Purpose: This study investigated acoustic–phonetic modifications produced in noise-adapted speech (NAS) and clear speech (CS) by children, young adults, and older adults.

Method: Ten children (11–13 years of age), 10 young adults (18–29 years of age), and 10 older adults (60–84 years of age) read sentences in conversational and clear speaking style in quiet and in noise. A number of acoustic measurements were obtained.

Results: NAS and CS were characterized by a decrease in speaking rate and an increase in 1–3 kHz energy, sound pressure level (SPL), vowel space area (VSA), and harmonics-to-noise ratio. NAS increased fundamental frequency (F0) mean and decreased jitter and shimmer. CS increased frequency and duration of pauses. Older adults produced the slowest speaking rate, longest pauses, and smallest increase in F0 mean, 1–3 kHz energy, and SPL when speaking clearly. They produced the smallest increases in VSA in NAS and CS. Children slowed down less, increased the VSA least, increased harmonics-to-noise ratio, and decreased jitter and shimmer most in CS. Children increased mean F0 and F1 most in noise.

Conclusions: Findings have implications for a model of speech production in healthy speakers as well as the potential to aid in clinical decision making for individuals with speech disorders, particularly dysarthria.

This study is the first of two articles examining speaking style adaptations that children, young adults, and older adults implement when speaking in response to noise (noise-adapted speech, NAS) and when instructed to speak clearly (clear speech, CS). The focus of the current research article is on identifying some of the acoustic–articulatory modifications characterizing a response to the two communicative barriers, both separately and in conjunction. In the companion article (Smiljanic & Gilbert, 2017), we examine whether intelligibility for young adult listeners is enhanced through these adaptations. Combined, these findings may be useful in guiding clinicians in choosing treatment for individuals whose intelligibility is compromised and in situating these treatment options within cognitive–perceptual models of intelligibility (Cooke, King, Garnier, & Aubanel, 2013; Lansford, Liss, Caviness, & Utianski, 2011).
been established (Cooke et al., 2013; Gilbert, Chandrasekaran, & Smiljanic, 2014; Godoy, Koutsogiannaki, & Stylianou, 2014). For instance, increase in the distinctiveness of phonological vowel and consonant contrasts has been shown much more reliably in CS (Ferguson & Kewley-Port, 2002; Lu & Cooke, 2008; Maniwa, Jongman, & Wade, 2008; Smiljanic & Bradlow, 2005, 2008), whereas increased spectral energy in a band spanning the range of formants is found more commonly in NAS (Krause & Braida, 2004; Lu & Cooke, 2009; Summers et al., 1988). It is important to note, though, that NAS and CS studies do not always compare the same set of acoustic–phonetic features, nor do they use consistent instructions to elicit CS or NAS; hence, the cross-study comparisons are not possible (Pichora-Fuller et al., 2010). A few studies that compared acoustic modifications for NAS and CS directly reported mixed results (Godoy et al., 2014; Goy, Pichora-Fuller, & van Lieshout, 2013; Gilbert et al., 2014). Goy, Pichora-Fuller, et al. (2013) found a significantly lower mean F0 on target words in CS and higher F0 mean in loud speech compared with any other condition, and no differences in F0 range across the conditions. They also reported a significantly slower speaking rate in CS compared with loud and CO speech, Gilbert et al. (2014), on the other hand, showed significantly slower speech rates, higher mean pitch, and more energy in the 1–3 kHz range for both speaking style adaptations. Additionally, a significantly larger F0 range in CS compared with CO speech but smaller F0 range in NAS compared with quiet speech was found. With an eye toward using cue enhancement algorithms for improving speech intelligibility, Godoy et al. (2014) reported trends for spectral energy boosting in NAS only and vowel space expansion for CS only. Note that Goy, Pichora-Fuller, et al. (2013) and Gilbert et al. (2014) examined speech of only one talker, whereas Godoy et al. (2014) used different materials to examine NAS and CS.

Given that both speaking in noise and with listeners who are hard of hearing or nonnative speakers of the target language characterize everyday communication interactions, the current investigation is relevant from both the theoretical and clinical perspectives. Within the hypo- and hyper-articulation (H&H) theory (Lindblom, 1990; Perkell, Zandipour, Matthies, & Lane, 2002), CO speech is at the hypo-articulation end of the continuum, and CS and NAS are at the hyper-articulation end. Young healthy adults can successfully modify their output from hypo- to hyper-articulated speech in response to different communication challenges (Hazan & Baker, 2011). Lam, Tjaden, and Wilding (2012) demonstrated that young adult talkers produced gradient acoustic–articular modifications in response to different instructions to clarify their speech (overnunciate vs. hearing impairment conditions). Focusing on suprasegmental characteristics, Hazan, Tuomainen, and Pettinato (2016) showed that children were still developing a full range of intelligibility-enhancing strategies in response to the degraded transmission condition (vocoded speech) at the ages of 9–14. This suggests that adolescents do not yet have a fully developed skill set to produce gradient responses on the hypo- to hyper-articulation continuum. The current study builds on this work by examining whether talkers produce similar hyper-articulatory modifications when speaking in response to noise and when speaking clearly and whether developmental and age-related factors affect hypo- and hyper-articulation abilities.

CS and loud speech have also been used as behavioral treatment techniques for maximizing intelligibility in speakers whose production of voice pitch, loudness, speaking rate, vowel formants, and distribution of energy are disrupted (Beukelman, Fager, Ullman, Hanson, & Logemann, 2002; Hustad & Weismer, 2007; Park, Theodoros, Finch, & Cardell, 2016; Sadagopan & Huber, 2007) and in aural rehabilitation programs (Schum, 1997). In response to CS instructions, talkers with dysarthria produced increases in vowel space area (VSA), vocal intensity, and mean F0 and a decrease in speaking rate (Goberman & Elmer, 2005; Tjaden, Kain, & Lam, 2014; Tjaden, Lam, & Wilding, 2013; Tjaden, Sussman, & Wilding, 2014). Lam and Tjaden (2016) found that, similar to healthy adults, talkers with Parkinson’s disease (PD) also produced different degrees of modifications in response to the varied CS instructions. Talkers with PD were found to successfully increase vocal effort in response to masking noise (Stathopoulos, Huber, & Sussman, 2011) although the response differed between the control participants and individuals with PD (Adams, Moon, Page, & Jog, 2006; Darling & Huber, 2011). Evidence is emerging that these speech-oriented, behavioral treatments may lead to enhanced intelligibility for talkers with dysarthria (Park et al., 2016; Tjaden, Sussman, et al., 2014). CS and NAS thus have great potential as therapy techniques for addressing speech impairment in patients with dysarthria. Examining the similarities and differences between the two adaptations across the same set of acoustic–articular features and with consistent elicitation methods can provide useful benchmarks for clinicians in treatment goals for individuals with dysarthria. Furthermore, better understanding of the impact of the treatment targets on the speech output and their consequences for speech processing can provide support for a cognitive–perceptual framework within which to conceptualize treatment goals (Cooke et al., 2013; Lansford et al., 2011).

**Age Effect on Speaking Style Adaptations**

Children, young adults, and older adults differ significantly in their speech production systems (e.g., vocal tract length, speech motor control) and auditory and cognitive abilities, leading to qualitatively different speech patterns (Benjamin, 1982; Gordon-Salant & Fitzgibbons, 1997; Goy, Fernandes, Pichora-Fuller, & van Lieshout, 2013; Lee, Potamianos, & Narayanan, 1999; Pichora-Fuller, Schneider, & Daneman, 1995; Schneider, Daneman, & Pichora-Fuller, 2002). Children have more variable sound productions, greater within-category dispersion, and smaller distances between contrasting phonemes compared with
adults (Lee et al., 1999; Nissen & Fox, 2005; Romeo, Hazan, & Pettinato, 2013). Some of the differences may be due to the children’s still-developing motor speech control, which can extend well into adolescence (Murdoch, Cheng, & Goozee, 2012; Walsh & Smith, 2002). Similarly, the quality of the speech signal is affected by aging for older adults. Listeners can easily distinguish between older and younger talkers, revealing that voice quality and speech patterns change with age (Goy, Fernandes, et al., 2013; Ryan & Burk, 1974). These age-related differences can arise from an interaction of sensorimotor changes and declines in auditory feedback (Liss, Weismer, & Rosenbek, 1990). Older adults have also been shown to have increased segment durations, slower overall speaking rate, spirantized stops, and more centralized vowels (Benjamin, 1982; Halle & Myerson, 1996; Liss et al., 1990).

Speech patterns with these characteristics (both in healthy and clinical populations) may lead to lower accuracy in perception and intelligibility as well as more effortful processing on the part of the listener (Hazan, Romeo, & Pettinato, 2013; Liss et al., 1990; McAuliffe, Wilding, Rickard, & O’Beirne, 2012; Newman, Clouse, & Burnham, 2001; Torre & Barlow, 2009).

While some work has examined nonnative (Rogers, DeMasi, & Krause, 2010; Smiljanic & Bradlow, 2011) and gender-related differences in CS production (Ferguson, 2004), large gaps remain in our understanding of age-related changes in the production of intelligibility-enhancing speaking styles. Previous work with young children showed that 3-, 4-, and 5-year-olds showed similar or larger vocal intensity increases in response to noise compared with adults (Garber, Speidel, & Siegel, 1980; Garber, Speidel, Siegel, Miller, & Glass, 1980; Siegel, Pick, Olsen, & Sawin, 1976). Redford and Gildersleeve-Neumann (2009) found that listeners were not able to differentiate CS from casual productions in 3-year-olds but could do so in 4- and 5-year-olds, even though they exhibited different acoustic-articulatory adjustments compared with adults (e.g., faster speaking rate and lower overall F0 in CS compared with casual speech). Syrett and Kawahara (2013) showed that 3- to 5-year-old children produced conversational-to-clear speech modifications typically found for adults (e.g., expanded F0 range; longer, louder, and more dispersed vowels). In contrast, Pettinato and Hazan (2013); Pettinato, Tuominen, Granlund, and Hazan (2016); and Hazan et al. (2016) found that differences between children’s and adult’s CS productions (e.g., vowel hyper-articulation) persisted even for 9–10 and 13–14 years old. These findings suggest that many features of adultlike speaking style adaptations continue to develop into adolescence. Since few of these speaking adaptation studies examined children over age 6, the current study seeks to provide additional findings on older children who exhibit more developed articulatory control and planning and who can participate in more comparable tasks to the adults.

A number of studies examined acoustic-articulatory modifications that older adults produced when instructed to speak clearly, to speak slowly, or in response to different levels of background noise (Adams, Dykstra, Jenkins, & Jog, 2008; Adams, Winnell, & Jog, 2010; Darling & Huber, 2011; Sadagopan & Huber, 2007; Tjaden et al., 2013; Turner, Tjaden, & Weismer, 1995). In these studies, older adults served as age-matched controls to the groups of individuals with PD, amyotrophic lateral sclerosis (ALS), or multiple sclerosis (MS). Older adults were shown to significantly modify their speech (e.g., increased vocal intensity, VSA, abdominal effort, and decreased speaking rate) in response to these varied challenges. While the older adults were more successful in implementing some of the adjustments compared with the individuals with PD, ALS, or MS, it is not clear from these studies how they would compare to the healthy young talkers. Schum (1996) found that older adults provided the same CS intelligibility benefit as young adults, whereas Smiljanic (2013) found a significantly smaller benefit for older adult talkers. In terms of the acoustic-phonetic adjustments, Smiljanic (2013) provided only speaking rate analysis, which revealed that older adults produced overall slower CO speech as well as a smaller CS speaking rate decrease compared with the young adults. Schum (1996) only reported that “similar acoustic effects” were found for older adults as reported in other CS literature (cf., Picheny et al., 1986; Schum, 1996, p. 215). Examining CS and NAS differences in children, young adults, and older adults will add to our understanding of the scope of variation within the neurologically healthy talkers related to developmental and aging factors.

The aim of the current production study was to examine acoustic-articulatory modifications that characterize CO and CS sentences produced both in quiet and in noise by children, young adults, and older adults. Understanding adaptations to different communicative challenges may be important in clinical decision making for individuals with speech disorders, particularly dysarthria. This examination can also enhance our understanding of whether auditory, articulatory, and cognitive changes across life span affect how talkers respond to the noise in the environment and when they attempt to increase their intelligibility by speaking clearly. In the companion article (Smiljanic & Gilbert, 2017), we report on the effect that NAS and CS have on intelligibility for the three talker groups. The two studies contribute toward a deeper understanding of the effects of inter- and intra-talker variability and environmental factors on intelligibility.

**Method**

**Talkers**

Ten children (11–13 years old, mean 12.3 years), 10 young adults (18–29 years old, mean 21.0 years), and 10 older adults (60–84 years old, mean 70.2 years) participated in the study. All talkers were native monolingual speakers of English and balanced for gender within each group. The hearing thresholds were obtained for both ears using an ascending method in 5-dB steps for octave frequencies of 500, 1000, 2000, and 4000 Hz. Children and young adults had typical hearing (thresholds below 25 dB HL.
at octave frequencies between 500 and 4000 Hz). Four out of 10 adult talkers had hearing within normal limits through 4000 Hz. Three out of 10 older adults had hearing within normal limits through 2000 Hz with 4000-Hz thresholds ranging between 40 and 45 dB HL. Three talkers had some degree of high-frequency hearing loss typical for presbycusis (ranging between 35–45 dB HL at 2000 Hz and 40–75 dB HL at 4000 Hz). None of the older adults were hearing aid users. The young adults were all University of Texas at Austin undergraduate students. Children and older adults were recruited from the Austin, Texas, community. All materials and procedures were approved by the Institutional Review Board at the University of Texas at Austin.

### Stimuli and Masker

Each talker produced 60 meaningful sentences (e.g., *Farm animals stay in a barn*) developed for use with children (Fallon, Trehub, & Schneider, 2002). Each talker produced all of the sentences in CO speech first and then in CS. The conversational and clear sentences were first elicited in quiet. Next, both styles were elicited in response to speech-shaped noise (SSN) interference presented diotically over Sennheiser HD 280 pro headphones (80 dB SPL). SSN was generated by obtaining the long-term average spectrum of 6-talker babble (three males, three females; Van Engen & Bradlow, 2007) and shaping white noise to match that spectrum. For CO speech, talkers were instructed to speak in a casual manner as if they were talking to a friend or a family member. For CS, they were instructed to speak as if they were talking to someone who has a low proficiency in English and has difficulty following them conversationally. When eliciting NAS, the talkers read on a slide: “Now, you will hear some background noise. Pretend that you are in a noisy place, talking to your friends or family.” This announcement was followed by the same instructions for CO and CS as provided when recordings were done in quiet. Participants were asked if they had any questions, and none of them indicated that they did not understand the instructions. The same elicitation order was maintained for all talkers.3

This set of instructions follows elicitation methods used previously in our work and work by other researchers examining CS and NAS (Gilbert et al., 2014; Pichora-Fuller et al., 2010; Smiljanic & Bradlow, 2009). It is well established that varied instructions to elicit speaking style modifications will result in different acoustic–phonetic modifications implemented by talkers (Hazan & Baker, 2011; Lam & Tjaden, 2013; Lam et al., 2012). It is important to keep this in mind when comparing the results of this study with previous work that may have used different instructions. However, in line with the goals of this study, the consistent set of instructions used here allowed us to compare speaking style adaptation strategies across our talker groups and across the two communication barriers.

### Procedure

A hearing test with a portable screening audiometer (GSI 18, Grason-Stadler) followed by the recording took place in a sound-attenuated booth in the phonetics lab at the University of Texas at Austin. Sentences were presented to the talker one at a time on a computer monitor. Recordings were made using a Shure SM10A head-mounted microphone and a MOTU UltraLite-MK3 Hybrid recorder. All four sets (CO and CS in quiet and in noise) of 60 sentences per talker were recorded in one session lasting approximately 1 hr. Only one repetition of each sentence in each speaking style condition was elicited. Talkers were instructed to repeat a sentence if they mispronounced it. Breaks were provided as needed. The total set of recordings contained 7,200 sentences (60 unique sentences × 4 speaking styles × 30 talkers). The recorded sentences were segmented into individual files for acoustic analyses.

### Acoustic Analyses

The specific global acoustic–phonetic parameters that we targeted in this analysis were F0 range (difference between the highest and lowest tonal targets in the sentence), F0 mean, speech rate (overall sentence duration, number and duration of pauses), energy in the 1–3 kHz range, and SPL. Speaking rate was calculated as the

---

2. While the use of headphones to play masking noise could introduce frequency-dependent own-voice attenuation (Arlinger, 1986; Boëi, Boëi, & Pollak, 2006), it is a commonly used method in eliciting Lombard speech (e.g., Summers et al., 1988). Lu and Cooke (2008) compared the closed headphone setup with a compensated transmission channel and found no significant differences in the production modifications. Furthermore, in our study, all talker groups heard masking noise using the same setup, and the attenuation characteristics would have impacted all equally. Finally, the acoustic-articulatory modifications reported here are in line with what has been reported previously (Cooke et al., 2014; Pichora-Fuller et al., 2010; Smiljanic & Bradlow, 2009). This, coupled with the intelligibility benefit produced by the two modifications reported in the companion article (Smiljanic & Gilbert, 2017), suggests that this effect would have been rather negligible.

3. The decision to keep the elicitation in the same order and constant across participants was based on our previous work (Gilbert et al., 2014; Smiljanic & Bradlow, 2005) and piloting which revealed that it is more difficult for talkers to produce more conversational and casual speech after they had been asked to speak clearly or spoken in response to noise. If there was any effect of fatigue resulting in smaller or fewer adaptations, it would have likely impacted the most hyper-articulated speaking style (NAS CS). This suggests that any across-style differences we find in the present study may be more pronounced in real-life situations.

4. This measure was chosen since energy increases in this frequency band have been shown to characterize CS modifications (Hazan & Baker, 2011; Hazan et al., 2016; Krause & Braid, 2004). Energy increase in this frequency range reflects a reduction of spectral tilt, which characterizes speech produced with increased vocal effort (e.g., Glave & Rietveld, 1975; Sluijter & van Heuven, 1996).
number of syllables produced per second after the pauses were excluded. A pause was defined as a period of silence of at least 100 ms in duration, excluding silent periods before word-initial stop consonants where it would be impossible to determine the end of a pause and the beginning of the stop closure (similar to Smiljanic & Bradlow, 2005). Energy in the 1–3 kHz range was measured by averaging the long-term average spectrum energy between 1 and 3 kHz across each sentence. The mean dB SPL of each sentence was obtained by averaging the root-mean-square (rms) intensity for each sentence. Praat scripts (Boersma & Weenink, 2014) were used to obtain acoustic values automatically from each sentence file. Any measurements that deviated by 2 standard deviations from the mean for the particular talker/speaking condition were hand checked and corrected for errors.

Vowel measurements were obtained from a subset of vowel tokens. These consisted of corner vowels (/i, a, ae, u/) in monosyllabic words embedded between two obstruents. Vowel durations were measured from the onset of voicing and periodicity, as seen in the spectrogram and the waveform, to the beginning of the closure for the following stop, as marked by the cessation of regular periodicity and a substantial decrease in the amplitude of the waveform, or a start of the aperiodic waveform in case of fricatives. Measurements were based on two tokens per vowel per style per speaker (960 total), one from a sentence-final content word and one from a mid-sentence content word. The speaking style effects on vowel characteristics were always compared across the same vowel token (e.g., vowel in cheese produced in CO and CS in quiet and in noise), thus controlling for the sound environment and prosodic context effects. F1 and F2 frequencies were automatically extracted from the midpoint of each vowel using a linear predictive coding (LPC) formant tracking algorithm in Praat, thus precluding errors introduced by hand measurements. Values that differed by more than 200 Hz from the mean of the category were hand checked and corrected if necessary. Additionally, VSA was calculated for each condition using the formula for the area of an irregular quadrilateral in Vorperian and Kent (2007), as follows:

\[
\text{Area} = 0.5 \times \left\{ (\frac{i}{F2} + \frac{a}{F1} + \frac{a}{F2} + \frac{a}{F1}) \right. \\
\left. + (\frac{i}{F2} + \frac{u}{F1} + \frac{u}{F2} + \frac{i}{F1}) \right) \\
\left. - (\frac{i}{F1} + \frac{a}{F2} + \frac{a}{F1} + \frac{a}{F2}) \right) \\
\left. + (\frac{a}{F1} + \frac{u}{F2} + \frac{u}{F1} + \frac{i}{F2}) \right).
\]

Following Goy, Fernandes, et al. (2013), voice measurements, harmonics-to-noise ratio (HNR; dB), jitter percent, and shimmer percent were obtained from all /a/ tokens (2 per style per speaker, 240 total). Seventeen vowels were excluded from these analyses due to the presence of creak throughout most of the vowel duration. All measures were obtained automatically using Praat. Jitter was measured as the average absolute difference between consecutive periods, divided by the average period, using the jitter (local) method. Shimmer was measured as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude, using the shimmer (local) method. Harmonicity was measured as the mean HNR.

**Data Analysis**

Each acoustic feature was analyzed using mixed-effect linear regression in SPSS with Talker Age Group (children, young adults, or older adults), Noise (produced in quiet or in response to noise), Style (conversational or clear), and their interactions as fixed effects. To account for talker and item variability, random intercepts for Talker and Sentence were included as well. Random slopes were included in the model for both Style and Noise at the level of Talker, since this level showed the greatest variance. For F1 and F2 analyses, Vowel Type (/i, a, ae, u/) was added as an additional fixed effect to examine how formant frequencies differed for each of the four corner vowels. These models determined the impact of age and communicative barrier on production.

**Results**

**Global Measurements**

Table 1 lists F0 range and mean, speaking rate, pause rate (number of pauses per talker), pause duration, energy in the 1–3 kHz range, and SPL in CO and CS produced in noise and in quiet for each talker group.

F0 mean was significantly affected by Talker Age Group, \(F(2, 27) = 8.830, p = .001\). As expected, children showed a significantly higher mean F0 than both adult groups. A significant main effect of Noise was also found, \(F(1, 54.001) = 122.514, p < .001\), with higher F0 mean in NAS compared with speech produced in quiet. There was no significant main effect of Style, \(F(1, 54.001) = 1.130, p = .292\). A significant two-way interaction between Noise and Talker Age Group, \(F(2, 54.001) = 8.184, p = .001\), indicated that children made a significantly larger F0 increase than adults in response to noise. Noise also interacted with Style, \(F(1, 7041.013) = 8.385, p = .004\). Pairwise comparisons revealed that while F0 range was increased in quiet-to-noise adjustments for both CO and CS, the increase was larger for CO. Finally, a three-way interaction was found between Talker Age Group, Noise, and Style, \(F(2, 7041.013) = 9.311, p < .001\). All talker groups showed significantly higher F0 in noise relative to quiet; however, the quiet-to-noise F0 increase for older adults’ CS was smaller than any of the other quiet-to-noise increases across talker groups. The interaction between Noise and Style (in which the quiet-to-noise changes were larger for CO than for CS) arose from the older adult’s significantly smaller quiet-to-noise increase in CS. F0 range analyses showed a significant main effect of both Noise, \(F(1, 53.995) = 6.444, p = .014\), and Style, \(F(1, 53.995) = 5.627, p = .021\), but not of Talker Age Group, \(F(2, 27) = 1.572, p = .226\). NAS and CS had wider F0 range compared with quiet speech and CO speech.
Table 1. Mean (SD) F0 range (Hz), F0 mean (Hz), 1–3 kHz energy, speaking rate (syllables per second), pause rate (number of pauses), pause duration (s), and sound pressure level (dB SPL) for conversational (conv) and clear speech produced in noise and in quiet for children, young adults, and older adults.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Talker</th>
<th>Style</th>
<th>Noise</th>
<th>F0 Range</th>
<th>F0 Mean</th>
<th>1–3 kHz Energy</th>
<th>Speaking rate</th>
<th>Pause rate</th>
<th>Pause duration</th>
<th>SPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
<td>Conv</td>
<td>Quiet</td>
<td>204.08 (56)</td>
<td>201.90 (41)</td>
<td>21.86 (2.39)</td>
<td>4.56 (0.47)</td>
<td>17.2 (8.48)</td>
<td>0.11 (0.04)</td>
<td>68.03 (2.51)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>202.76 (59)</td>
<td>207.96 (29)</td>
<td>25.71 (5.01)</td>
<td>3.73 (0.87)</td>
<td>35.1 (29.47)</td>
<td>0.13 (0.06)</td>
<td>70.42 (4.98)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>210.83 (52)</td>
<td>249.01 (50)</td>
<td>28.33 (2.51)</td>
<td>4.06 (0.52)</td>
<td>16.7 (9.35)</td>
<td>0.16 (0.17)</td>
<td>70.39 (2.61)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>221.98 (61)</td>
<td>256.28 (49)</td>
<td>28.02 (2.57)</td>
<td>3.56 (1.03)</td>
<td>31.1 (31.45)</td>
<td>0.14 (0.08)</td>
<td>69.67 (2.48)</td>
<td></td>
</tr>
<tr>
<td>Young adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>186.81 (60)</td>
<td>146.22 (36)</td>
<td>21.61 (2.25)</td>
<td>4.71 (0.44)</td>
<td>7.3 (10.31)</td>
<td>0.12 (0.07)</td>
<td>66.50 (2.54)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>217.83 (53)</td>
<td>184.96 (37)</td>
<td>23.47 (2.46)</td>
<td>3.41 (0.55)</td>
<td>61.6 (49.06)</td>
<td>0.14 (0.03)</td>
<td>67.62 (2.50)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>194.91 (62)</td>
<td>174.37 (41)</td>
<td>27.27 (1.69)</td>
<td>4.04 (0.60)</td>
<td>12.0 (22.52)</td>
<td>0.13 (0.09)</td>
<td>70.39 (2.42)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>229.34 (69)</td>
<td>175.74 (40)</td>
<td>27.36 (2.23)</td>
<td>3.03 (0.59)</td>
<td>81.6 (70.68)</td>
<td>0.14 (0.04)</td>
<td>69.86 (2.20)</td>
<td></td>
</tr>
<tr>
<td>Older adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>216.38 (61)</td>
<td>157.01 (40)</td>
<td>21.52 (3.26)</td>
<td>4.07 (0.37)</td>
<td>11.1 (7.22)</td>
<td>0.12 (0.04)</td>
<td>68.16 (2.39)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>229.09 (68)</td>
<td>160.12 (39)</td>
<td>21.91 (3.10)</td>
<td>2.86 (0.41)</td>
<td>85.2 (49.14)</td>
<td>0.16 (0.04)</td>
<td>68.20 (2.00)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>248.28 (65)</td>
<td>179.95 (47)</td>
<td>24.75 (3.75)</td>
<td>3.60 (0.46)</td>
<td>16.9 (17.65)</td>
<td>0.12 (0.06)</td>
<td>70.32 (3.12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>278.08 (62)</td>
<td>177.71 (42)</td>
<td>24.56 (2.83)</td>
<td>2.63 (0.50)</td>
<td>101.7 (54.37)</td>
<td>0.18 (0.05)</td>
<td>69.94 (2.76)</td>
<td></td>
</tr>
</tbody>
</table>

respectively. There was a significant two-way interaction between Noise and Style, $F(1, 7041.042) = 5.060$, $p = .025$, revealing a wider F0 range for CS compared with CO in noise, but not in quiet. Furthermore, F0 range was significantly wider when producing quiet-to-noise adjustments for CS, but not for CO.

With regard to speaking rate, there was a significant main effect of Talker Age Group, $F(2, 27.004) = 8.526$, $p = .001$, with younger talkers overall speaking faster than the older adults. Significant main effects of Noise, $F(1, 53.981) = 26.801$, $p < .001$, and Style, $F(1, 53.981) = 135.177$, $p < .001$, revealed that both NAS and CS were significantly slower compared with speech produced in quiet and CO, respectively. Only one significant interaction, between Noise and Style, was found, $F(1, 7008.697) = 416.141$, $p < .001$. While the speaking rate decreased in both NAS and CS, the quiet-to-noise decrease was larger for CO compared with CS. Similarly, the conversational-to-clear speaking rate decrease was larger for speech produced in quiet than in noise. Style marginally interacted with Talker Age Group, $F(2, 53.981) = 3.048$, $p = .056$. When producing CO, older adults were significantly slower than both young adults and children ($p < .009$), but young adults and children did not differ. In CS, older adults were still significantly slower than both young adults and children ($p < .001$), and young adults were also marginally slower than children ($p = .035$). This revealed that children slowed down less than adults when producing CS.

Pause analyses showed that children produced the fewest pauses, followed by young adults and older adults: 25, 41, and 54 pauses, respectively. More pauses were produced in CS (66) compared with CO (13). An increase in the number of pauses was smaller in response to noise compared with quiet (36 vs. 43). Analysis of pause durations showed that there were no significant main effects of Talker Age Group, $F(2, 27.016) = 1.449$, $p = .252$, or Noise, $F(1, 54.064) = 1.122$, $p = .294$. There was a significant main effect of Style, $F(1, 54.064) = 27.923$, $p < .001$, with pause durations being significantly longer in CS compared with CO. There was a significant two-way interaction between Style and Talker Age Group, $F(2, 7000.855) = 4.321$, $p = .018$, with older adults producing significantly longer pauses than younger talkers in CS. Style also interacted with Noise, $F(1, 7000.855) = 17.408$, $p < .001$. The conversational-to-clear speech modifications were larger in NAS than in quiet. A three-way interaction between Talker Age Group, Noise, and Style, $F(2, 7000.839) = 6.502$, $p = .002$, revealed that older adults, but not children and young adults, produced longer pauses for CS in noise compared with CS in quiet. In CO and for the younger talker groups in general, there were no significant quiet-to-noise increases in pause duration.

Energy in the 1–3 kHz range was significantly affected by Talker Age Group, $F(2, 27) = 5.444$, $p = .010$, and Noise, $F(1, 53.997) = 58.312$, $p < .001$, but not Style, $F(1, 53.997) = 3.205$, $p = .079$. Children (young adults at $p = .051$) spoke with significantly more energy in the 1–3 kHz region compared with older adults. Energy of 1–3 kHz increased significantly in NAS compared with speech produced in quiet. A significant two-way interaction between Noise and Style, $F(1, 7041.002) = 416.141$, $p < .001$, revealed that the conversational-to-clear speech increase in 1–3 kHz energy was significant only for speech produced in quiet. A significant three-way interaction between Talker Age Group, Noise, and Style, $F(2, 7041.002) = 9.311$, $p < .001$, indicated that children and young adults, but not older adults, significantly increased energy in the 1–3 kHz range for conversational-to-clear speech modifications in quiet. None of the talker groups made significant energy increases in the 1–3 kHz range for conversational-to-clear speech modifications when speaking in noise. Additionally, children exhibited significantly more 1–3 kHz energy than older adults in every speaking style except CO produced in quiet.

Finally, Noise, $F(1, 53.825) = 15.238$, $p < .001$, and Style, $F(1, 53.825) = 3.626$, $p < .001$, but not Talker Age Group, $F(2, 27) = 1.132$, $p = .089$, significantly affected SPL. SPL was higher in both NAS and CS compared with the baseline quiet and CO speech, respectively. A significant two-way interaction between Talker Age Group and Noise, $F(2, 7119) = 13.925$, $p < .001$, showed that children had significantly higher SPL than young adults when producing speech in quiet. A significant two-way interaction between Noise and Style, $F(2, 7119) = 7.971$, $p < .001$, respectively.
indicated that SPL was higher in CS compared with CO in quiet, but not in noise. There was a significant three-way interaction of Talker Age Group, Noise, and Style, F(2, 7119) = 4.934, p < .001. Pairwise comparisons revealed that both children and young adults had significantly higher SPL when producing CS compared with CO in quiet. Children furthermore had a significantly higher SPL than young and older adults when producing CS in quiet. In noise, children and young adults had significantly lower SPL in CS compared with CO. Older adults had significantly higher SPL than young adults in CO quiet. There were no significant SPL differences between CS and CO in quiet or in noise for older adults.

**Segmental Measurements**

Table 2 lists mean vowel duration, and Table 3 lists F1 and F2 for each vowel category and all speaking styles (male and female talkers’ formant values are provided separately although the analyses collapse across gender). Figure 1 shows the average F1 × F2 data in CO and CS produced in quiet and in noise for female and male talkers separately in each age group. VSA can be inferred from the lines connecting vowel coordinates.

Analyses of vowel duration revealed a significant main effect of Talker Age Group, F(2, 27) = 4.305, p = .024. In line with the overall slower speaking rate, older adults produced significantly longer vowels than both younger groups. There were significant main effects of both Noise, F(1, 54) = 93.220, p < .001, and Style, F(1, 54) = 61.175, p < .001. Vowels were significantly lengthened in NAS and CS relative to speech in quiet and CO speech, respectively. Two interactions approached significance: Noise and Talker Age Group, F(2, 54) = 3.028, p = .057, and Noise and Style, F(1, 860) = 3.681, p = .051. While older adults produced significantly longer vowels than the younger groups in quiet, they were only marginally slower than young adults in noise (p = .056). That is, younger talker groups made larger vowel duration increases when producing NAS compared with older adults. Noise × Style interaction revealed that the quiet-to-noise lengthening was larger for CO than for CS. Additionally, the conversational-to-clear speech lengthening was larger for speech produced in quiet than for NAS.

As expected, F1 was significantly affected by Vowel, F(3, 4) = 171.994, p < .001, with low /æ/, /i, u/ vowels systematically differing in F1 frequency. There was also a significant main effect of Talker Age Group on F1, F(2, 27) = 9.458, p = .001. Consistent with smaller vocal tracts, children had higher F1 frequencies than adults. In terms of speaking styles, there were significant main effects of both Noise, F(1, 54) = 66.935, p < .001, and Style, F(1, 54) = 6.552, p = .013. F1 was significantly higher in NAS compared with speech produced in quiet, and in CS compared with CO. There was a significant interaction between Noise and Talker Age Group, F(2, 54) = 8.989, p < .001. All talker groups made significant quiet-to-noise increases in F1, but children raised F1 more than either adult group. Each of the three main factors interacted with Vowel Type: Vowel × Talker Age Group, F(6, 833) = 23.605, p < .001, Vowel × Noise, F(3, 833) = 5.740, p = .001, and Vowel × Style, F(3, 833) = 5.305, p = .001. These interactions revealed that young and older adult talkers had similar F1 frequencies for all vowels except /æ/ for which young adults had a significantly higher F1 (i.e., they produced a lower, more open /æ/). Children had different F1 frequencies from the adults for all vowels. While F1 was raised for all vowels in noise, F1 increase was larger for the low vowels /a, a/ than for the high vowels /i, u/. In CS, F1 for the low vowels /a, a/ but not the high vowels /i, u/ was significantly increased.

F2 was also significantly affected by Vowel, F(3, 4) = 26.231, p = .004. As expected, front /i, æ/ and back /u, a/ vowels systematically differed in F2. There was a significant main effect of Talker Age Group, F(2, 27) = 3.938, p = .032, with children showing significantly higher F2 values than older adults and marginally higher F2 values (p = .067) than young adults. There was a significant main effect of Noise, F(1, 54) = 15.827, p < .001, with higher F2 in NAS compared with quiet. There was no significant main effect of Style, F(1, 54) = 2.165, p = .147. The main effect of Talker Age Group significantly interacted with Vowel Type, F(6, 833) = 6.111, p < .001. Children produced a significantly higher F2 than older adults for /i/. All talker groups had significantly different F2 frequencies for /u/, with children having the highest and older adults the lowest F2 frequency. For /æ/, children produced a significantly higher F2 than both adult talker groups. There were no differences in F2 for /æ/ across talker groups. Analyses of the VSA revealed a significant effect of both Style, F(2, 27) = 12.071, p < .01, and Noise, F(2, 27) = 23.88, p < .001, but not Talker Age Group, F(2, 27) = 0.136, p = .873. VSA was significantly enlarged in CS and in NAS. None of the interactions were significant.

**Voice Measurements**

Table 4 lists mean HNR, jitter, and shimmer in CO and CS produced in quiet and in noise for each talker group. Voice analyses revealed that the effect of Talker Age Group on HNR approached significance, F(2, 27) = 4.929,
Table 3. Mean F1 and F2 (Hz) in conversational (conv) and clear speech produced in quiet and in noise for female and male talkers in each age group.

<table>
<thead>
<tr>
<th>Talkers</th>
<th>Style</th>
<th>Noise</th>
<th>F1</th>
<th>F2</th>
<th>F1</th>
<th>F2</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>i</td>
<td>i</td>
<td>u</td>
<td>u</td>
<td>æ</td>
<td>æ</td>
</tr>
<tr>
<td><strong>Female talkers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>Conv</td>
<td>Quiet</td>
<td>392.81</td>
<td>2478.40</td>
<td>436.20</td>
<td>1932.33</td>
<td>945.83</td>
<td>1763.49</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>378.51</td>
<td>2344.58</td>
<td>468.99</td>
<td>1950.79</td>
<td>1029.02</td>
<td>1930.56</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>478.41</td>
<td>2649.26</td>
<td>556.62</td>
<td>2026.12</td>
<td>1075.26</td>
<td>1873.75</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>483.63</td>
<td>2652.03</td>
<td>532.03</td>
<td>1896.47</td>
<td>1109.06</td>
<td>1939.75</td>
</tr>
<tr>
<td>Young adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>334.95</td>
<td>2486.16</td>
<td>389.73</td>
<td>1760.29</td>
<td>883.61</td>
<td>1843.54</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>350.51</td>
<td>2660.91</td>
<td>388.98</td>
<td>1783.08</td>
<td>926.82</td>
<td>1814.52</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>358.14</td>
<td>2658.67</td>
<td>398.93</td>
<td>1800.67</td>
<td>929.68</td>
<td>1835.12</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>346.55</td>
<td>2790.17</td>
<td>398.89</td>
<td>1713.01</td>
<td>980.78</td>
<td>1853.65</td>
</tr>
<tr>
<td>Older adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>358.72</td>
<td>2505.33</td>
<td>374.16</td>
<td>1777.28</td>
<td>783.14</td>
<td>1908.13</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>337.06</td>
<td>2652.98</td>
<td>394.34</td>
<td>1682.59</td>
<td>805.86</td>
<td>1955.49</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>377.90</td>
<td>2520.29</td>
<td>406.13</td>
<td>1759.61</td>
<td>822.28</td>
<td>2032.89</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>369.42</td>
<td>2708.21</td>
<td>398.87</td>
<td>1779.71</td>
<td>837.20</td>
<td>2004.81</td>
</tr>
<tr>
<td><strong>Male talkers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>Conv</td>
<td>Quiet</td>
<td>385.61</td>
<td>2590.98</td>
<td>389.71</td>
<td>1736.87</td>
<td>841.07</td>
<td>1782.77</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>392.91</td>
<td>2773.21</td>
<td>413.47</td>
<td>1783.16</td>
<td>853.02</td>
<td>1861.21</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>393.26</td>
<td>2732.32</td>
<td>414.64</td>
<td>1782.82</td>
<td>916.14</td>
<td>1946.13</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>422.07</td>
<td>2729.21</td>
<td>426.31</td>
<td>1835.56</td>
<td>985.55</td>
<td>1973.29</td>
</tr>
<tr>
<td>Young adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>279.16</td>
<td>2360.34</td>
<td>358.11</td>
<td>1710.93</td>
<td>688.36</td>
<td>1713.04</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>284.68</td>
<td>2439.41</td>
<td>360.44</td>
<td>1665.45</td>
<td>721.80</td>
<td>1716.66</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>311.99</td>
<td>2398.40</td>
<td>369.80</td>
<td>1689.47</td>
<td>737.12</td>
<td>1713.93</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>306.27</td>
<td>2487.34</td>
<td>360.19</td>
<td>1610.29</td>
<td>779.16</td>
<td>1790.57</td>
</tr>
<tr>
<td>Older adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>341.33</td>
<td>2200.19</td>
<td>329.24</td>
<td>1367.42</td>
<td>649.09</td>
<td>1636.17</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>383.17</td>
<td>2243.47</td>
<td>322.93</td>
<td>1346.19</td>
<td>683.73</td>
<td>1683.41</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>304.85</td>
<td>2224.05</td>
<td>359.56</td>
<td>1391.62</td>
<td>705.10</td>
<td>1714.04</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>317.52</td>
<td>2304.06</td>
<td>362.85</td>
<td>1413.20</td>
<td>717.17</td>
<td>1719.71</td>
</tr>
</tbody>
</table>

p = .058. Pairwise comparisons revealed that children had a higher HNR than the young adult (p < .002) and older adult (p < .003) talkers. There was also significant main effect of Noise, F(1,200) = 2.373, p = .0219, and Style, F(1, 200) = 2.348, p = .0218. HNR was significantly increased in NAS compared with quiet speech, and in CS compared with CO. A significant interaction between Style and Talker Age Group, F(2, 200) = 4.788, p < .001, revealed that children, but not adults, increased HNR through the conversational-to-clear speech modifications. As a result, the child–adult difference in HNR was significant in CS but not in CO. Jitter was significantly affected by Talker Age Group, F(2, 27) = 4.095, p = .049. Children overall had lower jitter than both the adult groups. There was a significant main effect of Noise, F(1, 200) = 2.438, p = .0174, with a significant jitter decrease in NAS compared with quiet speech. Style had no significant effect on jitter, F(1, 200) = 1.104, p = .275. Talker Age Group interacted with Style, F(2, 200) = 2.439, p = .0217, indicating that while children decreased jitter in conversational-to-clear speech modifications, young adults increased it. Finally, results revealed a significant main effect of Talker Age Group on shimmer, F(2, 27) = 7.397, p < .001. Children overall had significantly lower shimmer than the adult talker groups. Noise significantly affected shimmer, F(1, 200) = 2.721, p = .009, with shimmer being lower in NAS compared with quiet speech. There was no significant main effect of Style, F(1, 200) = 0.408, p = .686. A significant two-way interaction between Style and Talker Age Group, F(2, 200) = 2.210, p = .039, revealed that children, but not adults, had significantly lower shimmer in CS compared with CO.

**Discussion**

**NAS and CS**

A direct comparison of CS and NAS revealed that the two adaptations were similar but also differed in acoustic–articulatory modifications. Different features targeted by each adaptation may be attributed to somewhat different emphasis on augmenting signal loudness versus enhancing articulatory precision to help the listener access and decode the linguistic message (Cooke et al., 2013; Lansford et al., 2011). The enhancement of the global salience of the speech signal was found in the present study through pitch and energy modifications, namely, an increase in F0 mean, 1–3 kHz energy, and SPL, when speaking in noise. These modifications presumably augment auditory sensitivity of the speech signal, thus making it stand out more from the surrounding noise and facilitating its transmission (Godoy et al., 2014). Additionally, increases in F0 and

Note that all sentences had equalized SPL so that we can control the signal-to-noise ratio for intelligibility testing (Smiljanic & Gilbert, 2017). The results showed significant intelligibility benefit for CS and NAS without the loudness cue playing a role, suggesting that the other acoustic–articulatory modifications significantly contributed to the improved speech understanding in both speech adaptations.
the energy shift to higher frequencies found predominantly in NAS could reflect specific strategies on the part of the talkers to make their speech more audible in the presence of the SSN used to elicit NAS (Cooke & Lu, 2010; Lu & Cooke, 2008). Energy increase in the 1–3 kHz range and in SLP was found in CS as well, suggesting that speaking clearly also promotes audibility in a similar way to NAS. This is in contrast to Godoy et al.’s (2014) study, which reported no spectral energy boost in CS compared with loud speech. This difference may be due to a number of reasons, including the elicitation instructions, type of barrier used to elicit the adaptations, and the difference in the materials. Vowel lengthening found in both NAS and CS also points to shared strategy possibly contributing to the enhanced audibility by making the voiced sounds more robust to energetic masking (Junqua, 1993).

Both NAS and CS modifications included a decrease in speaking rate. In addition to enhancing global salience of the speech signal, slowing down may serve to decrease cognitive effort on the part of the listener via a reduction

---

**Figure 1.** Average $F_1 \times F_2$ for /i, u, æ, ɑ/ for children (top), young adults (middle), and older adults (bottom). Data for female talkers are shown in the left panels and those for the male talkers in the right panels. Solid dark lines connect vowels produced in conversational quiet speech (COQ). Dashed dark lines connect vowels produced in clear quiet speech (CSQ). Solid light lines connect vowels produced in conversational noise-adapted speech (CONAS). Dashed light lines connect vowels produced in clear noise-adapted speech (CSNAS).
in rate at which information is transmitted or through better instantiated phonetic categories (Cooke et al., 2013; Lansford et al., 2011). When speech signal is degraded, listeners are required to apply more cognitive resources to recognize words, drawing on higher order decision making (Davis & Johnsrude, 2007). Van Engen, Chandrasekaran, teners are required to apply more cognitive resources to terrestrial phonetic categories (Cooke et al., 2013; in rate at which information is transmitted or through bet-
ed more processing resources for encoding speech content in memory (cf., effortfulness hypothesis, McCoy et al., 2005 and ease of language understanding, Ronnberg, 2003; Ronnberg, Rudner, Foo, & Lunner, 2008). Increased frequency and duration of pauses may further afford more processing time to the listener, thereby decreasing cognitive effort. Interestingly, this modification was largely found for conversational-to-clear adjustments. Segment lengthening and enhanced intonation cues associated with pausing could aid listeners in lexical segmentation, thus further contributing to the ease of processing (Mattys, White, & Melhorn, 2005). The absence of significant pausing in NAS suggests that this feature may reflect a more deliberate adaptation when speaking clearly, rather than when speaking in noise.

Talkers also increased formant frequencies in both NAS and CS. However, a distinction in the strategies between the two types of adjustments appeared. The interactions with Vowel Type revealed that quiet-to-noise modifications increased F1 for both low and high vowels, albeit more so for the low vowels. Conversational-to-clear adaptation, on the other hand, involved only increasing F1 for low vowels. F2 was increased for all vowels in response to noise but not when speaking clearly. Despite these different strategies, VSA was significantly increased both in NAS and CS (58082 Hz2 in NAS and 56541 Hz2 in CS), demonstrating that the distance between the contrastive vowel categories was increased via both adaptations. This finding is in contrast with previous work which typically finds the VSA expansion for CS and less consistently for NAS (Davis & Kim, 2012; Ferguson & Kewley-Peort, 2002, 2007; Godoy et al., 2014; Lu & Cooke, 2008; Summers et al., 1988). These changes reflect greater articulatory precision in CS aimed at addressing a perceptual difficulty on the part of the listener (e.g., nonnative listener or listener with hearing impairment). The results here suggest that talkers can also increase articulatory precision when speaking in response to noise presumably also to aid the listener in decoding the language-specific sound categories. This specific acoustic-articulatory adjustment is thus not limited by environmental or physiological constraints to only overcoming the interlocutor-related barrier. This further illustrates that NAS, as elicited in the current study and in a direct comparison with CS, is a listener-oriented speaking style adaptation beyond an automatic response to the noise in the environment. This is important because greater VSA, although not necessary (cf., Krause & Braida, 2004), has been linked to enhanced intelligibility (Ferguson & Kewley-Peort, 2002; Hazan & Markham, 2004; Picheny et al., 1986).

Finally, the current study found significant changes of voice quality measures in response to speaking in noise (increased HNR, decreased jitter and shimmer) and to a lesser degree in CS (increased HNR). On all measures, voice quality seems to show more regularity and less instability of vocal fold vibration in the speaking style adaptations compared with the baseline speech. Increased HNR in NAS and CS can be linked to an increase in vocal intensity, as seen in increased SPL and 1–3 kHz energy in both modifications. Additional changes in shimmer and jitter in NAS were likely due to the larger degree of increase in vocal effort required to compensate for talking over noise. The current findings are in line with previous work showing increased HNR and reduced jitter and shimmer when children were instructed to produce sustained vowels loudly (Broekmann-Bauser, Beyer, & Bohlender, 2015; Glaze, Bless, & Susser, 1990). We show here that similar changes in voice characteristics are found when talkers are producing speech in response to the noise in the environment and when speaking clearly, thus modifying their speech more globally, that is, beyond isolated vowels. While these voice changes result from the physiological manipulations of the larynx and respiratory system when speaking loudly and are thus not deliberate speaking style adaptations, examining any

Table 4. Mean (SD) harmonics-to-noise ratio (HNR; dB), jitter percent, and shimmer percent in conversational (conv) and clear speech produced in noise and in quiet for each talker group.

<table>
<thead>
<tr>
<th>Talkers</th>
<th>Style</th>
<th>Noise</th>
<th>HNR</th>
<th>Jitter</th>
<th>Shimmer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
<td>Conv</td>
<td>Quiet</td>
<td>10.00 (4.41)</td>
<td>0.027 (0.02)</td>
<td>0.094 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>12.42 (4.06)</td>
<td>0.019 (0.01)</td>
<td>0.066 (0.03)</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>11.47 (4.72)</td>
<td>0.019 (0.02)</td>
<td>0.065 (0.02)</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>14.15 (4.18)</td>
<td>0.013 (0.01)</td>
<td>0.054 (0.02)</td>
</tr>
<tr>
<td>Young adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>10.54 (4.59)</td>
<td>0.023 (0.02)</td>
<td>0.091 (0.05)</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>8.32 (3.54)</td>
<td>0.034 (0.02)</td>
<td>0.079 (0.02)</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>11.92 (2.73)</td>
<td>0.020 (0.01)</td>
<td>0.063 (0.02)</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>9.76 (2.77)</td>
<td>0.024 (0.01)</td>
<td>0.077 (0.02)</td>
</tr>
<tr>
<td>Older adults</td>
<td>Conv</td>
<td>Quiet</td>
<td>9.26 (2.56)</td>
<td>0.032 (0.03)</td>
<td>0.088 (0.03)</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Quiet</td>
<td>9.08 (3.31)</td>
<td>0.025 (0.01)</td>
<td>0.085 (0.07)</td>
</tr>
<tr>
<td></td>
<td>Conv</td>
<td>Noise</td>
<td>10.21 (3.18)</td>
<td>0.024 (0.02)</td>
<td>0.079 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Noise</td>
<td>9.86 (3.17)</td>
<td>0.028 (0.02)</td>
<td>0.081 (0.03)</td>
</tr>
</tbody>
</table>
differences that may occur in various communicative situations is important for our understanding of normal and disordered voice characteristics. Future work on voice characteristics in speaking style adaptations should include cepstral- and spectral-based measures that have been shown to be reliable in analyzing continuous speech and exhibited higher correlation with auditory–perceptual judgments (Barsties & De Bodt, 2015; Lowell & Hylkema, 2016; Lowell, Colton, Kelley, & Mizia, 2013).

Age Effect on Speaking Style Adaptations

Older adult speech was characterized by less energy in the 1–3 kHz range and a slower speaking rate (including longer vowels) compared with that of the younger talkers. This is in line with previous research showing production differences between older and young adults (Brückl & Sendimeier, 2003; Linville & Rens, 2001; Shipp, Qi, Huntley, & Hollien, 1992; Winkler, Brückl, & Sendimeier, 2003; Xue & Hao, 2003). These differences can be attributed to the age-related anatomical changes, slower cognitive processing (Craig & Byrd, 1982; Salthouse, 1996), and neuromuscular changes affecting the rate and precision of articulator movement (Hartman & Danhauer, 1976). Older adults also had overall higher SPL in quiet CO speech compared with young adults, which is in contrast to previous work that either reported no age-related differences (Prakup, 2012) or found older male talkers to have lower SPL than young male talkers (Goy, Fernandes, et al., 2013; Hodge, Colton, & Kelly, 2001). This, along with the lack of significant younger–older adult differences on F0, F1, F2 and voice measurements found here, could be attributed to the good overall physiological condition of the older adult talkers in this study. Research has shown that a speech production mechanism is shaped by both chronological age and physiological differences (physical condition) among individuals (Ferrand, 2000; Ramig & Ringel, 1983; Ringel & Chodzko-Zajko, 1987). All of the older participants were active enough to volunteer for on-campus research, and several commented on their regular physical activities. Combined, the similarities and differences between young and older adult talkers suggest that not all components of the speech production mechanism are equally affected by age-related changes. In the case of these older adults, articulatory precision (similar F1 and F2 with young adults) and vocal intensity (higher SPL) could be maintained even though some other age-related physiological changes (less energy in 1–3 kHz range) may occur. How these various factors interact to determine speech production in individual talkers across the life span remains to be determined.

Voice analyses revealed significant differences between children and adult talkers, but not between young and older adults. Children had overall higher HNR and lower jitter and shimmer than the adult talkers. Previous work has shown a lower HNR in children 4–10 years old compared with adults arising from the anatomical differences in vocal fold structure (Ferrand, 2000) and no differences in jitter for children 7–15 years old (Linders, Massa, Boersma, & Dejonckere, 1995). In comparison with older adult talkers, children in this study had higher HNR, demonstrating that children’s voices at 11–13 years of age have less additive noise in the voice signal. Higher HNR could also be related to lower jitter and shimmer found here as the measures are interrelated (Hillenbrand, 1987). The marginally higher HNR and lower jitter and shimmer found in children compared with adults could reflect an overshoot in these acoustic parameters (similar to Lee et al., 1999) as children are tuning their productions to become adultlike. Additionally, these differences could arise from children producing more careful and hyper-articulated speech while reading compared with the adult talkers.

The three groups diverged in some of their responses in noise and when speaking clearly. Children slowed down less in CS than the adults. This is in contrast with the findings of Pettinato and Hazan (2013), who found similar word duration lengthening for children 13–14 years old and young adults in response to vocoded speech. Pettinato and Hazan (2013) also found, unlike in the current study, that children were overall slower than young adults. These differences could be a result of the task used in the two studies. In their study, children participated in the spot-the-difference picture task, which may require more cognitive effort to complete, thus slowing down their speech, whereas here, they read sentences from the computer monitor. Furthermore, vocoded speech may elicit a different response compared with the noise in the environment or when speaking clearly (Hazan & Baker, 2011; Lam et al., 2012). In the current study, children also produced fewest pauses overall. Compared with the young adults, children produced 27 and 50 fewer pauses for CS in quiet and in noise, respectively. When they did pause, children produced pauses of similar length to the adults. The differences in speaking rate and pause information between children and young adults could arise from the children’s inability to deliberately change their speech in the absence of the obvious communicative goal, such as the presence of a listener with perceptual difficulty (Junqua, Finkke, & Field, 1999; Summers et al., 1988). However, children, as well as young adults, increased SPL in CS relative to CO in quiet. This adultlike response showed that children did modify their speech in response to the instructions to speak clearly. It is thus possible that children may not have developed a full adultlike spectrum of CS strategies by the age of 11–13 years.

While older adults had the overall slowest speaking rate, young and older adults in the current study made similar decreases to speaking rate when producing CS (1.16 and 1.09 syll/s, respectively). Given their overall slower baseline and similar decrease, the resulting CS speaking rate was still significantly slower for older adults compared with the other two talker groups. This is in contrast to Smiljanic (2013) in which the older adults’ speaking rate decrease was smaller than that of young adults, resulting in a similar CS speaking rate. The difference in the speaking style rates between the two studies could reflect the materials used in the two studies. In the current study, talkers produced simple meaningful sentences, whereas in Smiljanic
(2013), they produced semantically anomalous sentences that may result in different speech patterns. The current results suggest that older adults can slow down by the same decrement as young adults when producing intelligibility-enhancing CS. Older adults also produced the most pauses overall (54 for older adults, 41 for young adults, and 25 for children). Compared with young adults, they produced more pauses for CS in quiet and for CS in noise as well as the longest pauses when speaking clearly in noise. While older adults produced significantly longer vowels than children and young adults in quiet, this difference was diminished in noise where older adults had only marginally longer vowels than young adults. This shows that older adults lengthened their vowels the least in response to noise (35 ms compared to 42 and 63 ms for young adults and children, respectively). The pausing patterns could reflect a deliberate strategy on the part of the older adults to enhance saliency of the speech signal and decrease cognitive effort for the listener by allowing them more time to process the incoming information (Cooke et al., 2013). The increased pausing, in addition to the overall slower speech, could alternatively reflect a speech planning and production difficulty in a cognitively demanding communicative situation (speaking clearly while, at the same time, trying to overcome the environmental noise). The precise source of this difference between young and older adults remains to be determined.

The VSA increase when speaking in noise (noise minus quiet) was similar for children and young adults (73862 Hz² vs. 73930 Hz²). Children increased the VSA less in CS (clear minus conversational) compared with the young adults (42678 Hz² vs. 96363 Hz²), although the difference was not significant. The smaller adjustment in VSA found in CS mirrors differences in VSA expansion found in Pettinato and Hazan (2013) and Pettinato et al. (2016) for children 9–14 years old and young adults in response to vocoded speech. The fact that children could increase VSA in noise suggests that the somewhat smaller VSA expansion in CS is not the result of the children’s inability to produce hyper-articulated vowels. Rather, it suggests that children do not have fully adultlike CS strategies and, furthermore, that speaking in noise seems to induce a more automatic repertoire of responses resulting from increased vocal effort. Similar to Pettinato et al. (2016), children in the current study had a slightly larger overall VSA than young adults in their baseline CO speech (271414 Hz² vs. 231831 Hz²), which could indicate, as Pettinato and colleagues argue, that children also need to learn how to produce hypo-articulated variants when appropriate. Older adults made smaller VSA increases in both noise (26453 Hz²) and CS (30583 Hz²) compared with young adults, albeit these differences were not statistically significant. This suggests that older adults were overall successful in modifying their vocal productions although the resulting vowel categories were less dispersed compared with younger talkers. Interestingly, older adults also had a somewhat larger baseline VSA than young adults (268510 Hz² vs. 231831 Hz²). It is unlikely that older adults do not know how to hypo-articulate vowel categories in CO speech. The difference in VSA between young and older adults could arise from their overall slower speaking rate, which may allow time for more extreme tongue movements.

Finally, children made significantly larger mean F0 increases in response to noise compared with older adult talkers. Older adults made the smallest mean F0 increase in quiet-to-noise modifications for CS. Children and young adults, but not older adults, increased energy in the 1–3 kHz range when producing conversational-to-clear speech adjustments in quiet. This absence of energy increase in CS may reflect age-related changes to the articulatory and respiratory control in older adults. The larger mean F0 modifications in children, on the other hand, could reflect an acoustic target overshoot (cf., Lee et al., 1999) as children are aligning their strategies when responding to noise to those of adults. Finally, voice analyses showed that children increased HNR and decreased shimmer and jitter when producing conversational-to-clear speech adjustments. This result is similar to Glaze et al. (1990), who found that children increased HNR and decreased shimmer when asked to speak loudly. In line with this, children here were also found to have higher SPL in CS compared with CO. Modifications that children implemented in CS led to a greater reduction in the amount of noise in their voice quality (harmonics, amplitude) compared with the adults whose modifications also included louder speech in NAS and CS.

**General Discussion**

The current study examined similarities and differences in acoustic–articulatory adjustments that talkers of three different age groups make in response to noise and when speaking clearly. The findings suggest that although there are overlapping NAS and CS modifications, each speaking style adaptation also targets a distinct set of acoustic–phonetic features. The common adaptation strategies could reflect the shared listener-oriented goals of promoting the audibility of the signal (e.g., F0 raising, boost in spectral energy, increase in SPL, longer voiced segments), enhancing linguistic structure (e.g., increased VSA), and decreasing cognitive effort (e.g., slower speaking rate) (Cooke et al., 2013; Lansford et al., 2011). On the other hand, the differences in acoustic–articulatory features characterizing the two adaptations (e.g., increased F0, jitter, and shimmer in NAS vs. increased pausing in CS) reflect the specific type of communicative barrier that the talkers are trying to overcome. In the NAS condition, talkers were producing speech in response to the actual noise which may induce a more automatic adaptation through increased vocal effort. In the CS condition, on the other hand, talkers were instructed to speak clearly, addressing a perceptual difficulty on the part of the listener with low proficiency, which may be a more intentional adaptation inducing, for instance, simplification of the prosodic structure. CS and NAS can thus be viewed as two varieties of hyper-articulated speech (H&H, Lindblom, 1990) with talkers being able to fine-tune their response to different communication barriers (Hazan & Baker, 2011; Lam et al., 2012).
The comparison of CS and NAS strategies across the three talker groups revealed a number of age-related differences. It is not straightforward to account for some of these differences. Auditory and cognitive changes and changes in the rate and precision of articulatory control and phonatory function associated with aging could underlie some of the observed differences between the young and older adults. Speaking style adaptation differences between children and adult talkers, on the other hand, may reflect the children’s on-going tuning of the acoustic–articulatory adjustments needed to produce hypo-articulated variants and for overcoming communicative barriers. This suggests that not all neurologically typical talkers have all of the hypo- and hyper-strategies available to them (Lindblom, 1990). Future work should further investigate the role that development and aging play in the acquisition versus loss of the ability to produce intelligibility-enhancing adaptations. A wider span of age groups should be included in this examination to more precisely determine when some of these differences occur.

An important question is how these CS and NAS modifications affect intelligibility. For these particular talkers, this question is addressed in the companion article (Smiljanic & Gilbert, 2017). More generally, speech-oriented, behavioral therapy techniques using rate reduction, increased vocal intensity, and CS (Darling & Huber, 2011; Duffy, 2013; Lam & Tjaden, 2016) aim to improve speech production and maximize speech intelligibility for talkers with dysarthria (Beukelman et al., 2002; Park et al., 2016; Tjaden, Sussman, et al., 2014). The findings here provide evidence that NAS and CS simultaneously affect articulatory and respiratory–phonatory behavior. As such, either method has the potential to induce intelligibility-enhancing adaptations. It is important to note though that a direct link between any one acoustic-articulatory modification and increased intelligibility remains tenuous (Cooke, Mayo, & Villegas, 2014; Godoy et al., 2014; Krause, 2001; Krause & Braida, 2004; Liu & Zeng, 2006; Picheny et al., 1986; Tjaden, Kain, et al., 2014; Uchanski et al., 1996). It is likely that a combination of a number of cue modifications that talkers make gives rise to the intelligibility benefit. To that end, using CS and NAS as global, speech-oriented techniques, rather than targeting individual speech parameters, may be an effective way of eliciting a number of modifications simultaneously (Yorkston, Hakel, Beukelman, & Fager, 2007). Future work needs to examine how exactly these enhancement strategies aid speech processing beyond word recognition in noise (Cooke et al., 2013; Lansford et al., 2011).

References


Downloaded From: http://jslhr.pubs.asha.org/ by a ReadCube User on 11/09/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx


