

Leone, D.*, & Levy, E. S. (2015). Children's perception of conversational and clear American-English vowels in noise. *Journal of Speech, Language, and Hearing Research*, 58, 213-226. http://dx.doi.org/10.1044/2015_JSLHR-S-13-0285
Accepted for publication Dec. 22, 2014

Children's Perception of Conversational and Clear American-English Vowels in Noise
Dorothy Leone¹ & Erika S. Levy²

Iona College¹

Teachers College, Columbia University²

Abstract

Purpose: Much of a child's day is spent listening to speech in the presence of background noise. Although accurate vowel perception is important for listeners' accurate speech perception and comprehension, little is known about children's vowel perception in noise. "Clear speech" is a speech style frequently used by talkers in the presence of noise. This study investigated children's identification of vowels in nonsense words in noise and examined whether adults' use of clear speech would result in the children's more accurate vowel identification.

Method: Two female American-English (AE) speaking adults were recorded producing the nonsense word /gəbVpə/ with AE vowels /ɛ-æ-a-ʌ/ in phrases in conversational and clear speech. These utterances were presented to 15 AE-speaking children (ages 5.0-8.5) at a signal-to-noise ratio of -6 dB. The children repeated the utterances.

Results: Clear speech vowels were repeated significantly more accurately (87%) than conversational speech vowels (59%), suggesting that clear speech aids children's vowel identification. Children repeated one talker's vowels more accurately than the other's, and front vowels more accurately than central and back vowels.

Conclusions: The findings support the use of clear speech for enhancing adult-to-child communication in AE in noisy environments.

Children's Perception of Conversational and Clear American-English Vowels in Noise

Background noise exists in almost every listening environment (Helfer & Wilber, 1990) and can impact a listener's ability to perceive the speech signal (Crandell & Smaldino, 2000; Helfer & Wilber, 1990). As a signal-to-noise ratio (SNR) becomes less favorable, a listener's accuracy in perceiving the signal decreases (Stelmachowicz, Hoover, Lewis, Kortekaas, & Pittman, 2000). The ability to perceive a speech signal in adverse listening conditions, such as in the presence of noise or reverberation, increases with age until early adulthood (Crandell & Smaldino, 2000; Neuman & Hochberg, 1983; Nishi, Lewis, Hoover, Choi, & Stelmachowicz, 2010; Soli & Sullivan, 1997). Nelson and Soli (2000) suggest that acoustic environments utilizing a +15 dB SNR allow children to perceive a signal fully. In United States schools, SNRs have been reported between +3.0 dB and -17.6 dB while class is in session (Larsen & Blair, 2008), suggesting that many children spend a considerable part of their day listening to speech in the presence of background noise that might obscure the speech signal.

A fundamental problem addressed in studies of spoken-word recognition is how stimulus information in an acoustic signal is mapped onto items in the mental lexicon, permitting humans to effortlessly isolate the speech pattern of a particular word from among the many possible alternatives (Luce, Pisoni, & Goldinger, 1990). A predominant model of spoken-word identification, the Neighborhood Activation Model (NAM) (Luce & Pisoni, 1998) posits that acoustic-phonetic patterns activate word-decision units. A "similarity neighborhood" of phonetically similar words is activated and words compete for recognition. The greater the similarity neighborhood and the lower the word frequency, the slower and less accurate word recognition will be. Similarity neighborhoods in younger children are thought to be sparser than those of older children and adults (e.g., Charles-Luce & Luce, 1990; 1995), but questions remain regarding the relationship between the density of similarity neighborhoods and children's developing lexicons (see Coady & Aslin, 2003; Dollaghan, 1994). According to the NAM (Luce & Pisoni, 1998), stimulus degradation impedes the complete processing of stimuli and thus the word-recognition system. As less information is obtainable when it is masked by noise, a decision would necessarily rely on any information available.

Accurate perception of words relies to a large extent on accurate vowel perception. Kewley-Port, Burkle, and Lee (2007) found that adults' accurate vowel identification has a greater impact on the intelligibility of words in sentences in American-English (AE) than does adults' accurate consonant identification. Thus, examining the effects of noise on children's vowel identification will likely illuminate challenges to speech perception faced by children daily as they attempt to map a degraded speech signal to competing words in their similarity neighborhoods.

Talkers vary their articulatory clarity according to listener need, hyperarticulating when the listeners require the extra acoustic properties yielded by such speech and otherwise reducing articulatory effort (Lindblom, 1990). Research has found that talkers modify their speech style in noisy environments in ways that are beneficial to the listener. In noisy environments, talkers respond with a Lombard effect, for example, by increasing their vocal intensity (Garnier, Henrich, & Dubois, 2010).

Clear speech

Clear speech, the focus of this study, is another intelligibility-enhancing style that talkers may utilize in background noise or with listeners with hearing loss (Picheny, Durlach, & Braida, 1986; Smiljanic & Bradlow, 2009; Uchanski, 2005). The articulatory care taken by the talker in clear speech can be thought of as enhancing the information made available to help the listener

make a lexical match in a similarity neighborhood when the speech signal is perturbed. Clear speech is contrasted with “conversational speech,” the speech style used with highly familiar interlocutors, such as a friends or family members (Bradlow, Kraus, & Hayes, 2003). While descriptions of the acoustic and phonetic characteristics of clear speech vary across studies, clear speech typically is characterized as having higher intensity (Bradlow et al., 2003; Picheny et al., 1986; Uchanski, 2005), slower speaking rate, more pauses, and a higher fundamental frequency (Bradlow et al., 2003; Krause & Braida, 2002; Picheny et al., 1986; Uchanski, 2005) than conversational speech.

Clear speech vowels in AE are characterized by longer duration (Ferguson & Kewley-Port, 2002, 2007; Picheny et al., 1986) and less vowel reduction (Picheny et al., 1986; Uchanski, 2005) than AE conversational speech vowels. Expanded vowel space has been found in clear speech when compared to conversational speech (Bradlow et al., 2003). Specifically, for mid and low vowels, the first formant (F1), which corresponds to tongue height, is higher when the vowels are uttered in clear speech than in conversational speech, suggesting a lower tongue position for these vowels in clear speech. In contrast, for high vowels, the F1 is lower when the vowels are uttered in clear speech than in conversational speech, suggesting a higher tongue position for these vowels in clear speech. For front vowels, moreover, the second formant (F2), which corresponds to anterior-posterior tongue movement, is higher when these vowels are uttered in clear speech, suggesting a more forward tongue position for front vowels in clear speech than in conversational speech (Ferguson & Quené, 2014). On the other hand, for back vowels, the F2 is lower for clear speech than for conversational speech, suggesting a more retracted tongue position (Ferguson & Kewley-Port, 2007). In addition, clear speech vowels are more dynamic than their conversational speech counterparts, as indicated by greater spectral change (Ferguson & Kewley-Port, 2002).

Gender differences have also been found, with adults more accurately identifying AE vowels produced by females than those produced by males (Ferguson, 2004). Similarly, children benefit more from AE clear speech produced by a female talker than by a male talker (Bradlow et al., 2003). Several studies have documented a clear speech intelligibility benefit for a variety of talkers, utterances, and listener groups. Generally, as the listening environment becomes more degraded, a clear-speech benefit becomes greater (Bradlow & Bent, 2002; Payton, Uchanski, & Braida, 1994).

One of the first reports of a clear-speech benefit involved adult listeners with hearing loss (Picheny, Durlach, & Braida, 1985). Subsequently, researchers have likewise demonstrated a clear-speech benefit for listening populations with normal hearing. For example, a handful of studies have documented adults’ greater identification accuracy of AE clear speech sentences than of AE conversational speech sentences (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Krause & Braida, 2002, 2009; Payton et al., 1994). For AE clear speech vowels, a few studies have examined adults’ perception (Ferguson, 2004; Ferguson & Kewley-Port, 2002; Rogers, DeMasi, & Krause, 2010), and a preliminary study was conducted examining children’s perception (Leone et al., 2012). All studies documented a clear-speech benefit for identification of some vowels. Specifically, Ferguson and Kewley-Port (2002) reported that adults identified /e, ε, æ, ʌ, o, ʊ, u/, but not /i, I, a/, more accurately when the vowels were produced in clear speech than in conversational speech. The authors suggest that /i/ and /a/ are robust vowels, less in need of clarification. However, Rogers et al. (2010) reported that clear speech /ε, a/, but not /i, I, e, æ/ were identified by adults significantly more accurately than their conversational speech

counterparts. (For a summary of findings from clear speech studies through 2004, see Uchanski, 2005.)

Although clear speech vowel studies (e.g., Ferguson, 2004; Ferguson & Kewley-Port, 2002; Rogers et al., 2010) have focused on overall vowel perception rather than on which vowels were confused with each other when misidentifications occurred, findings regarding adults' conversational vowel confusions may also be applicable to clear speech studies. Conversational vowel confusions tend to involve vowels proximal in acoustic-phonetic vowel space. For example, Neel (2008) reported that AE-speaking adults most often confused neighboring vowels when identifying conversational speech vowels in /hVd/ context in quiet. Similarly, adults confused vowels close in vowel space when identifying syllables presented in multitalker babble at 0 dB SNR (Cutler, Weber, Smits, & Cooper, 2004).

Children's perception of clear speech

The two studies that have examined *children's* perception of clear speech (Bradlow et al., 2003; Riley & McGregor, 2012) have also found intelligibility benefits of this speech style. Bradlow et al. (2003) examined real-word repetition accuracy of school-age children listening to sentences in noise, at -4 dB and -8 dB SNRs in both conversational and clear speech. Broadband white noise was used to mask the signal at all frequencies. The investigators presented adults' sentences containing three to four keywords over loudspeakers to children with and without learning disabilities. Children were instructed to repeat each sentence while experimenters who were seated next to the children noted keywords that were repeated accurately. The children repeated the sentences with approximately 9% greater accuracy when the sentences were produced in clear speech than when they were produced in conversational speech (Bradlow et al., 2003). Both groups benefited more from clear speech in the -8 dB SNR condition than in the -4 dB SNR condition, suggesting that as the listening condition becomes more adverse, the clear-speech benefit increases. Similarly, Riley and McGregor (2012) showed a clear-speech benefit for school-age children listening over loudspeakers to conversational and clear speech narratives that contained target words embedded in white noise in a +8 dB SNR produced by one female talker. The children selected the picture (from a field of four on a computer screen) that represented the target word they heard. They identified the pictures more accurately when the words were produced in clear speech than in conversational speech, indicating beneficial effects of clear speech for children's real word comprehension.

Young children's clear-speech production strategies in adverse listening conditions differ from those used by older children and by adults (Pettinato & Hazan, 2013). Thus, it is possible perceptual strategies in resolving clear speech in noise also continue to develop throughout childhood. A handful of studies have discussed children's vowel perception, with limited information provided about children's vowel perception in noise or in clear speech. There is some evidence that children perceive speech in noisy environments less accurately than do adults, and that the younger the child and the more challenging the SNR, the poorer the child's perceptual accuracy (Bradley & Sato, 2008; Stuart, Givens, Walker, & Elangovan, 2006). Nishi et al. (2010) documented similar results for children's AE consonant perception. Additionally, Johnson (2000) examined children's and young adults' AE consonant and vowel identification by presenting CVCV sequences in multitalker babble noise at +13 dB SNR. The children identified vowels in noise with adult-like accuracy by age 10 and consonants in noise with adult-like accuracy by age 14. Thus, children's ability to perceive vowels and consonants in noise appears to increase as they age. If adults' clear speech vowels increase intelligibility for child

listeners, use of clear speech may be supported as a strategy to enhance communication with children, who appear to produce and perceive speech in noise differently from adults.

The current study

The current study examined the accuracy with which monolingual AE-speaking school-age children perceive adults' AE vowels in conversational and clear speech nonsense words in noise. Specifically, it investigated whether typically-developing school-age children's repetition accuracy of vowels in noise varied as a function of the talkers' speaking style (conversational vs. clear), the particular AE vowel uttered ($/\varepsilon, \text{æ}, \text{ɑ}, \text{ʌ}/$), and the talker. Acoustic differences between conversational and clear vowels were also examined. It was predicted that the children's repetition accuracy would be higher in clear speech than in conversational speech (Bradlow et al., 2003; Ferguson & Kewley-Port, 2002). It was further hypothesized that $/\varepsilon/$ would be identified with the least accuracy (Ferguson & Kewley-Port, 2002; Leone et al., 2012), with confusion occurring among vowels proximal in acoustic vowel space (Neel, 2008; Cutler et al., 2004). Certain vowels were expected to be aided more by clear speech than others (Ferguson & Kewley-Port, 2002; Rogers et al., 2010). Talker differences were anticipated, with some talkers providing a larger intelligibility benefit for child listeners than others (Bradlow et al., 2003; Ferguson, 2004; Uchanski, 2005). Lastly, acoustic analysis was predicted to reveal that vowels in clear speech would be characterized by longer duration, larger F1/F2 vowel space, and greater spectral variation (Ferguson & Kewley-Port, 2002) than vowels in conversational speech.

The stimuli in the current study were nonsense words. According to the NAM (Luce & Pisoni, 1998), acoustic-phonetic information drives the system of word decisions by providing bottom-up information. Higher-level lexical information such as word frequency biases word decision units, providing contextual information and a priori probabilities that influence the decision units. Our two main comparison studies (i.e., Bradlow et al., 2003; Riley & McGregor, 2012), which investigated children's perception of clear speech, examined the children's recognition of real words, for which higher-level lexical information is available and influences word recognition. To examine intelligibility of speech in noise at the level of "acoustic-phonetic pattern activation," nonsense words can be used, reducing higher-level contextual and word-frequency effects (Dollaghan & Campbell, 1998; Luce & Pisoni, 1998; Neuman & Hochberg, 1983; cf Vitevitch, Luce, Pisoni, & Auer, 1999). Thus, the current study investigated children's perception of nonsense words in a degraded speech signal, and the effects of enhancing the speech signal through clear speech. Results at this "bottom up" level of processing complement findings yielded by studies of children's perception of real words influenced by higher-level lexical information.

Method

Participants

A total of 15 monolingual native New York English speaking children, ages 5.0-8.5, with normal hearing, were recruited from the New York City metropolitan area to serve as participants. This age range was selected to tap into young children's speech perception skills before the children reach adult-like accuracy in their vowel identification, which is thought to occur at approximately age 10 (Johnson, 2000). The native and only language spoken in the homes of the child listeners (henceforth "children") was AE, specifically, the New York variety. The children's legal guardians reviewed and signed Institutional Review Board (IRB) consent forms. All children passed a bilateral hearing screening at 20 dB HL (re: ANSI, 2004) at 500, 1000, 2000, and 4000 Hz and had no history of speech or language disorders per guardian

reports. All presented with typical articulation, as evidenced by a score no more than one standard deviation below the mean on the *Arizona Articulation Proficiency Scale, 3rd edition* (Fudala, 2001).

Stimulus materials and preliminary studies

Four native monolingual New York English speaking female adult talkers from the New York metropolitan area produced AE vowels / ε , æ , ɑ , ʌ / in the trisyllable nonsense word / $\text{g}\text{ə}\text{b}\text{V}\text{p}\text{ə}$ / embedded in the carrier phrase “Five _____ this time” (see Strange et al., 2007). Carrier phrases were utilized to be more representative of everyday speech than would vowels or words in isolation.

The talkers were monolingual speakers of AE and had minimal exposure to speaking and listening to other languages, had no history of speech and language disorders, and passed a bilateral hearing screening at 20 dB HL (re: ANSI, 2004) at 500, 1000, 2000, and 4000 Hz. A talker with a similar language profile to those in the experimental condition was recorded for the familiarization task. Female talkers were selected, as females have been found to produce more intelligible clear speech (Bradlow, Torretta, & Pisoni, 1996) and a larger clear-speech benefit (Bradlow et al., 2003; Uchanski, 2005) than males. In addition, because approximately 76% of teachers in the United States are female (National Center for Education Statistics, 2009), school-age children likely spend much of their day listening to female voices.

The talkers were recorded in a sound-treated booth with the experimenter in an adjoining room in visible contact. The experimenter provided the talker with directions using an intercom and listened to the recording over Sennheiser HD 280 pro headphones. Talkers were instructed to read four lists of utterances (Five / $\text{g}\text{ə}\text{b}\text{V}\text{p}\text{ə}$ / this time) in conversational and in clear speech. Instructions for producing conversational phrases were: “Speak at a normal rate, as if speaking with someone who is very familiar with your voice.” For clear speech, talkers were instructed: “Speak as if talking with someone with a hearing loss.” Protocols consisted of randomized lists of 12 utterances. The first utterance and the last utterance contained the same target vowel and the final utterance was discarded to control for list-final intonation effects. It should be noted that the target / bV / was the stressed syllable of the utterance because this was the syllable that changed from utterance to utterance as the talkers read aloud. (For a description of the temporal and acoustic properties of these vowels in AE adults’ utterances, see Strange et al., 2007.)

All conversational stimuli were recorded prior to clear speech stimuli. If an utterance contained irregular pronunciation, rate, prosody, vocal quality, or noise, the talker was asked to repeat the stimulus. Output was recorded through a Shure (SM58) microphone placed 15 cm from the talker’s mouth and passed through a Shure (Prologue 200M) mixer to a Turtle Beach Riviera sound card of a Dell Pentium 4 desktop computer using SoundForge8.0 software (Sony, 2009), with a sample rate of 22,050 Hz, 16-bit resolution, on a mono channel.

For each vowel, the second and third tokens recorded were utilized for the stimuli unless they were characterized by background noise or disfluent speech. Two tokens of each utterance were used to tap into categorial perception (Gottfried, 1984; Levy & Strange, 2008) rather than simple physical discrimination. To minimize amplitude differences between talkers and speech styles, the mean root-mean-square (RMS) ($\bar{x} = -20.3$ dB) value was calculated across all stimuli. All stimuli were subsequently scaled to this amplitude using SoundForge8.0 software. Speech-shaped noise was created using Akustyk v. 1.8 (Plichta, 2004), an add-on program for Praat (Boersma & Weenink, 2006). The program created a noise file that had long-term spectral characteristics shaped to closely resemble the signal. The stimuli were then mixed with speech-shaped noise at -6 dB SNR by means of the Praat v. 5.2.22 program (Boersma & Weenink,

2006).

The SNR of -6 dB was selected based, in part, on findings from a preliminary child study and previous literature, with the goal of being able to measure potential differences between identification of clear vs. conversational speech in a realistic listening environment. In the preliminary child study, data were collected from two 6-year-old females who listened to target stimuli in 0 dB and +2 dB SNRs in a sound-treated booth and were asked to repeat what they heard. An adult experimenter noted their response choices on a computer using Paradigm software. Results revealed a ceiling effect: both children repeated 100% of the conversational and clear speech vowels accurately (Leone et al., 2012); thus, it was concluded that the noise level would need to be higher to permit detection of any clear speech effects. Additional preliminary data from a 7-year-old boy tested as described above revealed difficulty identifying the target vowels in -8 dB SNR and greater accuracy in the -4 dB SNR condition (Leone et al., 2012).

The SNR for testing was further narrowed down by previous literature consulted, including Bradlow et al.'s (2003) examination of conversational and clear keywords at -4 dB and -8 dB SNRs for children of similar ages (8.1-12.5 years) to the current study's listeners (5.0-8.5 years), Stuart et al.'s (2006) analysis of monosyllabic words at -10 dB SNR for children (4.0-5.0 years) younger than the current study's listeners, and Rogers et al.'s (2010) review of conversational and clear /bVd/ stimuli at -8 dB SNR for adults. However, it should be noted that Riley and McGregor (2012), who included slightly older (9.0-10.11 years) participants, utilized a higher SNR (+8 dB). Lastly, a realistic listening condition for children was sought; Larsen and Blair (2008) reported SNRs between +3.0 dB and -17.6 dB while class is in session. A -6 dB SNR was thus selected. A 0 dB SNR condition served as a control measure to ensure that the children were attending to the task.

Selection of talkers was based on the results of a preliminary adult listener study revealing two talkers with the greatest clear vs. conversational speech differences. The reduction to two talkers was aimed to render the task more feasible, given children's relatively short attention spans (McKay, Halperin, Schwartz, & Sharma, 1994). In this preliminary study, eleven monolingual AE adult listeners from the New York City metropolitan area were presented with recordings of conversational and clear speech from four talkers. Stimuli were entered into the Paradigm software program v.1.0.2 (Tagliaferri, 2011) and presented at a -10 dB SNR over Sennheiser HD 280 pro headphones at a comfortable listening level. In an 11-alternative closed-set response paradigm, the following response options representing all AE vowels were displayed on the computer monitor: "gabeepra (/gəbɪpɹə/), gabuppa (/gəbʌpɹə/), gabeppa (/gəbɛpɹə/), gabappa (/gəbæpɹə/), gabippa (/gəbɪpɹə/), gabaypa (/gəbɛpɹə/), gaboopa (/gəbʊpɹə/), gabUpa (/gəbʊpɹə/), gaboapa (/gəbɒpɹə/), gabawpa (/gəbɔpɹə/), gaboppa (/gəbɒpɹə/)" (see Strange et al., 2007). The adult listeners clicked on the response option indicating the word they perceived.

Descriptive results indicate that clear speech vowels were identified more accurately than conversational speech vowels (78% and 45% for clear and conversational speech, respectively). Furthermore, McNemar's test for paired proportions indicated that clear speech was identified significantly more accurately than conversational speech for all talkers (Talker 1: $\chi^2= 105.35(1)$, $p < .001$, $OR= 8.11$; Talker 2: $\chi^2= 41.86(1)$, $p < .001$, $OR= 3.45$; Talker 3: $\chi^2= 97.59(1)$, $p < .001$, $OR= 6.36$; Talker 4: $\chi^2= 75.63(1)$, $p < .001$, $OR= 4.66$). Talkers 1 and 2 demonstrated a larger clear speech effect (clear minus conversational difference = 38% and 38%, respectively) than Talkers 3 and 4 (clear minus conversational difference = 33% and 22%, respectively) and were thus chosen as talkers for the study. In addition, when identification results from conversational

and clear stimuli were combined (Talkers 1 and 2 only), adults identified / ε / and / Λ / more accurately (64% each) than / æ / (61%) and / α / (43%). However, when identification results from only clear speech vowels produced by Talkers 1 and 2 were examined, adults identified / Λ / most accurately (95%), followed by / ε / (81%), / æ / (81%), and / α / (52%). In the Discussion section, the children's responses are compared with these results.

Acoustic Analysis of the Stimuli

Acoustic analysis was performed using Wavesurfer v. 1.8.5 (Sjölander & Beskow, 2006) and Praat v. 5.2.22 (Boersma & Weenink, 2006). For measuring utterance duration, the beginning of each utterance “Five g**ə**bV**p**ə this time” was defined as the first mark of frication energy of the waveform for /f/ and the end of each utterance was defined as the end of the voicing bar on the spectrogram for /m/. The onset and offset of the syllable containing the target vowel were determined manually on the basis of the following definitions: Syllable onset was defined as the release burst of the /b/, which was visually correlated with the spike of acoustic energy on the spectrogram, and syllable offset was defined as the beginning of closure of /p/, evidenced by a decrease in periodic energy in the higher formants (F2, F3) on the spectrogram. Thus, vowel duration included the entire gesture from release of the preceding consonant to the beginning of full closure of the following consonant. To obtain a measure of fundamental frequency, F_0 was measured at the vowel's temporal midpoint. Values for the first three formants for a 25 ms window were calculated using formant tracking in Wavesurfer with linear predictive coding (LPC) analysis at the temporal midpoint (50% point) between onset and offset of the syllable. Comparisons were made to normative data (Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952; Strange et al., 2007).

Greater utterance and vowel duration and F_0 and a larger vowel space were found for clear speech than for conversational speech for both talkers. Table 1 lists 6 acoustic-phonetic measurements for both talkers. The mean utterance and vowel durations in conversational and clear speech and the clear-minus-conversational difference for each talker are shown. Consistent with other clear speech studies (Bradlow et al., 2003; Ferguson & Kewley-Port, 2007), both talkers showed a large increase in utterance and vowel duration for clear speech relative to conversational speech. For productions by Talker 2, utterance and vowel length for clear speech were greater than for productions by Talker 1. A higher F_0 at the vowel's temporal midpoint was found for clear speech relative to conversational speech. In conversational speech, the mean F_0 at midpoint was higher for the vowels produced by Talker 2 than for those produced by Talker 1. For clear speech, both talkers' mean F_0 at the vowel midpoint were comparable. The F1 was higher in clear speech than conversational speech for all stimuli. This is partially consistent with clear speech vowel studies (e.g., Ferguson & Quené, 2014) that posit that the tongue lowers during clear speech mid and low vowel production and rises for clear speech high vowel production. The F2 was higher for clear speech front vowels (/ æ , ε /) than conversational speech front vowels and lower for clear speech back vowels (/ α /) than conversational back vowels, indicating a larger front-back dimension of vowel space. These findings are in agreement with clear speech vowel studies (e.g., Ferguson & Kewley-Port, 2007) suggesting that the tongue moves farther forward during clear speech front vowel production and farther back for clear speech back vowel production than for conversational vowel production. The low-mid central unrounded vowel / Λ / showed varying F2 values. The F2 for / Λ / was lower for clear than for conversational speech when produced by Talker 1, whereas the F2 was higher for clear than for conversational speech when produced by Talker 2.

Table 1

Acoustic Analysis of the Conversational and Clear Speech Utterances as Produced by the Two Talkers (Talker 1 and Talker 2) Included in the Children's Repetition Task for Each Target Vowel

| Acoustic measurement | Talker 1 | | | Talker 2 | | |
|--------------------------------------|----------|-------|-------|----------|-------|-------|
| | Conv. | Clear | Diff. | Conv. | Clear | Diff. |
| <i>/ɛ/</i> | | | | | | |
| Utterance duration (s) | 1.270 | 2.765 | 1.495 | 1.174 | 3.684 | 2.510 |
| Vowel duration (s) | 0.089 | 0.133 | 0.044 | 0.096 | 0.175 | 0.080 |
| Duration ratio (vowel/utterance) (s) | 0.070 | 0.048 | | 0.081 | 0.048 | |
| F ₀ at 50% point (Hz) | 184 | 276 | 92 | 216 | 286 | 70 |
| F1 value at 50% point (Hz) | 737 | 829 | 93 | 774 | 883 | 109 |
| F2 value at 50% point (Hz) | 2117 | 2254 | 137 | 2065 | 2243 | 178 |
| <i>/æ/</i> | | | | | | |
| Utterance duration (s) | 1.277 | 2.685 | 1.408 | 1.214 | 3.335 | 2.122 |
| Vowel duration (s) | 0.135 | 0.210 | 0.075 | 0.142 | 0.235 | 0.094 |
| Duration ratio (vowel/utterance) (s) | 0.105 | 0.078 | | 0.117 | 0.071 | |
| F ₀ at 50% point (Hz) | 183 | 260 | 77 | 198 | 262 | 64 |
| F1 value at 50% point (Hz) | 1008 | 1092 | 85 | 965 | 1067 | 102 |
| F2 value at 50% point (Hz) | 1783 | 2219 | 436 | 1882 | 2192 | 311 |
| <i>/ɑ/</i> | | | | | | |
| Utterance duration (s) | 1.210 | 1.705 | 0.495 | 1.260 | 3.750 | 2.490 |
| Vowel duration (s) | 0.127 | 0.177 | 0.051 | 0.142 | 0.216 | 0.074 |
| Duration ratio (vowel/utterance) (s) | 0.105 | 0.104 | | 0.113 | 0.058 | |
| F ₀ at 50% point (Hz) | 172 | 251 | 79 | 203 | 219 | 16 |
| F1 value at 50% point (Hz) | 895 | 947 | 52 | 954 | 1093 | 139 |
| F2 value at 50% point (Hz) | 1366 | 1228 | -138 | 1515 | 1486 | -29 |
| <i>/ʌ/</i> | | | | | | |
| Utterance duration (s) | 1.220 | 1.665 | 0.445 | 1.255 | 3.500 | 2.245 |
| Vowel duration (s) | 0.092 | 0.129 | 0.037 | 0.084 | 0.147 | 0.063 |
| Duration ratio (vowel/utterance) (s) | 0.075 | 0.077 | | 0.067 | 0.042 | |
| F ₀ at 50% point (Hz) | 194 | 210 | 16 | 256 | 284 | 28 |
| F1 value at 50% point (Hz) | 754 | 814 | 60 | 820 | 856 | 36 |
| F2 value at 50% point (Hz) | 1472 | 1401 | -72 | 1540 | 1621 | 81 |

Note: all values are averages of token 1 and token 2

Procedure

Testing took place in a sound-treated booth in the Speech Production and Perception Lab at Teachers College, Columbia University. Each child was tested separately and was assisted by

three experimenters. During familiarization and experimental tasks, stimuli were presented over two loudspeakers (Altec Lansing BXR 1320) placed approximately 2 feet away from the listener at a mean output level of 65 dB SPL as verified by a Galaxy CM-140 Sound Pressure Level meter placed 30 centimeters from the Shure (SM58) microphone before and after testing sessions. Output was recorded through a Shure (SM58) microphone placed approximately 15 cm from the child's mouth and passed through a Shure (Prologue 200M) mixer to a Turtle Beach Riviera sound card of a Dell Pentium 4 desktop computer using SoundForge8.0 software, with a sample rate of 22,050 Hz, 16-bit resolution, on a mono channel. Experimenter 1 was seated in an adjoining room and monitored the child's repetitions over Sennheiser HD 280 pro headphones. The child was seated across from Experimenter 2, while Experimenter 3 was seated at a computer behind the child.

Stimulus presentation was controlled by Paradigm v.1.0.2 software (Tagliaferri, 2011). The children were given the following directions: "We're going to listen to some sentences with silly-sounding words. I want you to listen and then say exactly what you heard." Experimenter 3 played each adult stimulus using Paradigm software. The child repeated the stimulus. Experimenter 3 clicked on the response indicating the word she perceived, using the same key word options as in the adult identification task. For reliability an additional adult listener identified the children's vowels in a sound treated booth over Sennheiser HD 280 pro headphones at a later time, using the same Paradigm program as Experimenter 3. Reliability between the two adult listeners (the experimenter who was present with the child and the experimenter who later listened to the child recordings and the target stimuli) was 100%.

To address the possible confound of the adult listener's responses being influenced by the adult stimuli preceding the child responses, at a later time, two additional adults were presented with the children's utterances only (and not the adult stimuli) and identified the children's vowels. Reliability between the experimenter who was present with the child and the two experimenters who later listened to only the child recordings was 99%. If a discrepancy occurred, an additional listener was asked to respond and the most frequent response was selected.

A repetition task, which has been used in other studies with participants of similar ages (e.g., Bradlow et al., 2003; Neuman & Hochberg, 1983; Nishi et al., 2010), was used in the present study in order to tap into the children's identification skills, as young children's reading, language, and cognitive skills may not allow them to complete identification tasks reliably, especially those involving nonsense words. In addition, Neuman and Hochberg (1983) suggest the use of a repetition task to help reduce memory and cognitive demands. Children were provided with as much time as needed to respond. If repetition of a stimulus was required because the child was not attending or because external noise was present, Experimenter 3 replayed the stimulus. Between blocks, Experimenter 2 provided encouragement both verbally and through games in the sound-treated booth.

Prior to the experimental condition, task and stimulus familiarization procedures trained the children to become familiar with the stimuli and repeat what they heard. Children completed a 24-trial task familiarization block, which included conversational and clear vowels /t, o, u/ in trisyllables /gəbVpə/ in quiet produced by a different talker from the talkers who produced the experimental stimuli. They were required to achieve 90% accuracy during the 24-trial task familiarization block. All met this requirement.

Each child then completed a 36-trial stimulus familiarization block consisting of one presentation of each stimulus. Data from this block were discarded. After the stimulus

familiarization task, children were presented 4, 36-trial blocks (4 vowels X 2 talkers X 1 experimental SNR X 2 speaking styles X 2 tokens X 4 repetitions + 1 vowel X 2 talkers X 1 control SNR X 2 speaking styles X 2 tokens X 2 repetitions), yielding a total of 144 responses. Thus, the child completed 16 repetitions of each vowel, 8 repetitions for each talker. All stimuli were randomized within the blocks. Between blocks, children were given a break during which they played a game with the experimenter for approximately 5 minutes. Total testing time was approximately 2 hours.

Results

Data analysis

A total of 2,160 responses were collected from the 15 children. All responses were totaled and a percent correct score was computed (i.e., number of accurate responses/total responses). Control data were then tallied separately (16 trials per listener). A repetition correct score using all 1,920 experimental trials (2,160 trials – 240 control trials) was obtained. A repetition correct score was also calculated for each subcategory of the two speech styles (i.e., conversational and clear). Data were further divided by talker as well as by each vowel. Each repetition correct score was converted to a percent correct score. Descriptive and nonparametric statistics were performed on all data. Statistical analyses were performed with R version 2.15.1 (R Core Team, 2012) using the *glmer* function from the *lme4* package (version 0.999999-0).

Children's responses were analyzed using mixed-effects logistic regression analysis with a logit link and binomial family. Mixed effects models have been reported to provide more reliable results for categorical outcome variables (e.g., the forced-choice variables used in this study) than analysis of variance methods on transformed data (e.g., arcsine-square-root transformation) (Jaeger, 2008). In addition, the use of mixed effects modeling has been suggested to demonstrate higher statistical power and robustness than repeated-measures analysis of variance techniques in speech perception studies (Ferguson, 2012). The model used in the present study included speaking style, talker, and vowel as fixed effects, and listener and trial as random effects. Identifying listeners as a random effect allowed the model to consider the sample of children used in this study as a random sample from a larger population. Identifying trials as a random effect allowed it to consider the adult talkers' vowel recordings as a sample of a larger population of the talkers' possible vowel productions (Baayen, Davidson, & Bates, 2008). The effects of the independent variables (both fixed effects and random slopes for speaking style, talker, and vowel) were tested individually and retained only if they improved the model fit significantly. A theory-guided model comparison approach permitted us to contrast alternative models that were progressively more complex. The likelihood-ratio test and the Akaike Information Criterion (AIC) were used to compare the fit of competing models. Only the final model that fit the data best is presented and discussed.

Speaking style effects on vowel repetition

Returning to the first research question, regarding speaking style, children repeated clear speech vowels more accurately than conversational speech vowels. Clear speech vowels were repeated with 87% accuracy ($SD = 12.0$) and conversational speech vowels with 59% accuracy ($SD = 14.5$), revealing a mean difference between speaking styles of 27%. Mixed effects logistic regression confirmed that this effect of speaking style (clear vs. conversational speech) on vowel repetition was statistically significant ($z = 6.34, p < .001$).

In a further investigation of the effects of speaking style on vowel repetition by younger vs. older children, a descriptive analysis separated children into two groups according to age (younger group = ages 5.0 – 6.7; older group = 6.8 – 8.5). The younger group, which included 5

children, repeated clear speech vowels with 83% accuracy and conversational speech vowels with 54% accuracy, yielding a mean difference between speaking styles of 29%. The older group, which consisted of 10 children, repeated clear speech vowels with 89% accuracy and conversational speech vowels with 65% accuracy, revealing a mean difference of 24%. This finding suggests that clear speech benefited both younger and older children similarly.

The analysis also separated the children into groups according to gender. The 8 boys repeated clear speech vowels with 85% accuracy and conversational speech vowels with 58% accuracy, yielding a mean difference between speaking style of 23%. The 7 girls repeated clear speech vowels with 89% accuracy and conversational speech vowels with 61% accuracy, resulting in a mean difference between speaking styles of 28%. Thus, clear speech benefited both boys and girls similarly.

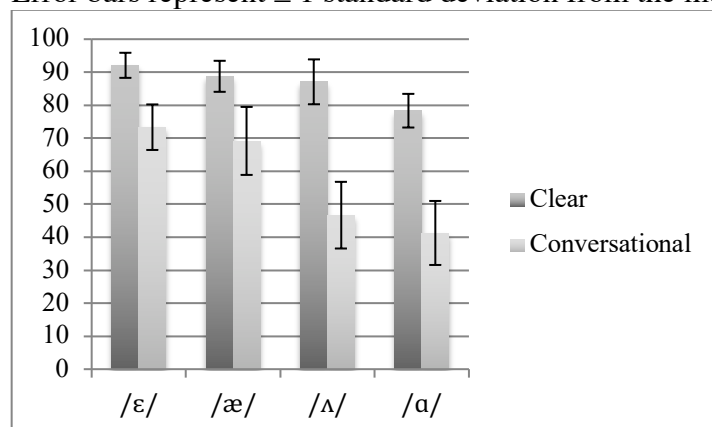
Particular vowel effects on vowel repetition

In response to the second research question, regarding particular vowel effects on repetition accuracy, listeners repeated / ϵ / with the highest accuracy (83%, $SD = 11.1$), followed by / \ae / (79%, $SD = 9.0$) then / Λ / (67%, $SD = 16.4$), and / a / with the least accuracy (63%, $SD = 12.2$). The control vowel / i / was repeated with 100% accuracy for all listeners, indicating that all children were on task. Pairwise comparisons revealed statistically significant differences in children's repetition accuracy for most of the target vowels. The vowel / ϵ / was repeated significantly more accurately than / Λ / ($z = 2.02, p < .044$) and / a / ($z = 3.04, p < .001$), but not than / \ae / ($z = 0.47, p = .641$). Furthermore, / \ae / was repeated significantly more accurately than / a / ($z = 2.13, p < .010$) and / Λ / ($z = 2.30, p < .021$). The difference between / Λ / and / a / was not statistically significant ($z = 0.47, p = .637$).

Particular vowel effects for each speaking style

Overall vowel repetition accuracy varied as a function of speaking style. Figure 1 depicts children's percent correct vowel repetition scores for conversational and clear speech, with target vowels along the x -axis and percent correct scores on the y -axis. Listeners repeated / ϵ / with the most accuracy and / a / with the least accuracy in both speaking styles (92%, 78% for clear speech vowels and 73%, 48% for conversational speech vowels, respectively). Repetition accuracy increased the most for / Λ / (clear speech vowel – conversational speech vowel difference = 40%) and thus showed the largest clear-speech benefit. There was no significant interaction between speaking style and individual vowel ($\chi^2(3) = 2.471, p = .481$). Thus, the clear speech benefit was not vowel-specific.

Figure 1. Children's percent correct vowel repetition scores for conversational and clear speech. Error bars represent ± 1 standard deviation from the mean.



The confusion matrix in Table 2 represents the children's identification (i.e., repetition) responses (options are in the top row) to the vowels intended (i.e., stimuli produced) by the adult talkers (listed in the left column). Responses are listed as percentages of the total repetitions for each target vowel for conversational vowels, clear vowels, and conversational and clear vowels combined. Consistent with preliminary studies (Leone et al., 2012) and conversational speech vowel confusions (Neel, 2008), vowels close in acoustic space were most often confused. When conversational and clear speech data were combined, children most frequently produced the front vowels /ɛ/ as /æ/ (8%) and /æ/ as /ɛ/ (12%). When conversational and clear speech vowel repetitions were separated (see Table 2 for confusion matrix), confusions between /ɛ/ and /æ/ remained, especially in conversational speech. Listeners repeated /ɛ/ as /æ/ (12% for conversational vowels; 3% for clear vowels) and /æ/ as /ɛ/ for each speaking style (17% for conversational vowels; 6% for clear vowels).

Table 2
Vowel Identification Rates and Confusion Matrix for Children Responding to Conversational and Clear Speech

| | <u>Vowel identified (repeated by child)</u> | | | | | | | | |
|---|---|------------|------------|-------------|-------------|-------------|-------------|------------|------------|
| | <u>/i/</u> | <u>/ɪ/</u> | <u>/e/</u> | <u>/ɛ/</u> | <u>/æ/</u> | <u>/ɑ/</u> | <u>/ʌ/</u> | <u>/ɔ/</u> | <u>/o/</u> |
| <u>Vowel intended</u> <u>(produced by adult)</u> | | | | | | | | | |
| <u>/i/ (control)</u> | | | | | | | | | |
| Conversational | 100.0 | | | | | | | | |
| Clear | 100.0 | | | | | | | | |
| Conversational and Clear Combined | 100.0 | | | | | | | | |
| <u>/ɛ/</u> | | | | | | | | | |
| Conversational | 4.6 | 3.8 | 0.4 | 73.3 | 11.7 | 3.3 | 2.5 | 0.4 | |
| Clear | 0.4 | 0.8 | | 92.1 | 3.3 | 1.7 | 1.7 | | |
| Conversational and Clear Combined | 2.5 | 2.3 | 0.2 | 82.7 | 7.5 | 2.5 | 2.1 | 0.2 | |
| <u>/æ/</u> | | | | | | | | | |
| Conversational | | | | 16.7 | 69.2 | 8.8 | 1.7 | | |
| Clear | | | | 6.3 | 88.8 | 3.3 | 0.4 | | 0.4 |
| Conversational and Clear Combined | | | | 11.5 | 79.0 | 6.0 | 1.0 | | 0.2 |
| <u>/ɑ/</u> | | | | | | | | | |
| Conversational | 5.0 | 2.1 | | 7.1 | 10.8 | 48.3 | 26.7 | | |
| Clear | 1.3 | | | 2.5 | 11.7 | 78.3 | 5.8 | | 0.4 |
| Conversational and Clear Combined | 3.1 | 1.0 | | 4.8 | 11.3 | 63.3 | 16.3 | | 0.2 |
| <u>/ʌ/</u> | | | | | | | | | |
| Conversational | 3.8 | 2.1 | | 34.6 | 5.4 | 7.1 | 46.7 | | 0.4 |
| Clear | | 0.4 | | 5.0 | 2.5 | 5.0 | 87.1 | | |
| Conversational and Clear Combined | 1.9 | 1.3 | | 19.8 | 4.0 | 6.0 | 66.9 | | 0.2 |

Greater variability was noted in the responses to /a/ than to other target vowels because listeners repeated this vowel with the lowest accuracy. Listeners repeated /a/ as /ʌ/ for 16% of the trials and as /æ/ for 11% of the trials. The confusion for /a/ with /ʌ/ was primarily unidirectional in that /ʌ/ was repeated as /a/ for 6% of the trials. This unidirectional confusion was consistent for both conversational and clear speech; listeners repeated /a/ as /ʌ/ more frequently in conversational speech (27%) and clear speech (6%) than /ʌ/ as /a/ in conversational speech (7%) and clear speech (5%).

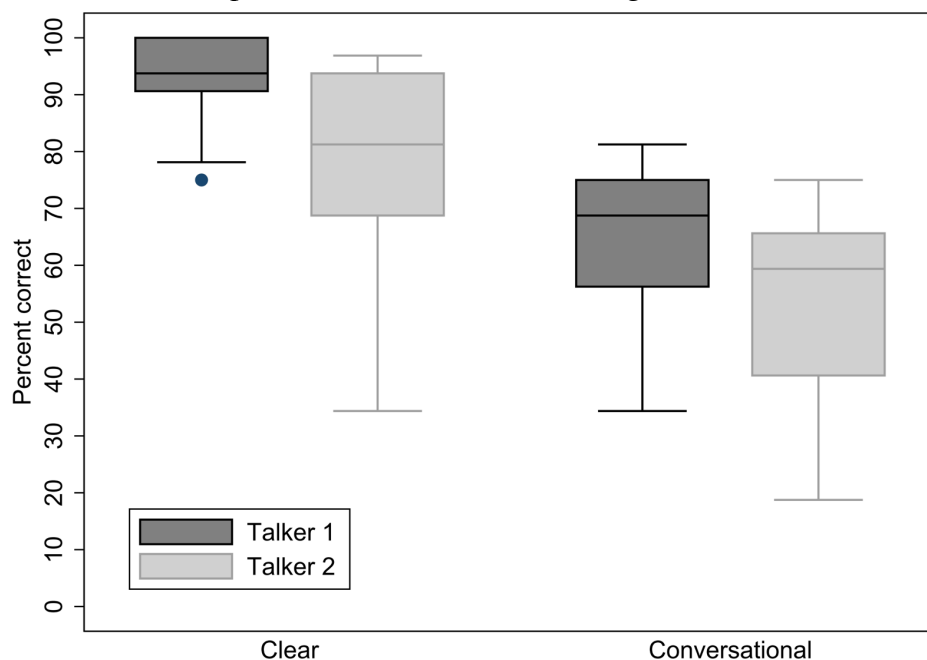
Talker effects on vowel repetition

On average, the children repeated Talker 1's vowels with 79% accuracy and Talker 2's vowels with 67% accuracy. The mixed effects logistic model confirmed a statistically significant fixed effect of the talker on vowel repetition ($z = 3.13, p < .001$). The children repeated vowels produced by Talker 1 significantly more accurately than those produced by Talker 2.

Talker effects for each speaking style

The mean difference between repetition accuracy of clear speech vowels and conversational speech vowels was 28% for Talker 1 and 27% for Talker 2. Even though these empirical findings suggest a small difference (1%) between talkers, the mixed effects logistic model indicated a statistically significant difference ($z = 2.10, p < .01$) after controlling for the random effects of listener and trial. Moreover, listeners' responses were relatively consistent. Figure 2 presents boxplots depicting the distribution of percent correct scores for the each talker in each speaking style, showing that almost all of the scores fell between the 25th and 75th percentile.

Figure 2. Box plots depicting the distributions of listeners' percent correct scores for conversational and clear speech split by talker. Horizontal lines within the rectangles represent median scores. The top and bottom of the rectangles correspond to the 25th and 75th percentiles. The solid circle represents scores outside this range.



Discussion

In summary, the present study's main finding was that children repeated clear speech vowels with higher accuracy than conversational vowels. Thus, clear speech benefited children's vowel perception in noise. Significant vowel and talker differences were also observed. Repetition accuracy depended on the vowel, with front vowels (/ɛ, æ/) repeated more accurately than central and back vowels (/ʌ, ɑ/). Furthermore, although a larger clear-speech benefit was observed for Talker 1 as compared to Talker 2, children benefited from clear speech for all vowels produced by both talkers.

Speaking style

The higher repetition accuracy for clear speech vowels as compared to conversational speech vowels is consistent with the clear speech literature and was predicted at the outset of the study. For example, the present study's finding of a 27% clear-speech benefit for child listeners is in line with Payton et al.'s (1994) finding that adults listening to sentences identified clear speech keywords with 21% higher accuracy than conversational speech keywords. The clear-speech benefit is also consistent with the two studies that have investigated children's perception of clear speech real words (Bradlow et al., 2003; Riley & McGregor, 2012).

The current study extended clear speech vowel research to children. A clear-speech benefit for *adults* identifying vowels in noise has likewise been reported in the literature (Ferguson 2004; Ferguson & Kewley-Port, 2002; Rogers et al., 2010). Ferguson and Kewley-Port (2002) documented a 14% vowel identification accuracy increase for clear speech, whereas Ferguson (2004) reported an 8.5% increase and Rogers et al. (2010) reported a 5-7% increase. Results from the present experiment suggest that clear speech also benefits school-age children. The benefit appears to be larger for adults (38%) than for children (27%) when results from the preliminary adult study and the present child study are compared. However, when the child study is compared to other adult studies that employed different procedures and stimuli (e.g., Ferguson & Kewley-Port, 2002), the benefit appears to be larger for children (27%) than for adults (8.5% or 5-7%).

The finding of a clear-speech benefit in the current nonsense word study is consistent with benefits found in the two similar clear speech studies on real words (Bradlow et al., 2003; Riley & McGregor, 2012), in which not only lexical influences, but also larger similarity neighborhoods would presumably have been activated (Luce & Pisoni, 1998). The care taken by talkers in clear speech to make the acoustic-phonetic information available to the listeners likely reduces the number of competitors in the listener's similarity neighborhood, yielding greater intelligibility. The current study complements previous studies in that in the real word studies, the lexical effects and the larger similarity neighborhoods activated by real words may have dampened the clear-speech benefits. Comparisons of real word and nonsense word studies, controlling for similarity neighborhood and changes in the lexical probabilities in children's developing lexicons (see Charles-Luce & Luce, 1990; 1995; Coady & Aslin, 2003; Dollaghan, 1994), may further reveal the specific influences of word frequency, lexical development, and similarity neighborhoods on word recognition.

Differences among vowels

The repetition accuracy differences for the four target vowels in this study (when conversational and clear vowels were analyzed together) suggest that children perceive /ɛ/ and /æ/ more accurately than /ʌ/ and /ɑ/. These results are partially consistent with the adult identification results obtained for Talkers 1 and 2 in the preliminary adult study suggesting that /ɛ/ and /ʌ/ are more easily perceived by adults than /æ/ and /ɑ/. However, adult identification

studies involving only conversational vowels in noise suggest that adults have difficulty perceiving both central and back vowels. Specifically, when listening to AE vowel syllables in noise, AE-speaking adults identified /ɛ/ and /æ/ more accurately than /ʌ/ and /ɑ/ (Cutler et al., 2004). Moreover, Bunton and Story (2009) found that when AE-speaking adults were presented with isolated synthetic productions of /ɛ, æ, ʌ/ in quiet, /æ/ was identified with the most accuracy and /ʌ/ with the least accuracy. Similarly, Neel (2008) reported that when AE-speaking adults identified vowels in /hVd/ context in quiet, they identified /ʌ/ with less accuracy than the other vowels presented (/ɛ, æ, ʌ/). Thus, children and adults demonstrated consistent difficulty perceiving the back vowel /ɑ/ and perceived the other target vowels (/ɛ, æ, ʌ/) with variable difficulty. Particular difficulty perceiving low back vowels has also been documented cross-linguistically and across listener populations (e.g., Bion, Escudero, Rauber, & Baptista, 2006; Higgins & Hodge, 2002; Levy et al., 2014; Lin, 2013), but is poorly understood, attributed variably to the cross-linguistic salience of high front vowels and the extremities of F2 frequencies.

Differences among clear speech vowels

Unlike for adults in previous studies (Ferguson & Kewley-Port, 2002; Rogers et al., 2010), for children in the present study, no particular vowel was more aided by clear speech than any other vowel. Children's identification accuracy for the 4 target vowels (as measured by their repetition accuracy) appeared to follow a different trend from adults' identification of the vowels. Adults in the preliminary study identified clear speech /ʌ/ with more accuracy (95%) than clear speech /ɛ/, /æ/, and /ɑ/ (81%, 81%, and 52% respectively) with /ʌ/ benefiting from clear speech the most (61%). Ferguson and Kewley-Port (2002) also reported adults' more accurate identification of clear speech /ʌ/ (94%), than clear speech /ɛ/ and /æ/ (76%, 85%, respectively). Furthermore, they noted the largest clear-speech benefit for /æ/ (54%) when compared to /ɛ, ʌ, ɑ/ (29%, 2.8%, 22%, respectively). In the present study, in clear speech, children repeated /ɛ/ with the most accuracy (92%) and the largest clear-speech benefit was evident for /ʌ/ (40%). However, it should be noted that, unlike the present study, Ferguson and Kewley-Port used monosyllabic real words in /bVd/ context presented in noise over headphones.

Thus, although children and adults benefit from clear speech, children appear to perceive speech in noise and clear speech differently from adults (Ferguson & Kewley-Port, 2002; Nishi et al., 2010). It follows that findings reported on adults' perception of clear speech vowels may have limited applicability to children. Children listening to clear speech vowels may attend to different articulatory or acoustic cues from those attended to by adults, perhaps as a function of the differences between children's and adults' similarity neighborhoods (Charles-Luce & Luce, 1990; 1995; Luce & Pisoni, 1998).

Vowel confusions

As noise degrades the speech signal, competition with the closest "neighbors" in a similarity neighborhood would be expected (Luce & Pisoni, 1998). Indeed, in the present study, vowels that were proximal in acoustic vowel space were most often confused. To the author's knowledge, no other study has reported adults' or children's confusions of clear speech vowels. When the present study's results were compared to conversational speech vowel confusions, they were found to be consistent with results from adult studies of conversational speech vowels in quiet (Bunton & Story, 2009; Hillenbrand et al., 1995; Neel, 2008; Peterson & Barney, 1952) and in noise (Cutler et al., 2004). Peterson and Barney (1952) reported that AE-speaking adults most often confused /ɛ/ with /æ/ and /æ/ with /ɛ/. Adults' confusion of /ɛ/ and /æ/ has also been documented in a number of other studies (Bunton & Story, 2009; Cutler et al., 2004; Hillenbrand

et al., 1995; Neel, 2008). Similarly, in the present study, children most frequently confused /æ/ with /ɛ/ and /æ/ was repeated in error as /ɛ/ most frequently.

When the present study's conversational speech vowel confusions were compared to the clear speech vowel confusions, clear speech /a/ was most often repeated as /æ/ more frequently than /ʌ/, whereas conversational speech /a/ was repeated as /ʌ/ more frequently than /æ/. Lastly, clear speech /ʌ/ was repeated as /ɛ/ and /a/, whereas conversational speech /ʌ/ was repeated more frequently as /ɛ/ than /a/. A possible explanation is that phonetically, the mid vowels remain mid vowels in clear speech, whereas the peripheral vowels become more peripheral. For example, /ʌ/ remained a mid vowel when produced in clear speech, with its acoustics not changing as extensively as did /ɛ, æ, a/ when produced in clear (as opposed to conversational) speech (see Table 1). Mid vowels appear to be less changed by clear speech than low vowels; thus, children's confusions in clear speech for mid vowels appear to follow different patterns from confusions involving low vowels. Acoustically, for example, F1 increased by approximately 94 Hz in clear speech for /æ/, but increased by only approximately 48 Hz for /ʌ/ in clear speech. It is possible that because both clear speech /æ/ and /a/ showed a greater increase in F1 than did clear speech /ʌ/, children confused /a/ with the "lowered" vowel /æ/ and not with the mid vowel /ʌ/. This change in acoustic vowel space for the more peripheral vowels may also clarify why children repeated /ʌ/ as /ɛ/ more frequently in conversational speech than in clear speech. Clear speech /ʌ/ remained a mid vowel and clear speech /ɛ/ was "lowered." As a result, acoustic vowel space between these two vowels increased; thus, children confused /ʌ/ with /ɛ/ less frequently in clear speech than in conversational speech.

Durational differences were noted between conversational and clear speech vowels. As predicted, both talkers produced longer clear speech vowels ($\bar{x} = 180$ milliseconds) than conversational speech vowels ($\bar{x} = 110$ milliseconds). The literature suggests that lengthening or shortening vowel duration increases conversational speech vowel confusions for adult listeners. Hillenbrand, Clark, and Houde (2000) asked AE adults to identify four sets of synthetic /hVd/ syllables that varied only in duration. Vowels that were manipulated to be longer or shorter in duration than the original recording were identified less accurately than the vowels with neutral duration. These results are consistent with those of the current study. The vowel-pairs /ɛ, æ/ and /a, ʌ/ are comprised of spectrally similar vowels that differ (within pair) in duration in conversational speech (Crystal & House, 1988). Thus, clear speech vowel confusions within the pairs might be expected because duration is increased during clear speech vowel production.

Talker differences

Children had more difficulty repeating conversational and clear speech vowels produced by Talker 2 than by Talker 1. Talker differences had been predicted based on conversational and clear speech vowel perception research (Ferguson, 2004; Hillenbrand et al., 1995; Neel, 2008; Uchanski, 2005) and on the findings of the preliminary study described here on adults' identification of the same stimuli. Uchanski (2005) expressed the need for the inclusion of multiple talkers in clear speech research because some talkers appear to produce a larger clear-speech advantage than others. Nonetheless, in the present study, even though significant differences between talkers were found, each talker provided listeners with a clear-speech benefit. These results suggest that, although the degree of benefit varies from talker to talker, an overall clear-speech benefit may be present for children despite talker differences.

The differences between effects of the two talkers' clear speech on listeners in the present study may be attributable, in part, to the different vowel durations produced by the talkers (as

shown in Table 1) for both tense and lax vowels. Talker 2 increased her vowel duration in clear speech more extensively than did Talker 1 and yielded a more limited clear-speech benefit. Similar findings have been reported for adults' clear speech perception. Talkers who produced significantly longer clear speech vowels provided a more limited clear-speech benefit than those who did not increase their vowel duration as extensively in clear speech (Ferguson & Kewley-Port, 2007).

Implications

In degraded listening conditions, such as in the presence of noise, listeners may rely more on durational than on spectral cues (Winn, Chatterjee, & Idsardi, 2012). Generally, a moderate rate of speech appears to facilitate intelligibility, with slightly slower rates thought to increase young children's comprehension (Berry & Erickson, 1973), although optimal vowel durations for child listeners in conversational speech or in clear speech are likely variable. Slight increases in vowel duration appear to yield the greatest intelligibility, permitting listeners to resolve a perturbed speech signal. Yet even "excessive" vowel lengthening yields greater intelligibility than does no lengthening (as in conversational speech vowels).

Limitations

Some limitations of the study should be noted. Only a subset of AE vowels (/ε, æ, α, Λ/) and only two talkers were included because of the need to restrict the number of trials for the children. Thus, results from adult studies that include a larger subset of AE vowels may not be comparable with those of the present study. Additionally, stimuli produced by more talkers would be more representative of the population at large, as clear speech is produced in diverse ways (Ferguson, 2004; Uchanski, 2005). Furthermore, results from the repetition task used in the present study (and in similar studies such as Bradlow et al., 2003, Neuman and Hochberg, 1983, and Nishi et al., 2010) may have been somewhat confounded by the children's production skills. The use of nonsense words in a sound-treated booth added the control of lexical effects (Dollaghan & Campbell, 1998) and noise to the study, permitting a focus on perception of the speech signal at the acoustic-phonetic level, but resulted in a less naturalistic listening setting than real words in a classroom, for example.

Conclusion and future directions

Research is just in the beginning stages of documenting the benefits of adults' use of clear speech for children's speech perception. Future studies may further investigate the effects of more vowels and consonants, along with suprasegmentals and diverse talkers. In addition, because children's ability to perceive speech in noise changes with age (Bradley & Sato, 2008; Stuart et al., 2006), a variety of age groups of listeners and SNRs should be incorporated into future studies.

Lastly, the extension of clear speech perception studies to school-age children with disabilities is warranted. Approximately 2.4 million school-age children in the United States have some type of learning disability, with this percentage typically increasing from year to year (National Center for Learning Disabilities, 2012). To date, the only study on this topic reported that children with learning disabilities demonstrated a strong clear-speech benefit when listening to sentences (Bradlow et al., 2003). As more information becomes available about children's perception of clear speech, the specifics of how to use this speech style effectively for children with and without disabilities will be better understood.

Acknowledgements

The authors wish to thank the Speech Production and Perception Lab of Teachers College, Columbia University, with special thanks to Miriam Baigorri, Sih-Chiao Hsu, Gemma Moya-Galé, Georgia Duan, Bethany Hetrick, and Elanna Seid. The authors are also thankful for the suggestions by Luca Campanelli, Robert Remez, Karen Froud, and Honor O'Malley.

References

- American National Standards Institute. (2004). Specifications for audiometers (ANSI S.36–2004). New York, NY.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Berry, M. D., & Erickson, R. L. (1973). Effects of children's comprehension of normal speech. *Journal of Speech, Language, and Hearing Research*, *16*, 367–374.
- Bion, R., Escudero, P., Rauber, A. S., & Baptista, B. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. *Proceedings of Interspeech*, Pittsburgh, PA.
- Boersma, P., & Weenink, D. (2006). Praat (Version 5.2.22) [Computer software]. Retrieved from <http://www.praat.org>
- Bradley, J. S., & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *Journal of the Acoustical Society of America*, *123*, 2078–2086.
- Bradlow, A. R. and Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, *121*, 2339–2349.
- Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, *112*, 272–283.
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for learning-impaired children: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, *46*, 80–97.
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*, 255–272.
- Bunton, K. & Story, B. H. (2009). Identification of synthetic vowels based on selected vocal tract area functions (L). *Journal of the Acoustical Society of America*, *125*, 19–22.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language*, *17*, 205–215.
- Charles-Luce, J. & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children. *Child Language*, *22*, 727–35.
- Coady, J. A. & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, *30*, 441–469.
- Crandell, C. C., & Smaldino, J. J. (2000). Room acoustics for listeners with normal-hearing and hearing impairment. In M. Valente, H. Hosford-Dunn, & R.J. Roeser (Eds.), *Audiology Treatment* (pp. 601–637). New York, NY: Thieme.
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, *116*, 3668–3678.
- Dollaghan, C. A. (1994). Children's phonological neighbourhoods: half empty or half full? *Journal of Child Language*, *21*, 257–71.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 5, 1136–1146.
- Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, *116*, 2365–2373.

- Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *Journal of Speech, Language, and Hearing Research, 55*, 779–790.
- Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*, 259–271.
- Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*, 1241–1255.
- Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired individuals. *Journal of the Acoustical Society of America, 135*, 3570–3584.
- Fudala, J. B. (2001). *Arizona Articulation Proficiency Scale, 3rd edition*. Los Angeles, CA: Western Psychological Services.
- Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 53*, 588–608.
- Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics, 12*, 91–114.
- Helfer, K., & Wilber, L. (1990). Hearing loss, aging, and speech perception in reverberation and in noise. *Journal of Speech and Hearing Research, 33*, 149–155.
- Higgins, C. M., & Hodge, M. M. (2002). Vowel space area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology, 10*, 271–277.
- Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108*(6), 3013–3022.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099–3111.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.
- Johnson, C. E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research, 43*, 144–157.
- Kewley-Port, D., Burkle, T. Z., & Hee Lee, J. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America, 122*, 2365–2375.
- Krause, J. C. & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America, 112*, 2165–2172.
- Krause, J. C. & Braida, L. D. (2009). Evaluating the role of spectral and envelope characteristics in the intelligibility role of clear speech. *Journal of the Acoustical Society of America, 125*, 3346–3357.
- Larsen, J. B., & Blair, J. C. (2008). The effect of classroom amplification on the signal-to-noise ratio in classrooms while class is in session. *Language, Speech, and Hearing Services in Schools, 39*, 451–460.
- Leone, D., Hsu, S., Baigorri, M., Moya-Galé, G., & Levy, E. S. (2012). Children's perception of clear-speech vowels. *Proceedings of Meetings on Acoustics, 12*, 060003.

- Levy, E. S. & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36, 141–157.
- Levy, E. S., Chen, W., Leone, D., Moya-Gale, G., Hsu, S.-C., & Ramig, L. O. (2014). Vowel intelligibility in children with dysarthria. Manuscript submitted for publication.
- Lin, C. Y. (2013). Perception and production of English front vowels by Taiwanese EFL learners. *Theory and Practice in Language Studies*, 3(11), 1952–1958.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Netherlands: Kluwer Academic Publisher.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, 19, 1–36.
- Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp.142–147). Cambridge, MA: MIT Press.
- McKay, K. E., Halperin, J. M. Schwartz, S. T., & Sharma, V. (1994). Developmental analysis of three aspects of information processing: Sustained attention, selective attention and response organization. *Developmental Neuropsychology*, 10(2), 121–132.
- National Center for Education Statistics. (2009). Program for international student assessment. Retrieved from <http://nces.ed.gov/surveys/pisa>
- National Center for Learning Disabilities. (2012). What are learning disabilities? Retrieved from <http://www.nclld.org/types-learning-disabilities/what-is-ld/what-are-learning-disabilities?start=1>
- Neel, A. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, 51, 574–585.
- Nelson, P. B., & Soli, S. (2000). Acoustical barriers to learning: Children at risk in every classroom. *Language, Speech, and Hearing Services in Schools*, 31, 356–361.
- Neuman, A., & Hochberg, I. (1983). Children's perception of speech in reverberation. *Journal of the Acoustical Society of America*, 73, 2145–2149.
- Nishi, K., Lewis, D. E., Hoover, B. M., Choi, S., & Stelmachowicz, P. G. (2010). Children's recognition of American English consonants in noise. *Journal of the Acoustical Society of America*, 127, 3177–3188.
- Payton, K. L., Uchanski, R. M., Braid, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America*, 95, 1581–1592.
- Pettinato, M. & Hazan, V. (2013). The development of clear speech strategies in 9-14 year olds. *Proceedings of Meetings on Acoustics*. 19, 060121.
- Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- Picheny, M. A., Durlach, N. I., & Braid, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28, 96–103.
- Picheny, M. A., Durlach, N. I., & Braid, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29, 434–446.
- Plichta, B. (2004). Akustyk (Version 1.8) [Computer software]. Retrieved from <http://bartus.org/akustyk/index.php>

- R Core Team. (2012). R: A Language and Environment for Statistical Computing (Version 2.15.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.
- Riley, K. G. & McGregor, K. K. (2012). Noise hampers children's expressive word learning. *Language, Speech, and Hearing in Schools, 43*, 325–337.
- Rogers, C. L., DeMasi, T. M., & Krause, J. C. (2010). Conversational and clear intelligibility of /bVd/ syllables produced by native and non-native English speakers. *Journal of the Acoustical Society of America, 128*, 410–423.
- Sjölander, K., & Beskow, J. (2006). Wavesurfer (Version 1.8.5) [Computer software]. Retrieved from <http://www.speech.kth.se/wavesurfer>
- Smiljanic, R. & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass, 10*, 236–264.
- Soli, S. D. & Sullivan, J. A. (1997). Factors affecting children's speech communication in classrooms. *Journal of the Acoustical Society of America, 101*, S3070.
- Sony. (2009). Soundforge (Version 8.0) [Computer software]. Madison, WI: Sony Pictures Digital Inc.
- Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R., & Pittman, A. I. (2000). The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children. *Journal of Speech, Language, and Hearing Research, 43*, 902–914.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America, 122*, 1111–1129.
- Stuart, A., Givens, G. D., Walker, L. J., & Elangovan, S. (2006). Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise. *Journal of the Acoustical Society of America, 119*, 1946–1949.
- Tagliaferri, B. (2011). Paradigm (Version 1.0.2) [Computer software]. Retrieved from <http://www.paradigmexperiments.com/>
- Uchanski, R. M. (2005). Clear speech. In D.B. Pisoni & R.E. Remez (Eds.), *The handbook of speech perception* (pp. 207–235). Malden, MA: Blackwell.
- Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation and lexical access for spoken words. *Brain and Language, 68*, 306-311.
- Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *Journal of the Acoustical Society of America, 131*, 1465–1479.