the adult listeners made at 10 dB SNR. Second, place confusions appeared to show a developmental trend as a function of age and SNR. At 0 dB SNR, all groups made place confusions. At 5 dB SNR, adults did not make any place confusions and, at 10 dB SNR, only the two youngest groups made place confusions.

FIG. 2. Frequency of place, manner, voice, pl/mn, and mn/vc confusions for 4–5, 6–7, and 8–9 yr. olds, and adults in 0 dB, +5, and +10 dB SNR conditions.

All groups made pl/mn confusions at 0 dB SNR, but only the youngest group made pl/mn confusions at the higher two SNRs. Lastly, the occurrence of voicing confusions were similar for all groups even at the poorest SNR. No above-chance voicing confusions were observed for any age group at 10 dB SNR, suggesting that the voicing confusions observed at 0 dB SNR were most likely due to the background noise. These results suggest that, regardless of listeners’ age, perception of manner cues are most vulnerable to the presence of background noise followed by the place cues. Children’s voicing perception was adult-like at ages 4–5 yrs.

These results differ from previous results for adult listeners, where the error rate for place cues was found to be highest in the presence of various background noise (e.g., [Miller and Nicely, 1955](#) for white noise; [Dubno and Levitt, 1981](#); [Cutler et al., 2004](#) for speech-shaped noise). Therefore, to examine whether confusions between specific consonants are responsible for the increased manner confusions and to determine whether there are age-related trends, manner and place confusions were examined in more detail. Table II presents the breakdown for the manner (top half) and the place (bottom half) error frequencies observed for all age groups across the three SNR conditions. To be consistent with Fig. 2, reciprocal confusions (e.g., /b/ responses for /v/ stimulus and vice versa) were pooled. The last row for each half shows the total number of confusions (as shown in Fig. 2) in each SNR condition for each listener group.

For all age groups, manner confusions between /b/ and /v/ (front place) were far more frequent than other confusions. Another notable point is that even though overall error rates decreased with improved SNR, this was not the case for the /b/-/v/ confusions. Interestingly, including the /b/-/v/ confusions, the majority of manner confusions were between consonants produced in the front place (/b/-/v/, /p/-/f/, /b/-/

TABLE II. Number of manner (top) and place (bottom) confusions observed for the four listener groups in the three SNR conditions (0, +5, and +10 dB). Missing values indicate no confusions.

Manner	4–5 yrs.			6–7 yrs.			8–9 yrs.			Adults		
	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
/b/-/v/	12	11	13	14	16	14	13	15	13	11	11	6
/p/-/f/	5	1		5			3			3	3	
/b/-/m/	2	1		2	2	1	1			4		
/d/-/n/	4		1	2			1			2		
/m/-/v/	2		2	1	2							
/z/-/d/			1									
/k/-/f/				1								
/g/-/r/				3								
/z/-/l/					1							
Total	25	13	17	28	21	15	18	15	13	20	14	6
Place	4–5 yrs.			6–7 yrs.			8–9 yrs.			Adults		
	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
/b/-/d/							1					
/p/-/t/				1							1	
/b/-/g/	1			1		1	1	2				
/p/-/k/	3	1		4		1	4				1	
/d/-/g/	3		2				1	1				
/t/-/k/	2			1			2			2		
/f/-/s/				1								
/s/-/f/	5	3	4	2	3	3		4		1	1	
/m/-/n/	2	2	1	2	2					1		
/l/-/r/	3	2	4		1	1		1		3		
Total	20	8	11	12	6	6	9	8	0	7	3	0

m/, and /m/-/v/). In addition, six out of 10 place confusions were between stops (/b/-/d/, /p/-/t/, /b/-/g/, /p/-/k/, /d/-/g/, and /t/-/k/), with no notable differences among the confusions across the three age groups of children. Three other place confusions (/s/-/f/, /m/-/n/, and /l/-/r/) observed for the younger children did not diminish substantially with increasing SNR. These results suggest that place confusions may be the result of an interaction between the effects of noise and age, whereas the high rate of manner confusions may be due to the combination of background noise and the structure of stimulus materials (VCV). Specifically, the most prominent acoustic distinction between /b/ and /v/ (i.e., stop closure versus frication) might have been weakened by the intervocalic position as compared to the CV stimuli used in the studies with adults (Miller and Nicely, 1955; Cutler *et al.*, 2004), which was further obscured by the presence of the background noise. This explanation is consistent with the fact that the /b/-/v/ confusions were not unique to the present study. That is, although the error rates were lower, previous studies with adults also have reported that the /v/ for /b/ confusion was most frequent in SNR conditions comparable to those in the present study (Miller and Nicely, 1955; Cutler *et al.*, 2004).

C. Recognition of acoustic-phonetic cues

To quantitatively evaluate the above observations, all responses were scored as correct or incorrect for each feature set across age group and SNR conditions. These results are summarized for each feature set and presented in Figs. 3–5.

In these figures, the four listener groups are represented by bars with different patterns. Error bars show 1 SD from the mean.

Figure 3 presents the results for the place cue. The results resembled the overall percent-correct data presented in Fig. 1, suggesting that the performance of all groups improved as a function of both SNR and age. In addition, some

FIG. 3. Percent-correct recognition of place features (front, mid, and back) for 4–5, 6–7, and 8–9 yr. olds, and adults in 0 dB, +5, and +10 dB SNR conditions. Error bars show 1 SD from the mean.

FIG. 4. Percent-correct recognition of manner features (stop, fricative, nasal, and liquid) for 4–5, 6–7, and 8–9 yr. olds, and adults in 0 dB, +5, and +10 dB SNR conditions. Error bars show 1 SD from the mean.

differences were observed among the three places such that the mid place cue was perceived most accurately, followed by the front place, and the back place was the poorest.²

Results for manner cues are shown in Fig. 4. Similar to the results for the place cue, listeners' performance improved with both age and SNR. Across all groups, the recognition of stop features appeared to be poorer than nasal or liquid features, possibly due to the high rate of confusions between /b/ and /v/.

Figure 5 presents the results for the voicing cue. Unlike place or manner cues, the overall performance for the voicing cue was relatively high for all groups. Still, some differences among the groups were observed in 0 dB SNR and possibly in +5 dB SNR conditions.

FIG. 5. Percent-correct recognition of voicing features (voiced and voiceless) for 4–5, 6–7, and 8–9 yr. olds, and adults in 0 dB, +5, and +10 dB SNR conditions. Error bars show 1 SD from the mean.

The above observations were confirmed by separate feature value $\times$ group $\times$ SNR ANOVA performed for each of the feature sets. Significant effects were further examined using Tukey's HSD test.

For all features, ANOVAs revealed significant effects of group [place: $F(3,92)=12.69$, $p<0.001$, $\eta_p^2=0.29$; manner: $F(3,92)=12.41$, $p<0.001$, $\eta_p^2=0.29$; voicing: $F(3,92)=5.96$, $p<0.001$, $\eta_p^2=0.16$] and SNR [place: $F(2,184)=31.15$, $p<0.001$, $\eta_p^2=0.25$; manner: $F(2,184)=46.03$, $p<0.001$, $\eta_p^2=0.33$; voicing: $F(2,184)=37.06$, $p<0.001$, $\eta_p^2=0.29$]. Across the features, 4–5 yr. olds performed more poorly than all other groups ($p<0.05$ against 6–7 yr. olds, $p<0.01$ against 8–9 yr. olds and adults for all features). The 6–7 yr. olds' performance was significantly poorer than adults for place and manner features ($p<0.05$ for both). As for the SNR effect, for all features, performance at 0 dB SNR was poorer than at other SNRs ($p<0.001$ for all). In addition, recognition of manner and voicing features at +5 dB SNR was poorer than at +10 dB SNR ($p<0.05$ for manner, $p<0.001$ for voicing). These results parallel those of the overall percent-correct data (Sec. III A). Additional significant effects are presented below.

The ANOVA for the place feature revealed a significant main effect of place [$F(2,184)=9.18$, $p<0.001$, $\eta_p^2=0.09$], and two interactions (place $\times$ SNR [$F(4,368)=3.40$, $p<0.05$, $\eta_p^2=0.04$] and group $\times$ place $\times$ SNR [$F(12,368)=1.81$, $p<0.05$, $\eta_p^2=0.06$]).

Interestingly, an examination of the group $\times$ place $\times$ SNR interaction revealed that the significant difference was associated with the front place only. Specifically, in the 0 dB SNR condition, 4–5 yr. olds recognized front place cues less accurately than adults ($p<0.001$) or 8–9 yr. olds ($p<0.05$). The 6–7 yr. olds' recognition of the front place cues was also less accurate than adults ($p<0.001$) at 0 dB SNR. At +5 dB SNR, only the 4–5 yr. olds performed more poorly on front place cues than adults ($p<0.01$).

The ANOVA for the manner feature also revealed a significant main effect of manner [$F(3,276)=8.49$, $p<0.001$, $\eta_p^2=0.08$], but no interaction reached significance. Further analyses of the manner effect showed that the recognition of stops was poorer than the other manner features ($p<0.05$ for stop versus fricative; $p<0.001$ for stop versus liquids or nasals).

The results of a voicing $\times$ group $\times$ SNR ANOVA revealed significant group $\times$ SNR interaction [$F(6,184)=3.60$, $p<0.005$, $\eta_p^2=0.10$] and voicing $\times$ SNR interaction [$F(2,184)=8.82$, $p<0.001$, $\eta_p^2=0.09$]. Tukey's HSD test revealed that the voiceless consonants were more accurately recognized than the voiced consonants at 0 dB SNR. An examination of the group $\times$ SNR interaction showed that 4–5 yr. olds performed more poorly than adults at 0 dB SNR ($p<0.001$) and 8–9 yr. olds at +5 dB SNR ($p<0.02$).

To summarize, these results revealed several developmental trends. The 4–5 yr. olds were less accurate than the other three groups in recognizing cues for all three phonetic features in noise. For place and manner cues, even 6–7 yr. olds did not perform as well as adults. These age differences were most prominent in the 0 dB SNR condition, while no significant age differences were observed for any of the fea-

tures at +10 dB SNR. Taken together, these results suggest that children younger than 7 yrs. of age tend to have difficulty perceiving stop consonants produced in the front place for SNRs < +5 dB.

IV. DISCUSSION

A. Summary and implications

The present study was conducted to document developmental changes in the perception of consonants presented in noise. The confusion patterns for typically developing children with normal hearing have no precedents. As for the effects of noise at different developmental stages, results showed that while overall performance of younger children is poorer than adults or older children in all SNR conditions, the *relative* decline in performance as a function of decreasing SNR was similar for the adults and children.

Consistent with previous studies, the present results showed that children's errors are similar to those of adults. However, the children in the present study did not show a uniform increase in the frequency of errors across all error types when compared to adults. Rather, they made some types of errors more frequently than others. Specifically, in the 0 dB SNR condition, younger children's recognition of stop consonants produced in the front place was considerably poorer than that of adults or older children. Detailed analyses of confusions associated with these features showed that the majority of confusions were between /b/ (voiced labial stop) and /v/ (voiced labio-dental fricative).

As was suggested in the Introduction, the confusion patterns reported in the present study may be used to evaluate whether children of various ages have acquired the ability to use phonetic cues. In fact, Moeller *et al.* (2007) suggested that the development of acoustic-phonetic awareness may be influenced by hearing status and some phonetic features may be more vulnerable than others. They followed early vocalizations of children with normal-hearing and hearing loss in a 5-yr. longitudinal study and reported that fricatives and affricates emerge later than the other phonemes, even for children with normal hearing. As expected, children with hearing loss showed an overall delay, but the rate of development was parallel to the normal-hearing peers for all phonemes except fricative/affricate consonants. For the normal-hearing group, there was a 19% increase in fricatives/affricates production between 10 and 24 months of age. In the same time period, the hearing-impaired children's production of fricatives/affricates increased only 3%. Although it is not clear how much of this delay is due to the development of acoustic-phonetic awareness, the development of articulatory motor control, or reduced audibility in the high frequencies, identifying atypical perceptual confusions should allow clinicians to distinguish confusions associated with hearing loss from those associated with a specific developmental stage and provide an adequate intervention to facilitate age-appropriate language development.

B. Limitations and future directions

Even though the present study was the first to compare normal-hearing adults and children on consonant recognition

in noise in a single study using a method comparable to those established for adults, there are some limitations. First, in the present study, a cross-sectional design was used. It is possible that different results might be observed in a study using a longitudinal design. Second, as was noted in Sec. II, multiple presentations were not provided for each consonant in each SNR condition as is typically done in studies with adults. This also precluded analyzing the confusion patterns using information theoretical analysis employed in studies with adults (Miller and Nicely, 1955; Cutler *et al.*, 2004; Donaldson and Kreft, 2006; Hu and Loizou, 2007). It is important to note, however, the pattern of confusions across the 72 children was relatively consistent, suggesting that the results are likely to be representative of young normal-hearing children in this age range. As such, the confusion patterns exhibited by the normal-hearing children in this study may be of value when assessing consonant confusions for children with hearing loss. Finally, the present study used VCV nonsense syllables that contained a single vowel /a/ and only 15 consonants known to be in the speech repertoire of children aged between 5 and 9 yrs. As previous studies with adults have suggested, if more consonants are included and more vowel contexts are used, or if other forms (e.g., CV, VC) of stimuli are used, different confusion patterns are likely to be observed.

V. CONCLUSIONS

The present study was the first attempt to compare consonant confusions by adults and children using the same methodology. Fifteen consonants in American English were embedded in VCV nonsense syllables and presented in noise at three SNRs. Results showed that performance of all groups improved at a similar rate as a function of SNR. As expected, children performed more poorly than adults, but their overall performance improved with age. However, detailed examinations of confusion patterns showed that some error types were observed more frequently for younger children than for older children or for adults. These results suggest that children do not acquire the ability to recognize all phonetic features at the same time, and that appropriate care should be taken when interpreting young children's speech perception performance.

ACKNOWLEDGMENTS

This research was supported by National Institute of Deafness and Other Communication Disorders under Grant No. R01 DC004300 to P.G.S. and Grant No. P30 DC004662 to Boys Town National Research Hospital. Merry Spratford and Rindy Ito assisted with data collection with support from Grant No. T35 DC008757.

APPENDIX: CONFUSION MATRICES

Tables III–XIV give confusion matrices for each group at the three SNR (0, +5, and +10 dB) conditions.

TABLE III. Percentages of responses pooled over 24 children (4–5 yr. olds) for 15 consonants at 0 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	66.7	4.2	4.2	12.5						4.2		4.2			
/t/		87.5						8.3							
/k/	8.3	8.3	66.7	4.2	8.3		4.2								
/f/	8.3		4.2	54.2			12.5			8.3					4.2
/s/			4.2		83.3	12.5									
/ʃ/					8.3	91.7									
/b/	4.2			20.8	4.2		29.2		4.2	8.3		8.3		4.2	4.2
/d/								87.5	4.2				8.3		
/g/			4.2					8.3	75.0			4.2			
/v/							41.7			50.0					
/z/					12.5	4.2					79.2				
/m/										8.3		58.3	8.3	4.2	16.7
/n/								8.3					87.5		
/l/	4.2						4.2			12.5		4.2		66.7	8.3
/r/	4.2						8.3							4.2	75.0

TABLE IV. Percentages of responses pooled over 24 children (6–7 yr. olds) for 15 consonants at 0 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	75.0	4.2	8.3	8.3									4.2		
/t/		100													
/k/	8.3	4.2	79.2	4.2		4.2									
/f/	12.5		4.2	50.0	4.2		4.2			20.8		4.2			
/s/					100										
/ʃ/					8.3	91.7									
/b/	4.2			12.5		4.2	12.5		4.2	37.5		4.2	4.2		16.7
/d/								87.5		4.2			8.3		
/g/									83.3	4.2					12.5
/v/							20.8			70.8				4.2	4.2
/z/											95.8				4.2
/m/							4.2			4.2		70.8	8.3	4.2	4.2
/n/													95.8		4.2
/l/										4.2		4.2		91.7	
/r/							4.2	4.2		8.3					75.0

TABLE V. Percentages of responses pooled over 24 children (8–9 yr. olds) for 15 consonants at 0 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	87.5		4.2	8.3											
/t/		100													
/k/	12.5	8.3	75.0						4.2						
/f/	4.2			83.3						12.5					
/s/					95.8						4.2				
/ʃ/						100									
/b/	8.3			20.8			41.7	4.2		16.7		4.2			4.2
/d/								91.7	4.2				4.2		
/g/							4.2		91.7						4.2
/v/				4.2			37.5			58.3					
/z/					4.2						95.8				
/m/												91.7			4.2
/n/													100		
/l/				4.2			8.3							87.5	
/r/							8.3			12.5					79.2

TABLE VI. Percentages of responses pooled over 24 adults for 15 consonants at 0 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	95.8			4.2											
/t/		95.8						4.2							
/k/		8.3	91.7												
/f/	8.3			83.3						8.3					
/s/					100										
/ʃ/					4.2	95.8									
/b/				8.3			45.8			33.3		12.5			
/d/								91.7					8.3		
/g/									100						
/v/				8.3			12.5			79.2					
/z/					8.3						91.7				
/m/							4.2					83.3		12.5	
/n/												4.2	95.8		
/l/										4.2		8.3		87.5	
/r/										4.2				12.5	83.3

TABLE VII. Percentages of responses pooled over 24 children (4–5 yr. olds) for 15 consonants at +5 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	87.5		4.2	4.2											
/t/		95.8													
/k/			83.3						12.5						
/f/				66.7			4.2			20.8					
/s/					91.7										
/ʃ/	4.2				12.5	83.3									
/b/							54.2			20.8	4.2	4.2			8.3
/d/								100							
/g/									95.8					4.2	
/v/	4.2						25.0			66.7					
/z/											95.8				
/m/												70.8	8.3	12.5	
/n/					4.2								95.8		
/l/										8.3		4.2		83.3	4.2
/r/					4.2					4.2				4.2	79.2

TABLE VIII. Percentages of responses pooled over 24 children (6–7 yr. olds) for 15 consonants at +5 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	100														
/t/		95.8						4.2							
/k/			95.8						4.2						
/f/				87.5			4.2			4.2	4.2				
/s/					95.8		4.2								
/ʃ/					12.5	87.5									
/b/		4.2					37.5			50.0		8.3			
/d/								100							
/g/									100						
/v/							16.7			83.3					
/z/											91.7			4.2	4.2
/m/										8.3		79.2	8.3		4.2
/n/													100		
/l/		4.2												95.8	
/r/	4.2										4.2			4.2	87.5

TABLE IX. Percentages of responses pooled over 24 children (8–9 yr. olds) for 15 consonants at +5 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	100														
/t/		100													
/k/			91.7						8.3						
/f/				87.5			12.5								
/s/					95.8	4.2									
/ʃ/					12.5	87.5									
/b/							45.8		8.3	45.8					
/d/								100	4.2						
/g/									95.8						
/v/							16.7			79.2					4.2
/z/											100				
/m/												100			
/n/													100		
/l/							4.2							95.8	
/r/														4.2	95.8

TABLE X. Percentages of responses pooled over 24 adults for 15 consonants at +5 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	100														
/t/	4.2	91.7						4.2							
/k/	4.2		91.7						4.2						
/f/	12.5			79.2			4.2							4.2	
/s/					100										
/ʃ/					4.2	95.8									
/b/				4.2			62.5			33.3					
/d/								100							
/g/									100						
/v/				8.3			12.5			79.2					
/z/					8.3						91.7				
/m/	4.2											95.8			
/n/													100		
/l/														100	
/r/															100

TABLE XI. Percentages of responses pooled over 24 children (4–5 yr. olds) for 15 consonants at +10 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	95.8						4.2								
/t/		100													
/k/			100												
/f/				95.8											4.2
/s/	4.2				87.5	8.3									
/ʃ/					8.3	87.5									4.2
/b/							70.8			29.2					
/d/								100	8.3						
/g/									91.7						
/v/							25.0			70.8					4.2
/z/								4.2			87.5				8.3
/m/										8.3		83.3		8.3	
/n/								4.2				4.2	91.7		
/l/										4.2				91.7	4.2
/r/														12.5	79.2

TABLE XII. Percentages of responses pooled over 24 children (6–7 yr. olds) for 15 consonants at +10 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	95.8					4.2									
/t/		100													
/k/	4.2		95.8												
/f/			4.2	95.8											
/s/	4.2				95.8										
/ʃ/					12.5	87.5									
/b/							37.5		4.2	54.2		4.2			
/d/								95.8							4.2
/g/									100						
/v/			4.2				4.2			91.7					
/z/											95.8				4.2
/m/									4.2			95.8			
/n/													100		
/l/														100	
/r/														4.2	95.8

TABLE XIII. Percentages of responses pooled over 24 children (8–9 yr. olds) for 15 consonants at +10 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	100														
/t/		100													
/k/			95.8												4.2
/f/				95.8											4.2
/s/					100										
/ʃ/						100									
/b/							66.7			29.2				4.2	
/d/								100							
/g/	4.2								95.8						
/v/							25.0			75.0					
/z/											100				
/m/												100			
/n/													100		
/l/										4.2				95.8	
/r/															100

TABLE XIV. Percentages of responses pooled over 24 adults for 15 consonants at +10 dB SNR.

Stimulus	Response														
	/p/	/t/	/k/	/f/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/z/	/m/	/n/	/l/	/r/
/p/	100														
/t/		95.8						4.2							
/k/			100												
/f/				100											
/s/					100										
/ʃ/						100									
/b/				4.2			75.0			20.8					
/d/								100							
/g/									100						
/v/				4.2			4.2			91.7					
/z/					4.2						95.8				
/m/												100			
/n/													95.8		4.2
/l/														100	
/r/															100

¹The present study was a part of a larger project (2 h of data collection/child) in which the development of phonological processing skills in normal-hearing children was examined. No multiple sessions were planned to minimize attrition.

²These difference among the place cues may be partly due to the number of consonants in each place category.

- Bankson, N., and Bernthal, J. (1990). *Bankson-Bernthal Test of Phonology* (Special Press, San Antonio, TX).
- Bell, T. S., Dirks, D. D., and Kincaid, G. E. (1989). "Closed-set effects in consonant confusion patterns," *J. Speech Hear. Res.* **32**, 944–948.
- Bilger, R. B., and Wang, M. D. (1976). "Consonant confusion in patients with sensorineural hearing loss," *J. Speech Hear. Res.* **19**, 718–748.
- Bradley, J. S., and Sato, H. (2008). "The intelligibility of speech in elementary school classrooms," *J. Acoust. Soc. Am.* **123**, 2078–2086.
- Cutler, A., Garcia Lecumberri, M. L., and Cooke, M. (2008). "Consonant identification in noise by native and non-native listeners: Effects of local context," *J. Acoust. Soc. Am.* **124**, 1264–1268.
- Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). "Patterns of English phoneme confusions by native and non-native listeners," *J. Acoust. Soc. Am.* **116**, 3668–3678.
- Danhauer, J. L., Abdala, C., Johnson, C., and Asp, C. (1986). "Perceptual features from normal-hearing and hearing impaired children's errors on the NST," *Ear Hear.* **5**, 318–322.
- Donaldson, G. S., and Kreft, H. A. (2006). "Effects of vowel context on the recognition of initial and medial consonants by cochlear implant users," *Ear Hear.* **27**, 658–677.
- Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (1982). "Evaluation of hearing-impaired listeners using a nonsense-syllable test: II. Syllable recognition and consonant confusion patterns," *J. Speech Hear. Res.* **25**, 141–148.
- Dubno, J. R., and Levitt, H. (1981). "Predicting consonant confusions from acoustic analysis," *J. Acoust. Soc. Am.* **69**, 249–261.
- Dunn, L. M., and Dunn, L. M. (1997). *Peabody Picture Vocabulary Test*, 3rd ed. (American Guidance Service, Circle Pines, MN).
- Finitzo-Hieber, T., and Tillman, T. W. (1978). "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing impaired children," *J. Speech Hear. Res.* **21**, 440–458.
- Garcia Lecumberri, M. L., and Cooke, M. (2006). "Effect of masker type on native and non-native consonant perception in noise," *J. Acoust. Soc. Am.* **119**, 2445–2454.
- Gelfand, S. A., Piper, N., and Silman, S. (1985). "Consonant recognition in quiet as a function of aging among normal hearing subjects," *J. Acoust. Soc. Am.* **78**, 1198–1206.
- Gelfand, S. A., Piper, N., and Silman, S. (1986). "Consonant recognition in quiet and in noise with aging among normal hearing listeners," *J. Acoust. Soc. Am.* **80**, 1589–1598.
- Graham, L. W., and House, A. S. (1971). "Phonological oppositions in children: A perceptual study," *J. Acoust. Soc. Am.* **49**, 559–566.
- Hu, Y., and Loizou, P. C. (2007). "A comparative intelligibility study of single-microphone noise reduction algorithms," *J. Acoust. Soc. Am.* **122**, 1777–1786.
- Johnson, C. E. (2000). "Children's phoneme identification in reverberation and noise," *J. Speech Lang. Hear. Res.* **43**, 144–157.
- Mayo, C., and Turk, A. (2004). "Adult-child difference in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," *J. Acoust. Soc. Am.* **115**, 3184–3194.
- Mayo, C., and Turk, A. (2005). "The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception," *J. Acoust. Soc. Am.* **118**, 1730–1741.
- Miller, G. A., and Nicely, P. E. (1955). "An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.* **27**, 338–352.
- Moeller, M. P., Stelmachowicz, P. G., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., Wood, S., Lewis, D., and Pittman, A. (2007). "Vocalizations of infants with hearing loss compared to infants with normal hearing: Part I—Phonetic development," *Ear Hear.* **28**, 605–627.
- Nelson, P. B., and Soli, S. (2000). "Acoustical barriers to learning: Children at risk in every classroom," *Lang. Speech Hear. Serv. Sch.* **31**, 356–361.
- Neuman, A. C., and Hochberg, I. (1983). "Children's perception of speech in reverberation," *J. Acoust. Soc. Am.* **73**, 2145–2149.
- Ross, M., and Lerman, J. (1971). *Word Intelligibility by Picture Identification* (Stanwix House, Pittsburgh, PA).
- Sander, E. K. (1972). "When are speech sounds learned?," *J. Speech Hear. Disord.* **37**, 55–63.
- Singh, S., and Black, J. W. (1966). "Study of twenty-six intervocalic consonants as spoken and recognized by four language groups," *J. Acoust. Soc. Am.* **39**, 372–387.
- Stuart, A., Givens, G. D., Walker, L. J., and Elangovan, S. (2006). "Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise," *J. Acoust. Soc. Am.* **119**, 1946–1949.
- Studebaker, G. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Sussman, J. E. (2001). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states of transitions?," *J. Acoust. Soc. Am.* **109**, 1173–1180.
- Wang, M. D., and Bilger, R. C. (1973). "Consonant confusions in noise: A study of perceptual features," *J. Acoust. Soc. Am.* **54**, 1248–1266.
- Wang, M. D., Reed, C. M., and Bilger, R. C. (1978). "A comparison of the effects of filtering and sensorineural hearing loss on patterns of consonant confusions," *J. Speech Hear. Res.* **21**, 5–36.