JSLHR

Article

Age-Related Changes to Spectral Voice Characteristics Affect Judgments of Prosodic, Segmental, and Talker Attributes for Child and Adult Speech

Laura C. Dilley,[a] Elizabeth A. Wieland,[a] Jessica L. Gamache,[a] J. Devin McAuley,[a] and Melissa A. Redford[b]

Purpose: As children mature, changes in voice spectral characteristics co-vary with changes in speech, language, and behavior. In this study, spectral characteristics were manipulated to alter the perceived ages of talkers' voices while leaving critical acoustic–prosodic correlates intact, to determine whether perceived age differences were associated with differences in judgments of prosodic, segmental, and talker attributes.

Method: Speech was modified by lowering formants and fundamental frequency, for 5-year-old children's utterances, or raising them, for adult caregivers' utterances. Next, participants differing in awareness of the manipulation (Experiment 1A) or amount of speech-language training (Experiment 1B) made judgments of prosodic, segmental, and talker attributes. Experiment 2 investigated the effects of spectral modification on intelligibility. Finally, in Experiment 3, trained analysts used formal prosody coding to assess prosodic characteristics of spectrally modified and unmodified speech.

Results: Differences in perceived age were associated with differences in ratings of speech rate, fluency, intelligibility, likeability, anxiety, cognitive impairment, and speech-language disorder/delay; effects of training and awareness of the manipulation on ratings were limited. There were no significant effects of the manipulation on intelligibility or formally coded prosody judgments.

Conclusion: Age-related voice characteristics can greatly affect judgments of speech and talker characteristics, raising cautionary notes for developmental research and clinical work.

Key Words: prosody, intelligibility, fluency, impairment, voice characteristics

Prosody is an important aspect of verbal communication that children must master to achieve adultlike speech-language competency. Prosodic cues such as pitch and timing are important for conveying a variety of information in spoken language, including semantic, lexical, syntactic, and emotional information (Lehiste, 1970; Scherer, 2003). Children's ability to effectively use prosody to communicate improves over time (Allen & Hawkins, 1980; Snow, 1995). Assessment of children's and adults' verbal communicative competency frequently involves characterizing prosodic attributes (e.g., speech rate, pausing, phrasing, and fluency) alongside speaker attributes (e.g., developmental age, language skill, and personality factors) in order to identify errors and develop treatment goals.

In this study, we investigated the perception of prosodic, segmental, and talker attributes, focusing on whether judgments about such attributes are dependent on a listener's knowledge or beliefs about a talker—in particular, the talker's probable age—and whether such dependence varies by assessment method. In addition to having clinical and research applicability, this investigation addressed the theoretical issue of whether prosodic (i.e., suprasegmental) judgments of speech are independent of other components of the speech signal, as predicted by theories of speech processing in which suprasegmental, segmental, and indexical components of the signal are treated as modular (Halle, 1985; Kuhl, 1991).

[a]Michigan State University, East Lansing
[b]University of Oregon, Eugene
Correspondence to Laura C. Dilley: ldilley@msu.edu

Editor: Jody Kreiman
Associate Editor: Ewa Jacewicz
Received July 23, 2011
Revision received March 26, 2012
Accepted June 9, 2012
DOI: 10.1044/1092-4388(2012/11-0199)

Accurate characterization of prosody is important for understanding speech-language development in children as well as for assessing speech-language disorders in both children and adults. Prosody is often affected in a variety of communication disorders (Peppe, 2009), and speech-language pathologists must judge prosodic attributes such as speech rate, pausing/phrasing, and fundamental frequency (F0) as part of diagnosing and treating speech-language disorders. Both in clinical and research settings, prosody judgments are often made impressionistically using ad hoc approaches, as opposed to using more formalized coding systems, rating procedures, or acoustic measurements.

Prior research suggests that impressionistic judgments of prosody (e.g., speech rate, F0) may be susceptible to listener biases. For example, Bond and colleagues (Bond, Simpson, & Feldstein, 1988) showed that listeners judged the same acoustic speech rate as faster or slower when presented at different levels of F0 or intensity. This suggests that perceived speech rate is not merely a function of acoustic rate but also depends on complex, interrelated acoustic factors, a proposition bolstered by similar findings in nonspeech auditory perception (Henry, McAuley, & Zaleha, 2009; Melara & Marks, 1990). If judgments of one attribute (e.g., speech rate) depend not only on the acoustics of that attribute (e.g., acoustic speech rate) but also on another (e.g., F0), then therapy might inadvertently target suboptimal characteristics. For example, if a talker's speech rate is perceived as too fast or too slow, then it may be more appropriate under some conditions to target F0 rather than articulation rate, per se, because of the interdependence of these acoustic attributes in perception (Bond et al., 1988; Henry et al., 2009; Melara & Marks, 1990).

Given that perceptual judgments of prosody may be subject to interdependencies of acoustic factors, a variable worthy of study is developmental age. Little research has examined the possibility of biases affecting assessment of children's speech; however, the presumed age and gender of a child have both been shown to influence perceptions of children's speech as well as ratings of accuracy and the quality of its production (Munson, Edwards, Schellinger, Beckman, & Meyer, 2010; Munson & Seppanen, 2009).

Understanding how prosodic competency changes over time is a long-term goal of our research. Given prior findings of perceptual interdependence among acoustic variables, impressionistic measures of prosody may rely on confounded acoustic variables that change as a child matures. If so, then changes attributed to prosodic development might reflect development in other areas. Thus, it is important to know if perception of prosodic attributes (e.g., perceived speech rate) is dependent merely on simple acoustic properties (e.g., acoustic speech rate) or a broader set of quasi-orthogonal acoustic characteristics.

Many acoustic characteristics change concurrently as children age, further underscoring the possibility of confounds between perception of prosodic and other acoustic attributes. For example, speech rate increases from ages 5 through 15 (Lee, Potamianos, & Narayanan, 1999). F0 tends to decrease with age, even in prepubescent children (Sussman & Sapienza, 1994). Likewise, formants tend to lower as children develop (Lee et al., 1999). Given findings that acoustic attributes underlying prosodic cues are perceptually interdependent (Bond et al., 1988), it is important to determine how perception of prosody may be affected by systematic changes in acoustic variables that are correlated with developmental age. Indeed, the converse relationship holds: For adults' speech, judgments of talker age are affected by manipulations to F0, F1, and/or speech rate (Harnsberger, Shrivastav, Brown, Rothman, & Hollien, 2008; Reubold, Harrington, & Kleber, 2010); however, no prior published work has examined the effect of F0 or formant manipulations on judgments of talker age for children's speech.

The critical manipulation in our experiments was a spectral manipulation of speech harmonics and formant frequencies that was expected to affect perceived talker age and, possibly, prosody judgments, if these are confounded with other perceptual characteristics of a speaker's voice. This spectral manipulation, which was expected to make children sound older and adults sound younger, critically permitted us to examine whether judgments of prosody would be affected by age-related voice characteristics while holding constant key acoustic–prosodic variables: acoustic speech rate; frequency, duration, and number of pauses; and the pattern of F0 ups and downs (i.e., the F0 contour), which conveys a variety of linguistic information (Lehiste, 1970). Therefore, the comparisons of greatest interest involved judgments of prosody for a talker's speech under different spectral manipulation conditions in which the talker was perceived as older or younger. In Experiment 1A, we compared such judgments for naïve participants and for participants who were told that the speech had been altered so that talkers may sound older or younger. Likewise, in Experiment 1B, we examined judgments made by speech-language pathology (SLP) master's students of speech, to determine whether individuals with relatively more SLP training would also be susceptible to age-related bias in prosody judgments.

By comparing judgments of different listeners, we could address questions regarding how awareness of potential bias and/or learning and experience might each affect judgments. The first question we addressed was whether age-related bias in prosody judgments is relatively automatic or, alternatively, whether age-related bias is a function of awareness. Prior research suggests that even when individuals are aware of biases (e.g., in racial attitudes), there can be long-lasting, subtle effects on processing (Dasgupta & Greenwald, 2001). Therefore, in Experiment 1A, we compared the judgments of listeners who were made aware of the spectral age manipulation with the judgments of listeners who were not informed of this manipulation. The second question we addressed was whether individuals with greater training and experience in speech development and assessment—here, SLP master's students—are susceptible to age-related bias in judgments of prosody and other characteristics. Less age-related bias might be expected for the SLP master's students than for the psychology student undergraduates, due to the former having substantially more experience in speech-language issues, which might permit more reliable assessment of prosody under different spectral profiles. On the other hand, speech perception is affected by sociodevelopmental characteristics of talkers, such as age and social class (Drager, 2011; Hay, Warren, & Drager, 2006; Johnson, Strand, & D'Imperio, 1999; Munson, Jefferson, & McDonald, 2006), suggesting that greater experience with different talker groups might affect speech perception. On this basis, we might expect *more* susceptibility to age-related bias in prosody judgments by SLP master's students than by undergraduate students in Experiment 1A.

Another question we addressed was how a talker's baseline degree of speech-language competency affects prosodic and other judgments. Talkers in our studies were either 5-year-old children or adults. Five-year-olds differ from adults in many speech-language characteristics, including phonology, syntax, and vocabulary (Bernthal, Bankson, & Flipsen, 2008). We predicted that baseline speech-language skill would affect prosody judgments, such that effects of our spectral "age" manipulation on such judgments would differently affect adults' versus children's speech. For example, we expected that, in the unmodified condition, children's speech rate should be judged as slower than that of adults because acoustic rate is slower for children than for adults (e.g., Lee et al., 1999). However, age-related bias about speech rate should trend in the direction of the expected *acoustic* speech rate, given the perceived age of the talker. Thus, children's speech should be judged as slower in the spectrally modified (i.e., older) condition than in the unmodified condition because a child's speech rate is age appropriate for a child's voice but too slow when paired with an adult's voice; the reverse should be true for adults' speech.

The talkers selected for this study were adults and 5-year-old children because of these groups' substantial differences in fluency and articulation (Bernthal et al., 2008; Redford, 2012). Fluency is strongly related to prosody, including duration, rate, pausing, and pitch, but it is not independent of segmental pronunciation (Bloodstein, 1995). Articulation is chiefly a segmental characteristic and greatly influences intelligibility. An advantage of our spectral manipulation method was its good preservation of segmental information—for example, distinctive feature contrasts for consonants and vowels. Thus, our manipulation permitted investigation of age-related bias in perception of the (quasi-prosodic) attribute of fluency as well as the (largely segmental) attribute of intelligibility. There is a dearth of research on potential age-related bias in impressionistic judgments of fluency and intelligibility. Such judgments are widely used in clinical practice (Bernthal et al., 2008; Bloodstein, 1995), even though they are notoriously unreliable and have been shown to have interdependencies with other perceptual variables (Bernthal et al., 2008; Southwood & Weismer, 1993). No research that we know of has investigated whether differences in judgments of intelligibility and fluency might be associated with differences in the perceived age of a talker.

We also predicted that bias about age-related speech-language performance expectations would be observed in judgments of typicality or impairment in talkers. For example, a relatively high disfluency or misarticulation rate would likely be interpreted as normal for a typically developing, nonstuttering, 5-year-old child. However, if the same (child) talker's voice is made to sound like that of an adult, then an identical disfluency and misarticulation rate might be deemed as indicating speech-language impairment (e.g., stuttering or a motor speech disorder), or perhaps cognitive impairment. A complementary spectral transformation for adults' speech was not predicted to have the same effect; because typical adults produce few disfluencies and misarticulations and have relatively high vocabulary and language skills, transforming the spectrum of (adult) speech to sound like that of a child should result in a talker's sounding extraordinarily competent (and unimpaired in speech, language, or cognition). Such a pattern of results would be consistent with a hypothesis that ratings of prosody and other speech attributes depend not only on isolated characteristics of a talker's speech but also on the listener's overall assessment of the talker's characteristics, including those related to probable impairment.

Finally, we assessed two additional properties related to communication impairment. The first was the degree of perceived anxiety of the talker, based on the observation that people who stutter tend to be judged as sounding more anxious than people who do not stutter (Menzies, Onslow, & Packman, 1999). We predicted that child talkers in the spectrally modified condition would be judged as sounding particularly anxious because of the likelihood that they will be judged as having a fluency-related disorder. Second, we investigated the likeability of a talker, based on findings showing that individuals with cognitive or communication impairments tend to be viewed more negatively than those without them (Franck, Jackson, Pimental, & Greenwood, 2003; Lass, Ruscello, Harkins, & Blankenship, 1993).

In summary, in this research we used a spectral manipulation to alter a talker's perceived age and to determine the subsequent effects on judgments of prosody and other speech and speaker characteristics. In Experiment 1, we examined judgments of prosodic and articulatory properties (speech rate, fluency, and intelligibility) and speaker characteristics (cognitive and/or communication impairment, anxiety, and likeability) as a function of spectral manipulation conditions by three groups of listeners: (a) naïve undergraduates versus (b) undergraduates who were explicitly told about the voice alteration (Experiment 1A), and (c) SLP master's students (Experiment 1B). Select comparisons among the groups permitted us to examine the extent to which such judgments are susceptible to age-related bias as a function of awareness of unreliability of voice age characteristics and/or listener experience. Experiment 2 was a control study aimed at determining the extent to which intelligibility was degraded due to the spectral manipulation, to determine the extent to which age-related bias affected impressionistic intelligibility judgments in Experiment 1. Finally, Experiment 3 involved use of formal prosody labeling to determine its comparative freedom from age-related bias relative to impressionistic judgments. We predicted that the more "analytical" method of formal prosody coding would result in smaller effects of perceived age on perception of prosody than impressionistic judgments of prosody from Experiments 1A and 1B.

Experiment 1A

Experiment 1A considered the issue of age-related bias in judgments of prosody, speech, and language and/or cognitive impairment. To investigate whether age-related bias is automatic, we compared the judgments of undergraduate participants who were unaware that the talker's speech may have been spectrally modified with the judgments of a group of undergraduate participants who were made aware of the potential unreliability of voice age characteristics. If age-related bias is a function of awareness of the unreliability of voice age characteristics, then a reduction in the effects of the spectral manipulation on judgments of prosody, speech, and impairment should be observed for the group that is made aware of voice age manipulations compared with the group that is not made aware.

Method

Participants and design. Fifty-six undergraduate students from Michigan State University participated in the experiment in return for partial credit in a psychology course or for monetary compensation. Participants were native speakers of American English who were at least 18 years of age with self-reported normal hearing. The design of the experiment was a 2 (awareness: unaware or aware) × 2 (talker age: child or adult) × 2 (modification: unmodified or modified) mixed factorial. Awareness was a between-subjects factor, and talker age and modification were within-subject factors. The unaware group consisted of 33 undergraduates (six men, 27 women) with a mean age of 19.5 years, and the aware group consisted of 23 undergraduates (nine men, 14 women) with a mean age of 21.5 years.

Stimuli. Stimuli for all experiments consisted of recordings of spontaneous speech produced during the telling of one of four "Frog Stories" picture books by Mercer Mayer (1967, 1969, 1973, 1975). Speech materials were produced by nine children (3 females, 6 males; M_{age} = 5;6 [years;months]) and their mothers (M_{age} = 31 years);[1] see Redford (2012) for details. Talkers were instructed not to interrupt the person telling the story, although minor interruptions occurred. The children were all in kindergarten and were tested within the first 4 months of starting school. All of the children were typically developing, with normal speech and hearing abilities. Recordings took place in a child-friendly, quiet experimental room in the Speech and Language Laboratory at the University of Oregon. The mean duration of the entire recording (including pauses) produced by adults was 154.6 s, and the mean duration produced by children was 136.1 s.

Each full recording was first segmented into three to 10 fragments (M = 5.8, total fragments = 105). Fragments for adult and child speech (including pauses) had mean lengths of 25.3 s and 24.5 s, respectively, and contained an average of 57.6 words and 33.6 words, respectively. To make talkers for each speech file sound older or younger, we used spectral modification tools in Praat (Boersma & Weenink, 2002). The first step was to change global spectral parameters for a talker using a pitch-synchronous overlap-add algorithm (Moulines & Charpentier, 1990). This involved setting a global value for each talker for two pitch measurement factors (i.e., the pitch floor and pitch ceiling) and two modification parameters (i.e., formant shift ratio and new pitch median); two other modification parameters (i.e., pitch range factor and duration factor) were left at the default value of 1.0 for all talkers. Specific values of global settings for the pitch measurement factors and modification parameters were restricted to particular ranges for adult versus child speech. Values were selected for which the talker seemed to be most convincingly transformed in age such that children's voices sounded maximally adultlike, and adults' voices sounded maximally childlike, and for which the speech was relatively free of artifacts. For the adults' speech, the F0 and formants were raised by a separate fixed factor for each speaker to values judged

[1]One mother had moved out of town, and her age could not be determined.

to make the speech sound as childlike as possible; in particular, the formant shift ratio was set to a value greater than 1.0, and adjustments were made to the new pitch median, the pitch floor, and the pitch ceiling (see Table 1 for details). For children's speech, the F0 and formants were lowered by a separate fixed factor for each speaker to values judged to make the speech of each talker sound adultlike; in particular, the formant shift ratio was set to a value less than 1.0, and adjustments were made to the new pitch median, the pitch floor, and the pitch ceiling (see Table 1). To create the most natural-sounding voices, the speech of all eight children was transformed into an adult male voice. The modifications left intact key prosodic attributes, including speech rate, F0 contour, pause timing, and length across spectral modification conditions.

After selection of global spectral parameters, the speech sounded very natural, with little if any perceptual indication of modification. To further minimize possible artifacts of modification, the speech was carefully checked auditorily. Occasional pitch errors were dealt with by hand-correction of the F0 contour using the pitch-synchronous overlap-add algorithm in Praat. The few other artifacts that occurred (i.e., crackling sounds or other unnatural-sounding isolated sections) were dealt with by first splicing out the short portion of modified speech evidencing the artifact, then subjecting the corresponding speech portion from the unmodified speech file to a different set of spectral modification parameters using a trial-and-error approach and obtaining the most natural-sounding result, and then (c) splicing the result back into the modified file.

To maintain a reasonable experiment length, we used a subset of the stimuli for Experiment 1A and 1B; all of the stimuli were used in Experiment 3. Two fragments were chosen from each child's and adult's retelling of the story; the fragments were at least 20 s long and contained at least 25 words. When multiple fragments met these qualifications, the two that contained the most words were chosen. This yielded 36 fragments (18 talkers × 2 fragments); each fragment furthermore occurred in both unmodified and modified forms. Selected fragments were 20.5 to 30.0 s long (M = 27.3) and contained 25 to 89 words (M = 54.5).

Four stimulus lists were created. First, the 36 speech fragments were divided into two blocks of 18 fragments; the first block contained one fragment from each of the 18 talkers, and the second block contained a second fragment from each talker. To create the first list, the 18 fragments within each block were arranged in a single, random order. Each fragment in the first block was then assigned pseudorandomly to either the unmodified or modified condition, and the complementary fragment by the same talker in the second block was assigned to the other condition. Sequencing and pairing were constrained so that the same modification condition did not appear more than four times in a row and the same combination of levels of talker age and modification did not occur more than three times in a row. A second stimulus list was created by inverting the modification condition from the first list. Two more lists were created by reversing the order of the first two lists. Each fragment occurred only once on a list, so participants never heard the same fragment in both unmodified and modified forms.

Task. Participants listened to a speech fragment and estimated the talker's age (0–100 years) and then provided judgments of seven measures while the scale for each question was displayed on the screen. Speech rate was assessed on a scale that ranged from 1 (*very slow*) to 6 (*very fast*). Fluency judgments were made on a scale that ranged from 1 (*always fluent*) to 8 (*always disfluent*). Participants were told that the term *fluency* referred to the broad spectrum of fluent speech characteristics, whether by normal talkers or people who stutter, and that *disfluency* referred to the broad spectrum of disfluent speech characteristics.[2] *Intelligibility* was defined as how understandable the talker was; participants gave an overall impressionistic estimate from 0 to 100 of the percentage of words they thought they could understand. For cognitive impairment, participants categorized talkers as either probably having normal cognitive abilities or probably having a cognitive impairment. For speech-language impairment, participants categorized talkers as probably having normal speech-language skills, probably having delayed speech-language skills, or probably having a speech-language disorder. For the purposes of later data analysis, the latter two categories were collapsed into a single category: "speech-language disorder or delay," given the difficulty in reliably distinguishing these categories by individuals trained in communication disorders (Bernthal et al., 2008), let alone laypersons. For likeability, participants judged the likeability of the talker on a scale that ranged from 1 (*extremely likeable*) to 6 (*extremely dislikeable*). Finally, participants judged the perceived anxiousness of the talker on a scale that ranged from 1 (*extremely anxious*) to 6 (*extremely calm*).

Procedure. Participants in the unaware condition estimated the talker's age and rated each of the seven attributes (see above) for each fragment. Participants in the aware condition were first familiarized with the set of speech fragments to be rated by estimating the talker's age for each fragment. After providing age estimates for fragments, participants in the aware condition were explicitly told the speech may have been modified so that talkers sounded older or younger than they actually were. They were then told that they would hear the set of speech fragments again and that they should ignore

[2]An expert well versed in fluency and stuttering approved our definition of fluency and disfluency for participants in Experiments 1A and 1B (S.-E. Chang, personal communication, August 11, 2010).

Table 1. Spectral modification parameters for experimental stimuli.

	Formant shift ratio		New pitch median (Hz)		Pitch floor (Hz)		Pitch ceiling (Hz)	
Nature of modification	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range
Adult → child	1.2 (0.03)	1.2–1.3	305.0 (32.8)	265–360	96.7 (22.8)	75–150	632.2 (33.5)	600–700
Child → adult	0.7 (0.05)	0.6–0.7	140.6 (34.8)	100–210	83.3 (14.8)	60–100	520.0 (113.7)	320–610

perceived age while making ratings of attributes. Approximately equal numbers of participants were randomly assigned to each of the four stimulus lists in the unaware and aware conditions. Before each experiment, participants in both conditions completed three practice trials using the same unmodified speech fragments so that they could become accustomed to each of the questions they would be asked during the session. Throughout the experiment, the scale associated with each question was displayed on the screen in order to reinforce consistent use of scale endpoint referents.

Apparatus. Participants listened to the speech fragments over Sennheiser HD 280 headphones. The experiment was presented to participants using E-Prime Version 2.0 Professional (Psychology Software Tools, Inc.) running on a Lenovo Intel Core2 Duo CPU E8500 with a 19-in. monitor.

Results

We conducted a 2 (awareness: unaware or aware) × 2 (talker age: child or adult) × 2 (modification: unmodified or modified) mixed-measures analysis of variance (ANOVA) for each dependent variable. The results are summarized in Table 2. Age estimates are shown in Figure 1, Panels A and B. There were significant main effects of both talker age and modification; the spectral modification was highly successful in creating different perceived ages for both awareness conditions. Unmodified speech sounded age appropriate (child talkers, $M = 6.3$ years; adult talkers, $M = 29.2$ years), whereas modified speech made child talkers sound older ($M = 19.4$ years) and adult talkers sound younger ($M = 13.6$ years), as evidenced by the significant interaction between talker age and modification. Effects of modification on estimated talker age were significant for both child talkers, $t(55) = 18.00$, $p < .001$, and adult talkers, $t(55) = –28.90$, $p < .001$. The three-way interaction among awareness, talker age, and modification revealed that having participants first estimate talker age for all fragments before making ratings of the different attributes yielded slightly more conservative age estimates (i.e., estimates that were closer to the actual ages of the talkers).

Speech rate ratings are shown in Figure 2, Panels A and B. Similar to the age analysis, there were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. For the child talkers, modifying speech to sound older yielded a slower perceived speech rate (unmodified, $M = 2.72$; modified, $M = 1.96$), $t(55) = 11.94$, $p < .001$, $d = –1.60$, whereas for adult talkers, modifying speech to sound younger yielded a faster perceived speech rate (unmodified, $M = 3.56$; modified, $M = 3.95$), $t(55) = –6.16$, $p < .001$, $d = 0.82$. There was no main effect of awareness or interaction with this factor. This means that making participants aware of the unreliability of voice age characteristics did not mitigate the age-related bias in judgments of speech rate.

Fluency ratings are shown in Figure 3, Panels A and B. Paralleling age and speech rate analyses, there were main effects of talker age and modification, as well as a reliable interaction between talker age and modification. For both child and adult talkers, age modification of speech yielded lower fluency ratings; however, there was a larger effect of modification on fluency ratings for child talkers (unmodified, $M = 4.99$; modified, $M = 3.83$), $t(55) = 10.31$, $p < .001$, $d = 1.38$, than for adult talkers (unmodified, $M = 7.30$; modified, $M = 6.88$), $t(55) = 6.80$, $p < .001$, $d = 0.91$. Similar to the speech rate analysis, there was no main effect of awareness or interactions with awareness: Making participants aware of the unreliability of voice age characteristics did not mitigate effects of the speech modification on perceived speech fluency.

Intelligibility ratings are shown in Figure 4, Panels A and B. As with the other measures, the ANOVA revealed main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. As with fluency ratings, spectral modification of a talker's speech to sound older or younger yielded lower ratings of intelligibility; however, effects of age modification on intelligibility ratings were significant for child talkers (unmodified, $M = 86.2$; modified, $M = 74.2$), $t(55) = –11.10$, $p < .001$, $d = –1.48$, but not for adult talkers (unmodified, $M = 97.8$; modified, $M = 94.8$), $t(55) = –1.64$, $p = .10$. The lack of a main effect of awareness or an interaction with awareness reveals that directing participants' attention to the unreliability of voice age characteristics did not affect intelligibility judgments; neither did it mitigate the effects of spectral modification on intelligibility judgments.

Table 2. Summary of results of three-way analyses of variance for Experiment 1A with variables of talker age, modification, and awareness.

	Main effect									Interactions											
	Talker age			Modification			Awareness			Talker Age × Modification			Awareness × Talker Age			Awareness × Modification			Awareness × Talker Age × Modification		
Dependent variable	*F*(1, 54)	*p*	η_p^2	*F*(1, 54)	*p*	η_p^2	*F*(1, 54)	*p*	η_p^2	*F*(1, 54)	*p*	η_p^2	*F*(1, 54)	*p*	η_p^2	*F*(1, 54)	*p*	η_p^2	*F*(1, 54)	*p*	η_p^2
Age	396.80	< .001	.88	9.24	.004	.15	*ns*	*ns*	*ns*	781.40	< .001	.94	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	5.90	.019	.10
Speech rate	325.40	< .001	.86	18.18	< .001	.25	*ns*	*ns*	*ns*	143.20	< .001	.73	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Fluency	483.00	< .001	.90	157.80	< .001	.75	*ns*	*ns*	*ns*	27.29	< .001	.34	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Intelligibility	133.00	< .001	.71	41.79	< .001	.44	*ns*	*ns*	*ns*	22.11	< .001	.29	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Cognitive abilities	855.40	< .001	.94	773.10	< .001	.94	5.11	.028	.09	738.20	< .001	.93	*ns*	*ns*	*ns*	5.95	.018	.10	4.33	.042	.07
Speech-language impairment	679.80	< .001	.93	152.00	< .001	.74	*ns*	*ns*	*ns*	108.90	< .001	.67	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Likeability	30.30	< .001	.36	78.61	< .001	.59	*ns*	*ns*	*ns*	15.17	< .001	.22	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Anxiety	37.30	< .001	.41	52.19	< .001	.49	*ns*	*ns*	*ns*	12.86	< .001	.19	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*

Figure 1. Age ratings, in years, as a function of talker age and modification for participants in the unaware condition (Panel A) or the aware condition (Panel B) of Experiment 1A, as well as speech-language pathology (SLP) master's student participants in Experiment 1B (Panel C). Error bars indicate ±1 standard error. Exp. = Experiment; yrs = years.

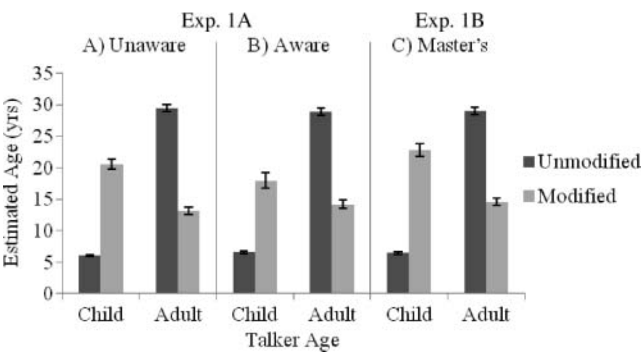
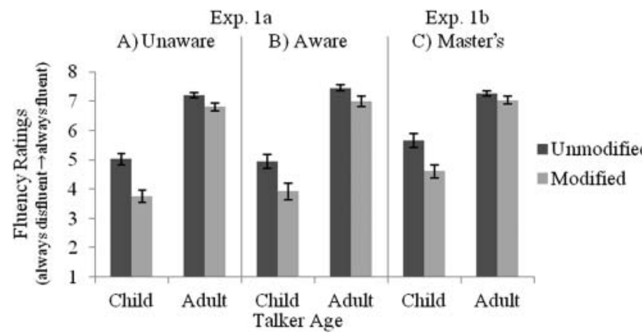


Figure 3. Fluency ratings as a function of talker age and modification for participants in the unaware condition (Panel A) or the aware condition (Panel B) of Experiment 1A, as well as SLP master's student participants in Experiment 1B (Panel C). Error bars indicate ±1 standard error.



Percentages of talkers classified with probable cognitive impairment are shown in Figure 5, Panels A and B. There were main effects of talker age, modification, and awareness, as well as Talker Age × Modification and Modification × Awareness interactions, and a three-way Talker Age × Modification × Awareness interaction. Modifying child talkers' speech substantially increased cognitive impairment classification (unmodified, $M = 11.0\%$; modified, $M = 86.0\%$), $t(55) = -27.92$, $p < .001$, $d = -3.73$. In contrast, modifying adult talkers' speech yielded no effect on cognitive impairment classification (unmodified, $M = 1.6\%$; modified, $M = 1.6\%$), $t(55) = 0.00$, $p > .99$. The effect of awareness and interactions with awareness suggest that making participants aware of the unreliability of age characteristics somewhat lessened the effect of the modification on cognitive impairment assessments of child talkers; however, effects of the modification on assessments of child talkers were reliable for both the unaware condition (unmodified, $M = 11.4\%$; modified, $M = 91.3\%$), $t(32) = -27.90$, $p < .001$, $d = -4.86$, and the aware condition (unmodified, $M = 10.6\%$; modified, $M = 78.4\%$), $t(22) = -14.20$, $p < .001$, $d = -2.96$.

Percentages of talkers classified with probable speech-language impairment are shown in Figure 6, Panels A and B. There were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. Children were much more likely to be judged to have speech-language impairments than were adults. Modifying child talkers' speech produced a large increase in assessed speech-language impairments (unmodified, $M = 44.0\%$; modified, $M = 93.5\%$), $t(55) = -12.50$, $p < .001$, $d = -1.67$, but the modification had no effect on assessed speech-language impairment in adult talkers (unmodified, $M = 5.5\%$; modified, $M = 5.5\%$), $t(55) = 0.01$, $p = .99$. The lack of main effect of or interaction with awareness reveals that directing participants' attention to the unreliability of voice age did not alter assessments of speech-language impairments. This lack of interaction is striking, given that the age modification more than

Figure 2. Speech rate ratings as a function of talker age and modification for participants in the unaware condition (Panel A) or the aware condition (Panel B) of Experiment 1A, as well as SLP master's student participants in Experiment 1B (Panel C). Error bars indicate ±1 standard error.

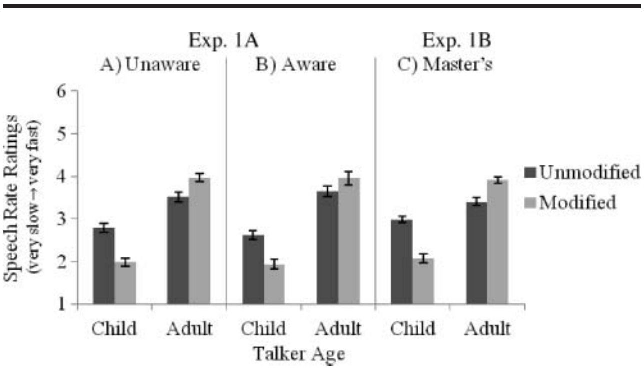


Figure 4. Intelligibility ratings as a function of talker age and modification for participants in the unaware condition (Panel A) or the aware condition (Panel B) of Experiment 1A, as well as SLP master's student participants in Experiment 1B (Panel C). Error bars indicate ±1 standard error.

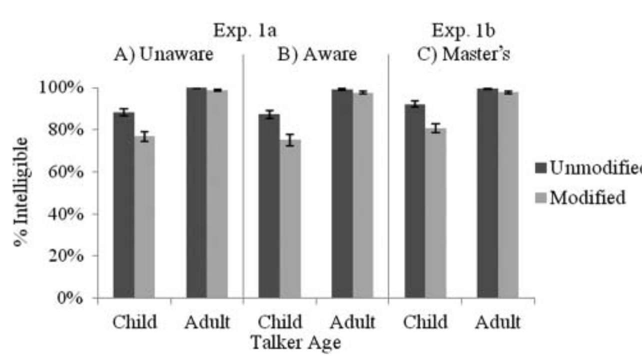
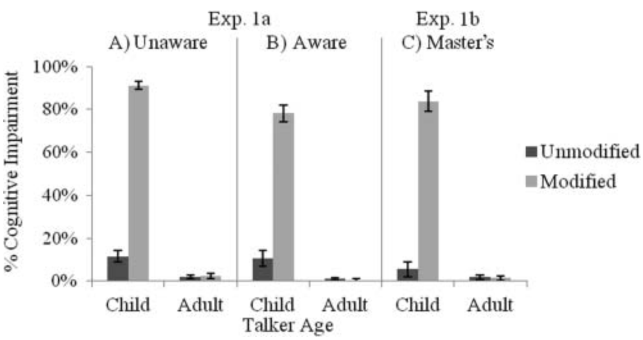


Figure 5. Percent judged to have a cognitive impairment as a function of talker age and modification for participants in the Unaware condition (Panel A) or the Aware condition (Panel B) of Experiment 1A, as well as SLP Master's student participants in Experiment 1B (Panel C). Errors bars indicate ±1 standard error.



doubled the rate of assessed speech-language impairments for child talkers.

Likeability ratings are shown in Figure 7, Panels A and B. As with the other measures, there were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. Spectral modification of the talker's speech yielded less likable talkers; however, there was a larger effect of modification on likeability ratings for child talkers (unmodified, $M = 2.68$; modified, $M = 3.62$), $t(55) = -10.15$, $p < .001$, $d = -1.36$, than for adult talkers (unmodified, $M = 2.38$; modified, $M = 2.69$), $t(55) = -2.79$, $p < .01$, $d = -0.38$. The lack of interaction with the awareness variable reveals that directing participants' attention to the unreliability of voice age did not alter perception of the rated attribute.

Anxiety ratings are shown in Figure 8, Panels A and B. There were main effects of talker age and modification as well as a reliable Talker Age × Modification interaction.

Figure 6. Percentage of talkers judged to have a speech-language delay or disorder as a function of talker age and modification for participants in the unaware condition (Panel A) or the aware condition (Panel B) of Experiment 1A, as well as SLP master's student participants in Experiment 1B (Panel C). Errors bars indicate ±1 standard error.

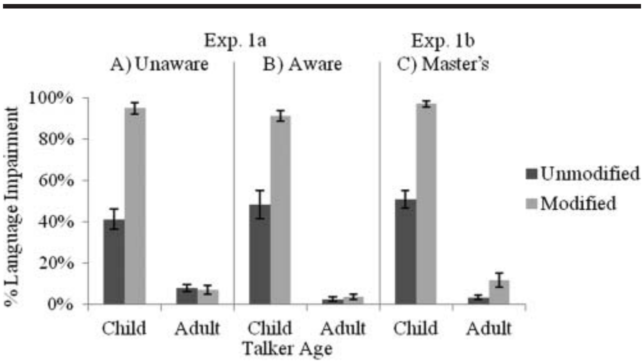
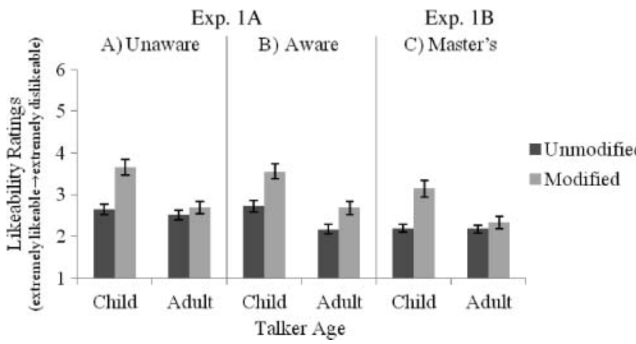


Figure 7. Likeability ratings as a function of talker age and modification for participants in the unaware condition (Panel A) or the aware condition (Panel B) of Experiment 1A, as well as SLP master's student participants in Experiment 1B (Panel C). Errors bars indicate ±1 standard error.

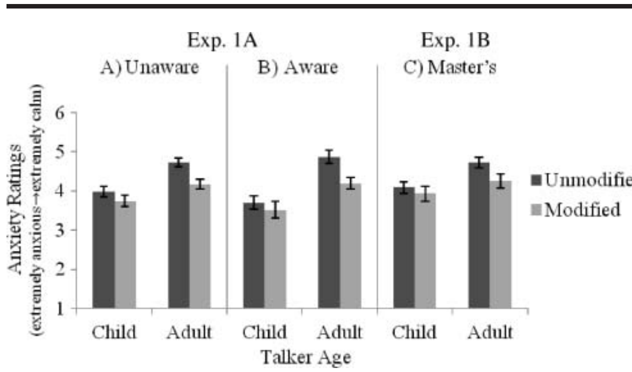


Adults were generally perceived to be less anxious than children, and spectral modification of the talker's speech yielded greater perceptions of talker anxiety. Unlike many other measures, however, effects of modification on the rated attribute were larger for adult talkers (unmodified, $M = 4.79$; modified, $M = 4.17$), $t(55) = 8.64$, $p < .001$, $d = 1.15$, than child talkers (unmodified, $M = 3.86$; modified, $M = 3.65$), $t(55) = 2.36$, $p < .05$, $d = 0.32$. As with most measures, there was no main effect of or interaction with awareness, indicating that directing participants' attention to the unreliability of voice age did not alter perceptions of talker anxiety.

Discussion

In Experiment 1A, we used a spectral manipulation to alter voice age characteristics of talkers to determine its possible effect on judgments of prosody, speech, and talker

Figure 8. Anxiety level as a function of talker age and modification for participants in the unaware condition (Panel A) or the aware condition (Panel B) of Experiment 1A, as well as SLP master's student participants in Experiment 1B (Panel C). Errors bars indicate ±1 standard error.



characteristics. The manipulation was very successful in altering perceived talker age. Modification made child talkers sound much older and adult talkers sound much younger. Moreover, altering perceived talker age influenced judgments on a wide range of dimensions. Speech of child talkers that was modified to sound like that of older talkers was judged to be slower, less fluent, and less intelligible. Child talkers with modified speech were judged to be less likeable, more anxious, and more likely to have cognitive and/or speech-language impairments. Effects of modification on judgments for adult talkers were less pronounced but still quite apparent. Speech of adult talkers that was modified to sound like it was produced by much younger talkers was judged to be faster and less fluent but not less intelligible. Adult talkers with modified speech were judged to be less likeable and more anxious but not more likely to have cognitive or speech-language impairments.

A second question addressed in this part of the experiment was whether awareness of the unreliability of voice age characteristics would reduce bias in judgments of talkers' speech and attributes. The answer to this question is generally "no." The results revealed no reduction in degree of age-related bias for most measures, including prosody-related measures (speech rate, fluency), as well as intelligibility, likelihood of a speech-language disorder or delay, likeability, and anxiety. Only for assessments of cognitive impairment was there a significant difference in degree of age-related bias observed for participants made aware of the unreliability of voice age, relative to the group that was not aware of this issue. Effects of awareness on cognitive abilities ratings were limited to reductions in the frequency of rating as cognitively impaired those child talkers in the modified speech condition. Overall, these findings reveal that age-related bias in judgments of prosody, speech, and impairment are relatively automatic, such that the listener has little or no control over such bias, an issue that has implications for developmental prosody research as well as clinical practice. These findings also support the hypothesis that ratings of prosody and other speech attributes depend not only on isolated characteristics of a talker's speech but also on the listener's overall assessment of the talker's characteristics of an individual, including those related to probable impairment.

One question is whether participants may have guessed the purpose of the experiment and were simply responding in a manner that conformed to how they were supposed to respond. To address this question, post-experiment surveys were administered to all participants that asked open-ended questions about their impressions of the talkers and files, as well as what they thought the purpose of the experiment was. The results revealed that 0% of participants in either group guessed the purpose, which lessens our concern about the role that demand characteristics may have played in these findings. In sum, the results of Experiment 1A show, first, that naïve listeners are influenced by the perceived age of the talker on many dimensions in judging speech and talker characteristics and, second, that awareness of the unreliability of voice age characteristics does little to mediate any of observed age bias.

Experiment 1B

In Experiment 1B, we considered whether the main finding of Experiment 1A—namely, that perceived talker age influences judgments of prosodic, speech, and impairment characteristics—also extended to a sample of SLP master's students. Because such individuals have had considerably more training and exposure to speech-language issues than the undergraduate psychology participants in Experiment 1A, they might be less susceptible to age-related bias in assessment of prosodic, speech, and impairment characteristics.

Method

Participants and design. Participants were 24 SLP master's students (all female) with self-reported normal hearing who were native speakers of American English and at least 18 years of age ($M = 24.7$). All participants were enrolled at Michigan State University and received course credit or monetary compensation for participation. Twenty-one of the students were in their first year of the master's program, and three were in their second year. All had completed a full sequence of prerequisite courses in communication sciences and disorders, including dedicated courses in language development, speech-language disorders, speech-language evaluation and treatment procedures, speech sciences, hearing sciences, phonetics, and other courses. The experiment implemented a 2 (talker age: child or adult) × 2 (modification: unmodified or modified) within-subject factorial design.

Stimuli, apparatus, task, and procedure. The stimuli and equipment were the same as in Experiment 1A. The task and procedure matched all details of the unaware condition in Experiment 1A.

Results

We conducted a 2 (talker age: child or adult) × 2 (modification: unmodified or modified) repeated measures ANOVA for each dependent variable; the results are summarized in Table 3. Age estimates are shown in Panel C of Figure 1. As in Experiment 1A, the modification was highly successful in creating different perceived

Table 3. Summary of results of two-way analyses of variance for Experiment 1B.

Dependent variable	Main effect: Talker age $F(1, 22)$	p	η_p^2	Main effect: Modification $F(1, 22)$	p	η_p^2	Interaction: Talker Age × Modification $F(1, 22)$	p	η_p^2
Age	133.99	< .001	.86	5.44	.029	.20	237.58	< .001	.92
Speech rate	168.48	< .001	.88	8.34	.009	.28	61.41	< .001	.74
Fluency	176.45	< .001	.89	73.32	< .001	.77	7.51	.012	.25
Intelligibility	59.04	< .001	.73	56.36	< .001	.72	34.98	< .001	.61
Cognitive abilities	243.66	< .001	.92	198.81	< .001	.90	189.02	< .001	.90
Speech-language impairment	315.05	< .001	.94	46.88	< .001	.68	33.53	< .001	.60
Likeability	16.93	< .001	.44	39.90	< .001	.65	*ns*	*ns*	*ns*
Anxiety	16.08	.001	.42	16.98	< .001	.44	7.83	.010	.26

talker ages. Unmodified speech sounded age appropriate (child talkers, M = 6.5 years; adult talkers, M = 29.1 years), whereas modified speech made child talkers sound older (M = 23.0 years) and adult talkers sound younger (M = 14.6 years), accounting for the significant Talker Age × Modification interaction. Post hoc paired-samples t tests showed that the modification affected estimated talker age for both child talkers, $t(23) = -15.40$, $p < .001$, $d = -3.13$, and adult talkers, $t(23) = 14.50$, $p < .001$, $d = 2.96$.

Speech rate assessments are shown in Panel C of Figure 2. Similar to Experiment 1A, there were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. Modifying child talkers' speech to sound older yielded slower perceived speech rates (unmodified, M = 2.98; modified, M = 2.07), $t(23) = 10.75$, $p < .001$, $d = -2.19$, whereas modifying adult talkers' speech to sound younger yielded faster perceived speech rates (unmodified, M = 3.40; modified, M = 3.91), $t(23) = -5.60$, $p < .001$, $d = 1.14$. The results showed that SLP master's students were susceptible to the same age-related bias in judging speech rate as the naïve participants tested in Experiment 1A.

Fluency ratings are shown in Panel C of Figure 3. Similar to Experiment 1A, there were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. Modification yielded lower ratings of fluency for both child and adult talkers, with a larger effect on fluency ratings for child talkers (unmodified, M = 5.64; modified, M = 4.60), $t(23) = 8.65$, $p < .001$, $d = 1.76$, than for adult talkers (unmodified, M = 7.26; modified, M = 7.04), $t(23) = 2.39$, $p < .05$, $d = 0.49$.

Intelligibility ratings are shown in Panel C of Figure 4. Similar to Experiment 1A, there were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. As with fluency ratings, modification yielded lower intelligibility ratings for both child and adult talkers. Modification effects were significant for both child talkers (unmodified, M = 91.9; modified, M = 80.6), $t(23) = -14.20$, $p < .001$, $d = -1.66$, and adult talkers (unmodified, M = 99.4; modified, M = 97.8), $t(23) = -2.99$, $p < .01$, $d = -0.61$, with the Talker Age × Modification interaction being due to the stronger effect of modification on ratings for child talkers than for adult talkers. These findings suggest that SLP master's students are strongly biased to interpret the same set of articulatory cues produced by children as less intelligible when those cues occurred with an adult's voice (modified condition) than with a child's voice (unmodified condition).

Percentages of talkers classified as cognitively impaired are shown in Panel C of Figure 5. Similar to Experiment 1A, there were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. Child talkers were more likely to be classified as having a cognitive impairment than adults. Moreover, modifying child talkers' speech produced a large increase in classification of talkers as having a cognitive impairment (unmodified, M = 5.6%; modified, M = 83.8%), $t(23) = -14.10$, $p < .001$, $d = -2.88$, whereas modifying adult talkers' speech had no effect on whether they received a cognitive-impairment classification (unmodified, M = 1.9%; modified, M = 1.4%), $t(23) = 0.30$, $p = .77$.

Percentages of talkers judged to have a speech-language impairment are shown in Panel C of Figure 6. Similar to most other measures, there were main effects of talker age and modification, as well as a significant Talker Age × Modification interaction. Child talkers were more likely than adults to be classified as having a speech-language impairment; modifying child talkers' speech almost doubled the likelihood of classifying a talker as having a speech-language impairment (unmodified, M = 50.8%; modified, M = 97.5%), $t(23) = -9.50$, $p < .001$, $d = -1.94$. Modification also had a significant, though weaker, effect on percentages of adult talkers classified with a speech-language impairment (unmodified, M = 3.2%; modified, M = 11.6%), $t(23) = -2.58$, $p < .05$,

$d = -0.53$, accounting for the Talker Age × Modification interaction.

Likeability ratings are shown in Panel C of Figure 7. Similar to Experiment 1A, there were main effects of talker age and modification, as well as a reliable Talker Age × Modification interaction. As in Experiment 1A, modification of speech generally produced less likable talkers. The Talker Age × Modification interaction revealed that the effect of modification on likeability was reliable for child talkers (unmodified, $M = 2.19$; modified, $M = 3.14$), $t(23) = -6.26$, $p < .001$, $d = -1.26$, but not adult talkers (unmodified, $M = 2.18$; modified, $M = 2.33$), $t(23) = -1.07$, $p = .29$.

Anxiety ratings are shown in Panel C of Figure 8. There were main effects of talker age and modification, but there was no reliable Talker Age × Modification interaction. Adults were judged as less anxious than children, and modification increased perceptions of talker anxiety. Although the Talker Age × Modification interaction was not reliable, post hoc *t*-tests revealed a significant effect of modification for adult talkers (unmodified, $M = 4.73$; modified, $M = 4.26$), $t(23) = 4.06$, $p < .001$, $d = 0.83$, but not child talkers (unmodified, $M = 4.08$; modified, $M = 3.93$), $t(23) = 0.99$, $p = .33$.

Discussion

The results of Experiment 1B replicated the general pattern of findings observed in Experiment 1A in showing that SLP master's students are also susceptible to similar age-related bias in assessing talker characteristics on a range of dimensions. As in Experiment 1A, when the speech of child talkers was spectrally modified to sound as if the talkers were older, it was judged to be slower, less fluent, and less intelligible than unmodified child speech. Child talkers with spectrally modified speech were also judged to be less likeable and were more likely to be judged to have cognitive and/or speech-language impairments. Also similar to Experiment 1A, effects of age-modified adult speech on judgments of talker characteristics were less pronounced, but still quite apparent. Modified speech of adult talkers was judged to be faster and less fluent, but not less intelligible. Adult talkers with age-modified speech were also judged to be less likeable and more anxious and were also more frequently judged to have speech-language impairments. The spectral modification, however, did not affect judgments of whether adult talkers had a cognitive impairment.

In sum, Experiment 1B showed that the specialized speech and language training and experience of the SLP master's students did not inoculate them against age bias in judgments of talker characteristics that are largely predicted to be unrelated to talker age. These findings indicate that age-related bias may be difficult to overcome even with additional training in speech-language issues. However, experience with and exposure to speech-language issues and language development did not appear to create more age-related bias in the SLP master's students' judgments compared with those of the undergraduates of Experiment 1A, as would have been predicted by findings that speech perception can be affected by experience with and exposure to particular talker groups or group characteristics (e.g., Drager, 2011). Overall, the findings support the hypothesis that ratings of prosody and other speech attributes depend not only on isolated characteristics of a talker's speech but also on the listener's overall assessment of an individual.

One question raised here is the extent to which intelligibility rating differences across spectral modification conditions in Experiments 1A and 1B could have been due to signal degradation associated with the spectral modification itself, as opposed to age-related bias. We considered this possibility in Experiment 2.

Experiment 2

In Experiment 2, we presented a new sample of naïve participants with a subset of the child talker and adult talker speech in unmodified and modified conditions and asked participants to transcribe the words they heard. If the intelligibility effects observed in Experiment 1A and 1B were due simply to signal degradation associated with the spectral modification, then transcription accuracy would be expected to be worse for the modified speech stimuli than for the unmodified speech stimuli. On the other hand, if the intelligibility effects observed in Experiment 1A and 1B reflected an age bias, then transcription accuracy in Experiment 2 would be expected to be similar in the modified and unmodified talker conditions.

Method

Participants and design. Participants were 16 undergraduate students (seven women, nine men) with a mean age of 20.8 who received course credit for their participation. A 2 × 2 within-subject factorial design was used with independent variables of talker age (adult or child) and modification (unmodified or modified). The dependent variable was the proportion of words correctly transcribed by participants in each condition.

Stimuli. The stimuli comprised a subset of those described in Experiment 1; in particular, the speech fragment with the largest word count for each talker was selected. The mean word counts for the child and adult files were 43.7 words and 65.1 words, respectively, and the mean durations of child and adult files were 27.0 s and 27.7 s, respectively.

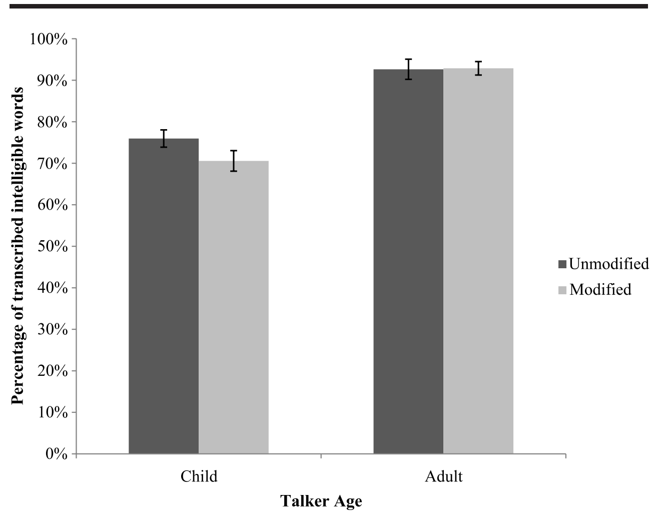
Apparatus and procedure. Lenovo Intel Core2 Duo CPU E8500 computers with 19-in. screens running Microsoft Word 2007 and Microsoft Media Player 10 connected to Sennheiser HD 280 Professional headphones were used for the experiment. Two counterbalanced lists of speech files were created, each of which consisted of 18 speech fragments (one from each of the nine child and nine mother talkers). Half of the fragments were in the unmodified condition, and half were in the modified condition. Each speech fragment occurred only once in the entire list in one type of modification. Participants were randomly assigned to one of the two lists, with equal numbers assigned to each list. Participants were directed to the folder on the computer that contained the stimuli for their list and were instructed to listen to the files in the order specified on the computer screen. Participants were asked to listen to a speech file in Microsoft Media Player as many times as necessary and then to transcribe the speech of the main speaker in each file into a template in a Microsoft Word document. They were instructed to use only lowercase letters in their transcription and not to use punctuation (to facilitate comparison of transcriptions). In addition, they were instructed not to transcribe any nonspeech noises (e.g., laughter, singing). Participants were also asked to transcribe partial words with the heard phonemes followed by a dash (e.g., *he*–, *wh*–), to transcribe filler words as *um* and *uh*, and to transcribe "X" for all consecutive unintelligible words. The entire experiment took 60 to 90 min.

Data analysis. A transcription key was first prepared for each speech fragment by two trained phonetic analysts from the acoustic–phonetic information in the fragment's unmodified version. Of particular interest was the identification of words in each fragment that were unanimously deemed intelligible by both of the trained phonetic analysts, as well as those words in the fragment that were unintelligible for reasons other than spectral modification (e.g., developmental misarticulation, soft speech, background noise, etc.). Our goal was to compare transcription accuracy by naïve listeners in unmodified versus modified speech conditions with words marked unanimously intelligible to the trained phonetic analysts in the unmodified speech fragment, which made it possible to determine the extent to which spectral modification per se degraded intelligibility. Phonetic analysts were given the same guidelines for transcription as participants but listened to the unmodified speech files while viewing spectrograms and waveforms in Praat. The use of combined acoustic and auditory cues by the trained phonetic analysts was expected to generate excellent agreement between the two analysts on words that were intelligible in the unmodified speech files. The transcription key for each file consisted of words that were transcribed by (and thus intelligible to) both analysts; words that were not agreed on exactly by both analysts and those transcribed as "X" were treated as unintelligible. The transcription key was then compared against the naïve listeners' transcripts to determine the rate with which naïve participants correctly identified each word in the unmodified versus modified conditions. This permitted an estimate of the extent to which the spectral modification might have reduced intelligibility of the speech. This metric was calculated for each participant for each talker age and modification condition by dividing the number of words correctly identified (i.e., which matched intelligible words unanimously agreed on in the transcription key) by the total number of words identified in the transcription key for that file.

Results

The proportions of intelligible words in each condition are shown in Figure 9. The mean percentage of intelligible words heard for the child speech was 76.0% for unmodified and 70.6% for modified, and for the adult speech, it was 92.6% for unmodified and 92.9% for modified. A 2 (talker age: child or adult) × 2 (modification: unmodified or modified) ANOVA by participants was conducted on these data; both talker age and modification were within-subject factors. There was a significant effect of talker age, $F(1, 15) = 791.50, p < .001, \eta_p^2 = .98$, but no effect of modification, $F(1, 15) = 1.69, p = .214$, and no interaction, $F(1, 15) = 1.89, p = .190$.

Figure 9. Average percentage of transcribed intelligible words as a function of talker age and modification for Experiment 2. Error bars indicate ±1 standard error.



Discussion

The main finding from this experiment was that the spectral modification used here did not significantly reduce intelligibility in a transcription task. In particular, the lack of a significant effect of modification and absence of an interaction with modification indicates that the modification was not responsible for degradation in the proportion of words understood. These results suggest that the effects of Modification on perceived intelligibility observed in Experiments 1A and 1B were primarily the result of age-related bias as opposed to actual differences in intelligibility per se between modified and unmodified speech. Therefore, the results of Experiment 2 lend support to our interpretation from Experiments 1A and 1B that listeners were biased to interpret the same articulatory cues produced by children as less intelligible when those cues occurred with an adult's voice (child-modified condition) than when they occurred with a child's voice (child-unmodified condition) because misarticulations are common for 5-year-old children but rare for typical adults. Returning to the issue of prosody, we asked in Experiment 3 whether a different method of reporting prosody might result in less age-related bias than the impressionistic judgments of Experiments 1A and 1B.

Experiment 3

Method

Participants and design. Participants were 10 prosody analysts (eight women, two men) at Michigan State University (M_{age} = 20.6 years) who volunteered or received pay for participation. All analysts had participated in a one-semester seminar on the Rhythm and Pitch (RaP) system of prosody transcription (Breen, Dilley, Kraemer, & Gibson, 2012; Dilley & Brown, 2005), which is a method of annotating prosody in speech, including phrasal boundaries, prominences, pitch accents, disfluencies, and other information. The seminar was taught by the first author; participants passed tests of proficiency at the end of the seminar. The seminar included weekly practice in transcription, and participants attended weekly group transcription practice sessions following the seminar. A 2 × 2 within-subject factorial design was used with the independent variables talker age (adult or child) and modification (unmodified or modified). The dependent variables were the number of phrasal boundaries and disfluencies marked by analysts in the four conditions (adult–unmodified, adult–modified, child–unmodified, child–modified).

Stimuli. Experiment 3 materials included the exhaustive stimulus set described in Experiment 1A. There were 105 unmodified fragments from the 18 talkers (nine child, nine adult) and their 105 modified counterparts.

Procedure. The analysts used the RaP labeling system to annotate the locations and types of disfluencies and phrasal boundaries. Before beginning work on the project, analysts reviewed the different types of disfluencies, listening to multiple examples of each type. Next, analysts were allowed to listen to a speech file as many times as necessary before providing an annotation. Two counterbalanced lists of speech files were created so that no analyst would ever listen to the same speech file in both modification conditions. Within a list, the sequential pairing of speech files with modification conditions was random but fixed for all participants assigned to that list. The analysts were instructed to listen to files in a designated folder sequentially from the first to the last file. Half the fragments on each list were from child talkers, and half were from adult talkers; in addition, half of the fragments were in the unmodified condition, and half were in the modified condition. This counterbalancing ensured that each analyst never heard the same speech file in both its unmodified and modified form.

Analysts were randomly assigned to one of the two speech file lists (five analysts per list). They labeled the phrasal boundaries and disfluencies using a modified version of RaP guidelines (Dilley & Brown, 2005), which involved indicating when a prosodic phrasal boundary occurred and how large it was. In particular, labels of ")?", ")", "))?", and "))" were used to indicate a possible minor phrasal boundary, a clear minor phrasal boundary, a possible major phrasal boundary, or a clear major phrasal boundary, respectively. Finally, analysts annotated each time a disfluency was perceived to occur.

Apparatus. The computers used for prosody analysis were either Lenovo Intel Core2 Duo CPU E8500 with 19-in. screens using Sennheiser HD 280 Professional headphones from the Michigan State University Speech Laboratory or personal computers owned by the analysts.

Analysis. The rate of marking a phrasal boundary—")", "))?", or "))"—was first calculated for each labeler in each condition, which corresponded to the number of times a phrasal boundary was marked, divided by the number of possible phrasal boundary locations (i.e., the number of word-final locations in the file, calculated as the number of words in the file minus 1 because the final word in the file was trivially a phrasal boundary; Breen et al., 2012). The rate of marking a disfluency corresponded to the number of times a disfluency was marked by each labeler in each condition, divided by the number of possible disfluency locations (i.e., the number of words or pauses minus 1; see above).

Analyses of interrater agreement were also carried out using the kappa (κ) metric, which adjusts for chance agreement levels on the basis of the number of coding distinctions and label frequency (Carletta, 1996). κ is calculated as in the following equation:

$$\kappa = \frac{A_O - A_E}{1 - A_E},$$

where A_E is expected chance agreement and A_O is observed (actual) agreement.

We calculated κ values following Breen et al. (2012). Here, specific labels were grouped into *label equivalence relations* to indicate how labels corresponded to the constructs of interest (i.e., presence and size of phrasal boundary and presence of disfluency). The label equivalence relations for this study are shown in Table 4. See Breen et al. (2012) for further details of calculation of κ.

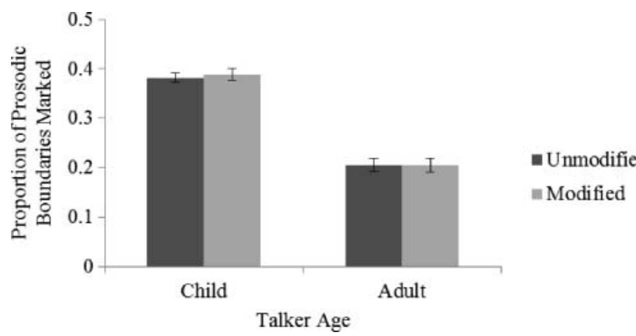
Results

We conducted a 2 × 2 repeated measures ANOVA by participants with talker age and modification as within-subject factors for each dependent variable. The proportion of phrasal boundaries marked is shown in Figure 10. More phrasal boundaries were identified for child speech ($M = 0.39$) than adult speech ($M = 0.21$). There was a significant effect of talker age, $F(1, 9) = 304.73, p < .001, \eta_p^2 = .97$, but no effect of modification, $F(1, 9) = 0.44, p = .52$, and no interaction, $F(1, 9) = 0.53, p = .49$. κ values ranged from .74 to .79, indicating high and substantial agreement.

The proportion of large phrasal boundaries marked is shown in Figure 11. More large phrasal boundaries were identified for child speech ($M = 0.31$) than adult speech ($M = 0.14$). There was a significant effect of talker age, $F(1, 9) = 142.56, p < .001, \eta_p^2 = .94$, but no significant effect of modification, $F(1, 9) = 1.87, p = .204$, and no interaction, $F(1, 9) = 0.06$, $p = .81$. κ values ranged from .82 to .86, indicating high and substantial agreement.

The proportion of disfluencies marked is shown in Figure 12. There was a significant effect of talker age, $F(1, 9) = 140.09$, $p < .001$, $\eta_p^2 = .94$; more disfluencies were identified for children's speech ($M = 0.14$) than for adult speech ($M = 0.02$). Moreover, there was a marginally significant effect of modification on the rate of labeling disfluencies, $F(1, 9) = 4.77$, $p = .057$, as well as a marginally significant interaction, $F(1, 9) = 3.45$, $p = .096$. κ values of interrater agreement across conditions ranged from .26 to .38, indicating moderate agreement.

Figure 10. Proportion of prosodic phrasal boundaries marked by labelers as a function of talker age and modification for Experiment 3. Error bars indicate ±1 standard error.



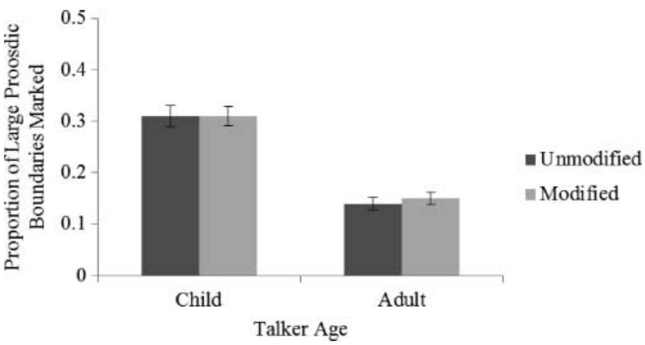
Discussion

In this experiment, analysts used a formal prosody labeling system to indicate prosodic phrasal boundaries and disfluencies in speech fragments produced by children and adults, where half of the fragments had been modified to make talkers sound older or younger, respectively. Careful counterbalancing ensured that no analyst heard any speech fragment more than once in a single modification condition. Analysts marked more prosodic phrasal boundaries and disfluencies in children's speech compared with adults' speech. Critically, spectral modification did not significantly affect rates of indicating prosodic phrasal boundaries or their perceived sizes. There was also no significant effect of the spectral manipulation on the rate of labeling disfluencies, although the effect approached significance. It is notable that effect sizes of the modification on rates of labeling disfluency were uniformly smaller than effect sizes of the manipulation on impressionistic fluency ratings in Experiments 1A and 1B. The results suggest that using a formal prosodic labeling system incurs less age-related bias in judging prosody than impressionistic judgments of prosody. This supports the usage of formal prosody labeling as a particularly valuable methodology

Table 4. Label equivalence relations used for calculating rates of labeling boundaries and disfluencies for Experiment 3.

Variable	Boundary	No Boundary	Large	Small	Disfluency	No disfluency
Presence of phrasal boundary	),))?,))	)?, no label				
Size of boundary			))?,))	)?,), no label		
Disfluency vs. no disfluency					Prolongation, truncation, repeat, mispronunciation, unfilled pause	Restart, no label

Figure 11. Proportion of large prosodic phrasal boundaries marked by labelers as a function of talker age and modification for Experiment 3. Error bars indicate ±1 standard error.



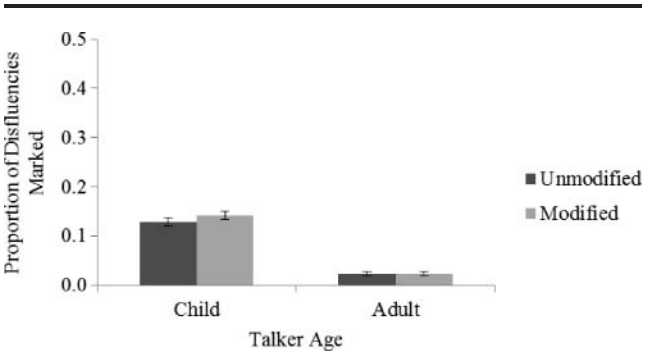
for investigation of developmental prosody issues while minimizing age-related bias.

General Discussion

In the present research, we used spectral manipulation to alter talkers' perceived ages and determine the subsequent effects on judgments of prosodic, segmental, clinical, and personal characteristics of speech and/or the talkers themselves. Understanding the effects of perceived age on talker and/or speech judgments is important for developmental work on speech and language. We investigated the extent to which talker age might affect such judgments in experiments using impressionistic ratings (Experiments 1A and 1B), orthographic transcription (Experiment 2), and formal prosody labeling (Experiment 3).

Experiments 1A and 1B demonstrated that differences in perceived talker age resulting from our spectral manipulation were associated with differences in every impressionistic measure investigated, including prosodic and articulatory variables (perceived speech rate, fluency, intelligibility) and variables related to talkers themselves (likelihood of speech-language and/or cognitive impairment, likeability, and anxiety). This is the first study to demonstrate that manipulations to F0 and formant frequencies affect judgments of talker age for children's speech; such effects have previously been demonstrated for adults' speech (Harnsberger et al., 2008; Reubold et al., 2010). Moreover, Experiment 2 confirmed that effects of perceived age on intelligibility ratings could not be attributed to signal degradation due to spectral modification. Finally, Experiment 3 showed that age-related bias in judgments of at least some prosodic variables may be mitigated by using a formal prosodic coding system. Overall, these findings confirm that perceived talker age affected listeners' standards for judging speech and talker attributes, demonstrating the existence of age-related bias in impressionistic judgments of these speech and talker characteristics.

Figure 12. Proportion of disfluencies marked by labelers as a function of talker age and modification for Experiment 3. Error bars indicate ±1 standard error.



Experiment 1A investigated whether the age-related bias in judgments of speech and talker characteristics can be reduced as a function of attention or whether such bias is automatic. We accomplished this by comparing ratings by undergraduate students who had been made aware of the spectral manipulation with ratings by naïve undergraduates. No reduction in degree of age-related bias was observed for the former group for the majority of measures. These findings suggest that age-related bias is relatively automatic.

In addition, investigating ratings in Experiment 1B by SLP master's students allowed us to determine whether additional learning and experience relating to speech-language issues affected age-related bias in judgments of prosody, speech, and talker characteristics. Although the majority of measures showed no differences in age-related bias across modification conditions, three measures showed a reduction in bias for SLP master's students relative to undergraduates: (a) age estimates, (b) fluency judgments, and (c) likeability. These findings suggest that additional learning and experience in speech-language issues may reduce age-related bias for at least some types of judgments but that, in general, age-related bias in judgments of prosodic, segmental, and impairment characteristics may be difficult to overcome. It is possible that had our participants been certified speech-language pathologists, who have had even more training and experience, even less age-related bias might have been observed, an issue that awaits future research.

It is worth considering whether an alternative explanation exists for effects of the spectral manipulation on judgments according to which listeners detected incongruity between the talker's true age and his or her perceived age and judged such "mismatched" speech more negatively than "matched" speech. Such an explanation predicts more negative judgments for "mismatched" speech on relevant dimensions than "matched" speech. In contrast to this prediction, for a number of properties for which one scale endpoint could be interpreted as more negative (i.e., cognitive impairment, language impairment, and intelligibility), "mismatched" adult speech was not assessed by listeners more negatively than "matched" speech by most listener groups across Experiments 1A and 1B. These findings argue against an explanation of the spectral modification manipulation effects based solely around incongruity detection. The age bias explanation is most consistent with the majority of results across these experiments; however, we cannot rule out the possibility that incongruity detection did not play a role in some conditions examined.

Our results suggest that age-related voice characteristics can affect a wide range of impressionistic judgments of attributes that are commonly assumed to be orthogonal or quasi-orthogonal to prosodic, segmental, and talker attributes. These findings are consistent with prior research demonstrating the interdependence of perceptual judgments of pitch, timing, and/or spectral profile (Bond et al., 1988) as well as findings showing that age estimates are affected by manipulations to F0, first formant (F1), and/or speech rate (Harnsberger et al., 2008; Reubold et al., 2010). In terms of theory, the present research suggests that prosody cannot be compartmentalized from other components of the speech signal (e.g., the segmental and indexical components, where the latter includes talker age). Therefore, our findings are consistent with a wide range of research demonstrating interdependency in perceiving and processing lexico-segmental, prosodic, and indexical components of speech (e.g., Ladefoged & Broadbent, 1957; Nygaard, Sommers, & Pisoni, 1994).

As expected, measures of speech and talker characteristics were also substantially affected by the baseline speech-language competency of talkers. Across studies, children were judged as being less competent (e.g., less fluent, slower in rate, less intelligible, more likely to be impaired) than adults. More interesting is that baseline speech-language competency interacted with spectral modification such that children's speech in general was judged very differently on many measures, depending on voice age characteristics, whereas adults' speech was less often and less substantially affected by the modification.

One consistent finding was that when children's speech was spectrally modified, talkers were particularly likely to be interpreted as having speech-language and cognitive impairments and to be perceived as significantly less likeable. These findings can be explained by the pairing of relatively high disfluency and misarticulation rates, limited vocabulary, and simpler syntax of children's speech, on the one hand, with an adult-sounding voice, on the other; such a situation is consistent with an adult with below-par speech-language and cognitive skills. Given that adult talkers with communication impairments are judged more negatively than talkers without such impairments (Franck et al., 2003; Lass et al., 1993), we can explain the decreased likeability of child talkers in the modified speech condition. The findings support the hypothesis that prosody and other speech ratings depend not only on perceived speech characteristics but also on an assessment of the talker as an individual, including possible impairment. Overall, the present research illustrates the potential of spectral modification as a tool for investigations of attitudes toward groups in relation to developmental, life span, and/or communication impairment issues.

Moreover, these findings raise cautionary notes for developmental research studies in which voice age characteristics co-vary with participant groups under study, or in which sufficient time elapses during the observation phase that maturation of voice age occurs. Our findings suggest that confounding may occur between primary prosodic variables of interest (e.g., intelligibility, speech rate) and voice age characteristics in longitudinal or cross-sectional developmental research designs, an issue that has seldom been considered in developmental studies of prosody. Many prosodic characteristics have been investigated in child speech, including phrasal boundary emergence and placement (Crystal, 1986), timing (Snow, 1997), disfluency (Gordon & Luper, 1989; Wexler & Mysak, 1982), speaking rate (Kelly & Conture, 1992), and phrase-level intonation (Chen, 2011; MacWhinney & Bates, 1978). In much of this work, children's prosodic behavior has been compared with that of adults or of children of other ages. This research suggests that in such comparative designs, differences in prosodic metrics could have partly reflected the influence of voice age characteristics, rather than merely the prosodic variables of interest. It would be wise for developmental speech-language research studies to include additional controls to guard against potential confounding between voice age characteristics and primary perceptual speech-related variables of interest if utilizing impressionistic judgments, or else to use more formalized prosodic coding methods.

The present results also have implications for assessment of speech in clinical situations. First, little research has examined the possibility of bias affecting clinical assessment of children's speech. Existing work suggests that a child's presumed age and gender influence adults' perceptions of children's speech and ratings of its accuracy and quality (Munson et al., 2010; Munson & Seppanen, 2009). These results suggest that a child's perceived age can bias perceptual judgments of speech in more far-reaching ways. Second, these results suggest that maturation of the voice could occur during clinical observations over time, which could be confounded with perceptions of other speech-language attributes of greater clinical interest (e.g., speech rate, fluency, intelligibility). Third, these results highlight a potential risk of unreliability in impressionistic perceptual judgments of speech variables of interest (e.g., speech rate, fluency, or intelligibility), due to potential misattribution of changes from voice age characteristics.

In addition to these cautionary notes, the present results suggest some positive new directions for therapeutic strategies aimed at altering how a talker's speech, degree of impairment, and/or self-presentation may be perceived by others. In particular, future research might explore whether some talkers might be coached to change their voice spectral profile or pitch to make themselves sound older or younger in certain situations in order to change others' perceptions of their speech rate, fluency, intelligibility, anxiety level, or the likelihood that they have cognitive or communicative impairments. For example, future work might explore the value of an adult with a communication impairment adopting something like a "Bart Simpson voice" (i.e., one that sounds more child-like) when talking over the telephone (e.g., Guo & Togher, 2008) in order to alter how others perceive her speech, impairment, or presentational characteristics. Candidates for such an intervention might be individuals who have difficulty controlling aspects of speech articulation or rate but who have good control over voice characteristics—for example, some individuals with dysarthria or people who stutter.

In summary, age-related spectral voice characteristics can affect judgments of speech and talker characteristics in somewhat surprising ways, given that these characteristics are often treated in theory and practice as orthogonal to one another. This research suggests cautionary notes for research and clinical settings when observing individuals of different ages or the same individual over time because voice age characteristics may be confounded with judgments of prosodic and segmental aspects of speech. Using a formalized coding system for prosody may minimize age-related bias relative to impressionistic ratings. Finally, this work points to the potential for developing intervention strategies for talkers with some types of communication impairment that target alterations in voice spectral characteristics to achieve changes in how others perceive the talker's speech, impairment, and/or self-presentation characteristics.

Acknowledgments

This work was supported by Grant R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health. We are grateful for the advice and assistance of Amanda Millett, Christine Collins, and other members of the Michigan State University Speech Laboratory.

References

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. H. Yeni-Komshien, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology: Vol. 1: Production* (pp. 227–256). New York, NY: Academic Press.

Bernthal, J. E., Bankson, N. W., & Flipsen, P., Jr. (2008). *Articulation and phonological disorders: Speech sound disorders in children* (6th ed.). Boston, MA: Allyn & Bacon.

Bloodstein, O. (1995). *A handbook on stuttering*. San Diego, CA: Singular.

Boersma, P., & Weenink, D. (2002). Praat: Doing phonetics by computer (Version 4.0.26) [Computer program]. Retrieved from http://www.praat.org

Bond, R. N., Simpson, S., & Feldstein, S. (1988). Relative and absolute judgments of speech rate from masked and content-standard stimuli: The influence of vocal frequency and intensity. *Human Communication Research, 14,* 548–568.

Breen, M., Dilley, L. C., Kraemer, J., & Gibson, E. (2012). Inter-transcriber agreement for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory, 8,* 277–312.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics, 22,* 249–254.

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in Dutch. *Journal of Child Language, 38,* 1055–1083.

Crystal, D. (1986). Prosodic development. In P. J. Fletcher & M. Garman (Eds.), *Studies in first language development* (pp. 174–197). New York, NY: Cambridge University Press.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800–814.

Dilley, L. C., & Brown, M. (2005). *The RaP (Rhythm and Pitch) labeling system, Version 1.0.* East Lansing: Michigan State University. Retrieved from http://speechlab.cas.msu.edu/rap-system.htm

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech, 54,* 99–121.

Franck, A. L., Jackson, R. A., Pimental, J. T., & Greenwood, G. S. (2003). School-age children's perceptions of a person who stutters. *Journal of Fluency Disorders, 28,* 1–15.

Gordon, P. A., & Luper, H. L. (1989). Speech disfluencies in nonstutterers. *Journal of Fluency Disorders, 14,* 429–445.

Guo, Y. E., & Togher, L. (2008). The impact of dysarthria on everyday communication after traumatic brain injury: A pilot study. *Brain Injury, 22,* 83–97.

Halle, M. (1985). Speculations about the representation of words in memory. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 101–104). New York, NY: Academic Press.

Harnsberger, J. D., Shrivastav, R., Brown, W. S. J., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice, 22,* 58–69.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-process. *Journal of Phonetics, 34,* 458–484.

Henry, M. J., McAuley, J. D., & Zaleha, M. (2009). Evaluation of an imputed pitch velocity model of the auditory tau effect. *Attention, Perception, & Psychophysics, 71,* 1399–1413.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27,* 359–384.

Kelly, E. M., & Conture, E. G. (1992). Speaking rates, response time latencies, and interrupting behaviors of young stutterers, nonstutterers, and their mothers. *Journal of Speech and Hearing Research, 35,* 1256–1267.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America, 29,* 98–104.

Lass, N. J., Ruscello, D. M., Harkins, K. E., & Blankenship, B. L. (1993). A comparative study of adolescents' perceptions of normal-speaking and dysarthric children. *Journal of Communication Disorders, 26,* 3–12.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 28,* 195–209.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behavior, 17,* 539–558.

Mayer, M. (1967). *A boy, a dog, and a frog*. New York, NY: Penguin Group.

Mayer, M. (1969). *Frog, where are you?* New York, NY: Penguin Group.

Mayer, M. (1973). *Frog on his own*. New York, NY: Penguin Group.

Mayer, M. (1975). *One frog too many*. New York, NY: Penguin Group.

Melara, R. D., & Marks, L. C. (1990). Interaction among auditory dimensions: Timbre, pitch and loudness. *Perception & Psychophysics, 48,* 169–178.

Menzies, R. G., Onslow, M., & Packman, A. (1999). Anxiety and stuttering: Exploring a complex relationship. *American Journal of Speech-Language Pathology, 8,* 3–10.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453–467.

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M., & Meyer, M. K. (2010). Deconstruction phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of *vox humana*. *Clinical Linguistics & Phonetics, 24,* 245–260.

Munson, B., Jefferson, S. V., & McDonald, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *The Journal of the Acoustical Society of America, 119,* 2427–2437.

Munson, B., & Seppanen, V. (2009, November). *Perceived gender affects ratings of the quality of children's spoken narratives*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5,* 42–46.

Peppe, S. J. E. (2009). Why is prosody in speech-language pathology so difficult? *International Journal of Speech-Language Pathology, 11,* 258–271.

Redford, M. A. (2012). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics, 1,* 1–21. doi:10.1017/S0142716411000877

Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication, 52,* 638–651.

Scherer, K. R. (2003). Vocal communications of emotion: A review of research paradigms. *Speech Communication, 40,* 227–256.

Snow, D. (1995). Formal regularity of the falling tone in children's early meaningful speech. *Journal of Phonetics, 23,* 387–405.

Snow, D. (1997). Children's acquisition of speech timing in English: A comparative study of voice onset time and final syllable vowel lengthening. *Journal of Child Language, 24,* 35–56.

Southwood, M. H., & Weismer, G. (1993). Listener judgments of the bizarreness, acceptability, naturalness, and normalcy of the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology, 1,* 115–118.

Sussman, J. E., & Sapienza, C. (1994). Articulatory, developmental and gender effects on measure of fundamental frequency and jitter. *Journal of Voice, 8,* 145–156.

Wexler, K., & Mysak, E. C. (1982). Disfluency characteristics of 2-, 4-, and 6-year-old males. *Journal of Fluency Disorders, 7,* 37–46.