Volume 46 http://acousticalsociety.org/

182nd Meeting of the Acoustical Society of America

Denver, Colorado

23-27 May 2022

Speech Communication: Paper 4pSC7

Regional dialect effects on the perception of gender in children's voices

Christopher Edward Holt, Ewa Jacewicz, Robert Fox and Lian Arzbecker

Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio, 43210; holt.465@osu.edu; ceholt80@gmail.com; jacewicz.1@osu.edu; fox.2@osu.edu; arzbecker.1@osu.edu

Studies investigating perception of a child's gender have found that adults can identify the gender of young children accurately before developmental changes in vocal fold size and vocal tract length provide reliable acoustic cues, including fundamental frequency (f_0) and formant frequencies. This suggests that judgments about gender in boys' and girls' voices may also be influenced by sociocultural and behavioral factors that shape children's production patterns. We tested whether gender identification in children can be influenced by regional dialect of American English. 26 adults from Central Ohio listened to 620 speech stimuli (comprised of isolated words, read sentences, and spontaneous talks) from 90 children ages 8-11 years representing three dialect regions in the United States (Central Ohio, Southeastern Wisconsin, and Western North Carolina). As predicted, gender was identified more accurately from sentences and spontaneous talk samples than from isolated words. Significant dialect effects were found, but they were manifested differently for younger and older children, and differently as a function of stimulus type. Overall, the results suggest that variations between regional dialects can provide additional cues to gender identification in children's voices, but these effects are variable and depend on a child's age, listening context, and dialect of a listener.

Published by the Acoustical Society of America

© 2022 Acoustical Society of America. https://doi.org/10.1121/2.0001603

1. INTRODUCTION

Adults can accurately identify the gender of children as young as four years of age,[1] which is well before the developmental changes in vocal fold size and vocal tract length give rise to reliable acoustic cues in fundamental frequency (f_0) and formant frequencies, respectively. These two acoustic cues are dominant in gender identification in adults as males typically have lower f_0 and lower formant frequencies than females.[2] But these cues are not yet prominent in young children, as developmental changes within the oral and pharyngeal sections of the vocal tract are manifested only later in childhood.[3] The reported high gender identification rates for young children, despite the absence of reliable acoustic cues in their voices, prompt questions about the ways in which we perceive the gender of a talker based on their voice and speech characteristics.

Possibly, these adult listeners in Perry et al.[1] found some cues to gender in 4- and 8-year-olds in vowel formant frequencies in single syllables that were about 9% lower in boys than in girls (a similar 10%-difference was also reported elsewhere for 7- and 8-year-olds)[4]. However, it is also possible that boys and girls around 8 years of age begin to demonstrate learned characteristics of male and female speech that are influenced by behavioral factors. Support for this claim comes from another study[5] that asked adult listeners to identify the gender of 5- and 6-year-olds from their spontaneous speech samples. The identification rates were high, 78% for the boys and 71% for the girls, which suggests that cues in longer speech samples such as variation in vocal pitch, phrasal intonation, stress patterns, or voice breathiness can also influence perceptual differentiation of children's gender.

In the current study, we test the hypothesis that gender information in children's voices is influenced by sociocultural factors including regional dialect. While previous studies have investigated regional dialect variation in American English-speaking children,[6,7,8] there is still a gap in the literature regarding whether dialect features can provide cues to gender, reflecting the regional patterns of sociolinguistic behavior and cultural factors that children may imitate. In particular, speech communities may differ in the speech characteristics of males and females and these behavioral aspects can shape children's way of talking to sound more like a male or a female. Such cues to gender may be more salient in longer speech samples than in single words, reflecting dialect-specific prosodic and emotional acoustic cues, among others.

Here, we are investigating these dialect effects using a relatively comprehensive set of stimuli: single monosyllabic words, read sentences, and spontaneous talks. The selection of speech materials to elicit children's productions can significantly affect listeners' decisions about the children's gender. In the past, investigators have used a variety of stimuli including single words,[1,9] isolated vowels and sentences,[10] isolated words and sentences,[11,12,13] or sentences only.[14] The speech stimuli in these studies varied in the number of vowels, words, and sentences used. A consistent finding was that longer speech samples provided the needed information for more accurate gender identification than did shorter samples. Presumably, in longer stretches of speech, factors other than the acoustic cues in f_0 and formants contribute to identification accuracy.[11] We expect that spontaneous talks will provide the richest set of cues to gender that could be related to factors such as the regional use of prosody, discourse characteristics, or temporal variation in speech. Consequently, we expect the highest identification accuracy of children's gender from talks, followed by read sentences and single words, respectively.

2. METHODS

A. LISTENERS

There were 26 listeners naïve to the experiment. They ranged in age from 18–30 years (M = 21.65, SD = 2.92). All listeners were native speakers of American English, born in the United States, and lived in Ohio. Exclusionary criteria for this study included: a) being born outside of the United States; b) history of speech and/or language disorders; c) being over the age of 30 yrs. Participants were recruited from local communities in Columbus, OH. This study protocol was approved by the Institutional Review Board (IRB) at The Ohio State University.

B. SPEAKERS

Speech samples utilized for this experiment were selected from a larger corpus of children's speech (n = 90) ages 8–11 years, from three distinct dialect regions in the United States: two in the Midwest (Central Ohio and Southeastern Wisconsin) and one in the South (Western North Carolina). Gender of the children was reported by their parents as male (M) or female (F). In terms of their distribution, 30 children (15 M, 15 F) were from Ohio (OH) (*M* = 10.31 yrs., *SD* = 1.30 yrs.), 30 children (15 M, 15 F) from Wisconsin (WI) (*M* = 10.15 yrs., *SD* = 1.21 yrs.), and 30 children (15 M, 15 F) from North Carolina (NC) (*M* = 10.27 yrs., *SD* = 1.31 yrs.). For each dialect, children were matched as closely as possible for chronological age (in years; months) at each of four age levels: 8, 9, 10, and 11.

C. STIMULI

A total of 620 stimuli were used. There were three types of stimuli: single words (n = 360), read sentences (n = 180), and phrases from spontaneous talks (n = 80). Each child contributed four /hVd/-words with /i, u, ɑ, æ/ vowels (representing both high and low, and front and back vowels) and two read sentences. Children from OH did not produce spontaneous talks, and thus the spontaneous speech samples were only available from WI and NC children. The stimuli were pilot tested in a preliminary experiment because previous literature reported generally high accuracy of gender identification in even younger children. Based on this pilot testing, we decided to mask the utterances by using signal-correlated noise to increase task difficulty and minimize the expected ceiling effects. Signal-correlated noise (SCN) is essentially a noise signal, modulated by the amplitude envelope of the original speech signal[15,16]. The signal-correlated procedure introduces noise into the target signal without changing the waveform envelope by reversing the sign of randomly selected speech samples; the percentage of samples whose signs are randomly switched is based on how degraded the investigators want the signal to become. In this study, the percentage of samples whose sign was randomly switched was 30%.

D. PROCEDURE

Listeners completed a questionnaire regarding their language background and demographic information. For the experiment, each listener was seated in a sound-attenuating booth and was facing a computer monitor. After hearing a stimulus over Sennheiser 600 laboratory headphones, one at a time, the listener indicated with a mouse response whether the speaker was male (a boy) or female (a girl). The experiment was executed using a custom program written in MATLAB. The response interface is shown in Figure 1. The listeners completed this one-interval two-alternative forced-choice identification task at their own pace. The stimuli were randomized and presented in five blocks: 2 x 180 words, 1 x 180 sentences, 1 x 32 WI talks, and 1 x 48 NC talks. Listeners were allowed to take breaks between blocks if needed. The experiment lasted about 45 minutes.

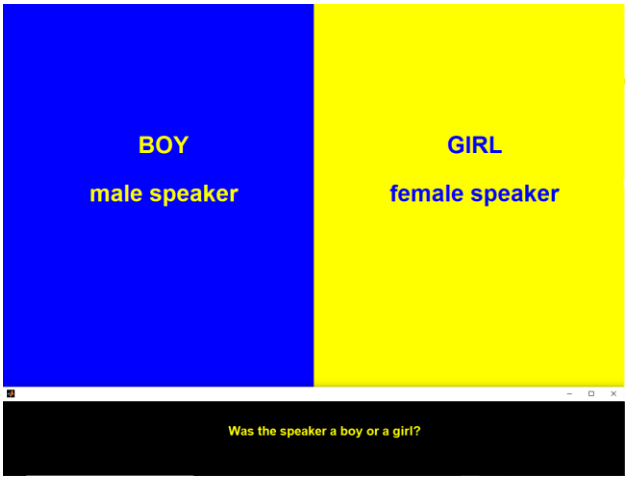


Figure 1. A screenshot of response boxes used in the experiment.

E. DATA ANALYSIS

Identification responses were analyzed using Signal Detection Theory[17,18] to separate sensitivity in gender judgments from response bias. A nonparametric measure of sensitivity (A')[19] was used as the data were not normally distributed. The sensitivity analyses were followed by within-subject ANOVAs and post-hoc t-tests.

3. RESULTS

A. ACCURACY

Overall, gender was identified more accurately from sentences (67% correct) and talks (62%) than single words (57%). Somewhat surprisingly, spontaneous talks provided fewer gender cues than did read sentences, although we note that gender in the talks was identified slightly more accurately for male speakers (65%) than female speakers (59%). Figure 2 shows the accuracy broken down by stimulus type, gender, and age level (8, 9, 10, and 11 years). As can be seen, listeners were consistent in identifying gender from words and sentences: They were more accurate when responding to the younger female speakers (ages 8 and 9) than younger male speakers (Figure 2a, 2b). However, the accuracy results for talks were more variable across age levels. A strikingly different pattern was found for 8-year-olds, where listeners were mostly incorrect in identifying female speakers whereas male speakers were more correctly identified in the talks than in either words or sentences (see Figure 2c).

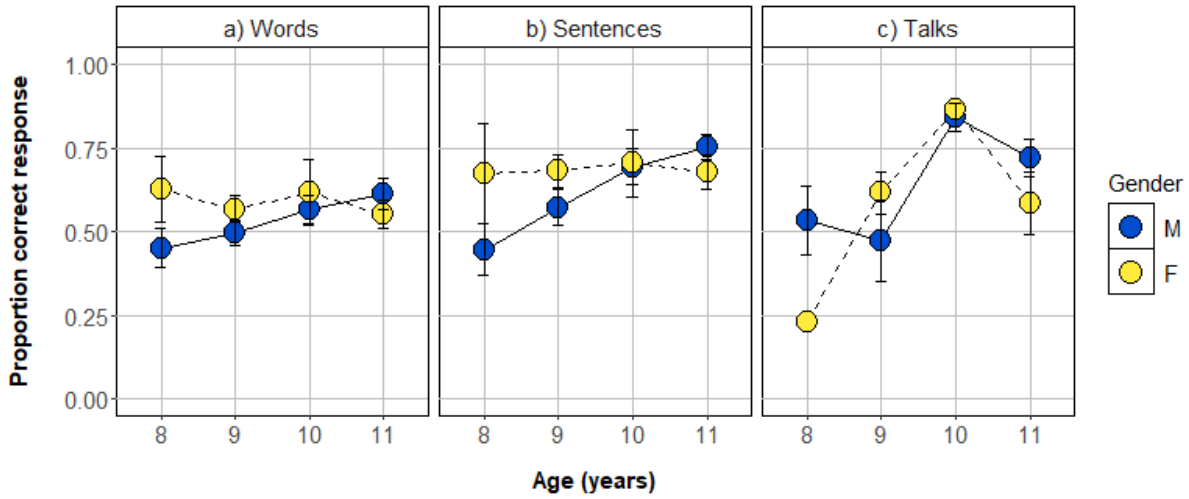


Figure 2. Accuracy by stimulus type, gender, and age level.

B. SENSITIVITY

Sensitivity data (A') were analyzed in two ways. The first ANOVA was used to analyze listeners' sensitivity to gender judgments in words and read sentences only, because these speech samples were obtained from the children in all dialect groups (n = 90). Based on the response patterns in Figures 2a and 2b, the age levels were collapsed into younger (ages 8 and 9) and older (ages 10 and 11) children. In this ANOVA, the within-subject factors were stimulus type (words, sentences), age group (younger, older), and dialect (OH, WI, NC). The second ANOVA was used to assess the differences between sentences and talks, but analyses were conducted for the effects of WI and NC dialects only as no talks were available from OH children.

In the first ANOVA, all three main effects were significant. Listeners' sensitivity in gender judgments was significantly greater for sentences than for words [$F(1, 25) = 117.49$, $p < 0.001$, $\eta_p^2 = 0.825$] and for older than for younger children [$F(1, 25) = 10.01$, $p = 0.004$, $\eta_p^2 = 0.286$]. The significant effects of dialect were comparatively weaker [$F(2, 50) = 5.23$, $p = 0.009$, $\eta_p^2 = 0.173$], and pairwise comparisons indicated that sensitivity was significantly greater only in response to OH speakers (M = 0.70) compared with NC speakers

(M = 0.65); the remaining comparisons were not significant. Importantly, there was a significant interaction between dialect and age group [$F(2, 50) = 19.57$, $p < 0.001$, $\eta_p^2 = 0.439$]. As illustrated in Figure 3, the interaction arose because listeners' sensitivity in gender judgments was significantly greater for younger (as compared to older) OH children and for older (as compared to younger) NC children. Also, within each age group, all comparisons were significant for the younger children, but none were significant for the older children. This indicates that dialect did not contribute significant gender cues for older children, but for the younger children, listeners were most sensitive in gender judgments in response to OH children, followed by WI and NC children, respectively.

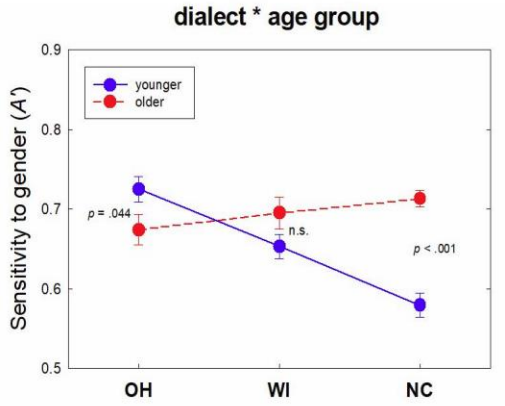


Figure 3. Sensitivity in gender judgments as a function of speakers' age and dialect.

The second significant interaction was between stimulus type and age group [$F(1, 25) = 5.13$, $p = 0.032$, $\eta_p^2 = 0.170$]. As shown in Figure 4, there were no significant differences between age groups for words, but for sentences, older children provided significantly more gender information than younger children.

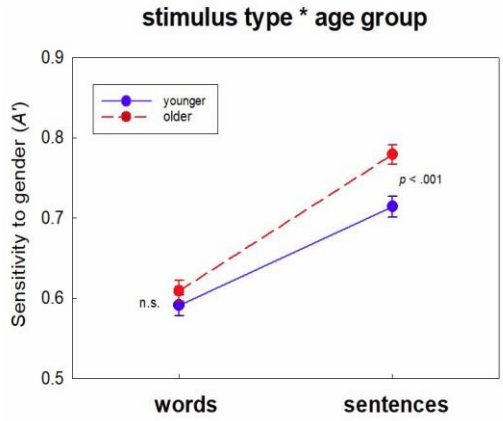


Figure 4. Sensitivity in gender judgments as a function of stimulus type (words, sentences) and speakers' age.

In the second ANOVA that focused only on WI and NC sentences and talks, all three main effects were significant. Listeners' sensitivity in gender judgments was significantly greater for sentences (M = 0.73) than for talks (M = 0.66) [$F(1, 25) = 14.84$, $p < 0.001$, $\eta_p^2 = 0.372$], for older than for younger children [$F(1, 25) = 14.80$, $p < 0.001$, $\eta_p^2 = 0.372$], and for NC speakers than WI speakers [$F(1, 25) = 7.78$, $p = 0.010$, $\eta_p^2 = 0.237$]. However, there was also a significant stimulus type by dialect interaction [$F(1, 25) = 7.36$, $p = 0.012$, $\eta_p^2 = 0.228$]. This interaction, shown in Figure 5, clarified that NC dialect provided more gender information than WI dialect only for talks but not for sentences, where speaker dialect did not influence listeners' sensitivity in gender judgments. Also, WI talks provided fewer cues to gender than WI sentences, whereas gender information in NC dialect was present with similar strength in either sentences or talks.

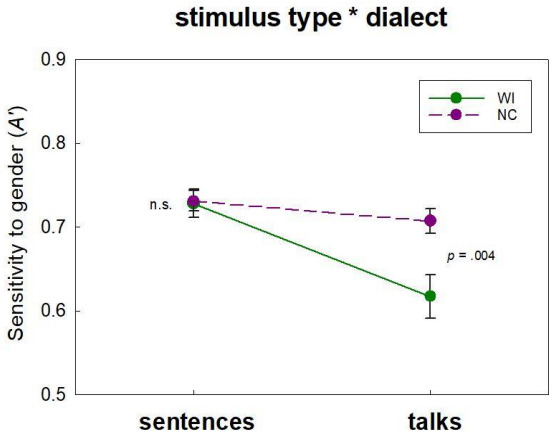


Figure 5. Sensitivity in gender judgments as a function of stimulus type (sentences, talks) and speakers' dialect.

4. DISCUSSION

This study investigated whether sociocultural background and behavioral characteristics of children's speech contribute useful information to the perception of their gender. We hypothesized that regional dialect variations can provide cues to voice gender as speech contains social, prosodic, and behavioral features that are typical of a given community which may influence listeners' identification decisions. Based on the previous studies, we expected that longer speech samples would provide more gender information than single words. In addition, we expected that performative aspects of unconstrained child productions in spontaneous conversations could enhance masculine and feminine voice characteristics, and that listeners would be more accurate in gender identification when listening to spontaneous talks rather than read sentences.

Our hypothesis that regional dialect can influence listeners' judgments about gender was confirmed in two different ways. First, when presented with read utterances (words and sentences), dialect was a significant factor for younger children but not for older children. Presumably, voice characteristics of the latter provided more anatomically related source (f_0) and spectral cues, and these cues became dominant in guiding listeners' identification decisions. However, in the absence of the salient, anatomically based cues in younger children, listeners may have relied on additional behavioral or prosodic features that could be associated with male or female speech. It is possible that the current listeners from Central Ohio were attuned to all possible gradations between phonetic and prosodic features in the regional variety of English spoken in that area and were able to associate a specific combination of features as reflecting male and female speech. Consequently, these listeners could be more accurate in detecting these "composite" cues in the voices of children from Central Ohio than in any other regional variety used in the experiment. Our results support this assertion, showing that listeners were most sensitive in their judgments about gender in response to children from Ohio, followed by Wisconsin and North Carolina, respectively. We can further speculate that listeners were more familiar with and could

associate more gender features of mainstream American English spoken in the Midwest (Ohio and Wisconsin) than in the variety of Southern American English spoken in Western North Carolina. It could be that the perceptual attunement to a particular variety of English that emerges in the first year of life[20] also includes attunement to gender in voices. As adults, listeners may make a greater use of the complex interdependencies of perceptual cues in detecting fine-grained differences in the voices of young boys and girls. This possibility needs to be explored in future research.

Second, when listeners were presented with spontaneous talks samples from only WI and NC children (younger and older), they were more accurate when listening to children from NC. Importantly, when compared with read sentences, the NC talks provided as many gender cues as sentences (see Figure 5) whereas WI talks provided significantly less gender information. This indicates that, in longer stretches of speech, cues in the NC speakers' dialect were overall stronger than in the WI dialect, regardless of whether children were reading sentences or engaged in a conversation. We have no explanation at present why listeners found fewer gender cues in WI talks than in WI sentences, but the current results show that, overall, gender cues in phrases extracted from spontaneous conversations were not stronger than in read sentences. This could be due, in part, to the nature of stimulus material used and the way the sentences were produced. The stress patterns in the read sentences were highly structured and required emphasis of words in the focal positions, which may have provided more acoustic information even in noise-masked speech. Thus, the emphasized words bearing sentential focus could provide comparatively more robust gender cues than what could be detected in more variable spontaneous productions. The availability of gender cues in read versus spontaneous productions awaits further testing in a more focused design.

The current study also contributes a new finding that, although longer speech samples provide more gender information than single words, the gender cues in either stimulus type are distributed differentially as a function of children's age. That is, we found that listeners found fewer gender cues in single words produced by both younger and older children, whereas the difference between the two age groups showed up in read sentences; listeners were more correct when listening to longer speech samples in older than in younger children. This indicates that the expected age effects (i.e., voices of older children are expected to provide more gender cues than younger children) are not always robust and may vary as a function of stimulus type.

Overall, the study found that, on average, gender of children between 8-11 years of age can be identified accurately above chance, even when their speech is masked by noise. However, it needs to be pointed out that the accuracy results obtained here are generally lower than those reported for children of that age[1,11,21] or adults tested under similar stimuli and testing conditions[22] due to the signal degradation, which was not used in previous studies. This by itself is a significant finding, as we anticipated the accuracy for words to fall well below chance. This suggests that cues to gender can be distributed over a wide range of spoken language characteristics, and listeners who are familiar with the dialect of the speakers can detect more of them. As our work continues, the current preliminary report will be expanded by additional analyses of response bias and application of different statistical models to the dataset.

REFERENCES

[1]T. L. Perry, R. N. Ohde, and D. H. Ashmead, "The acoustic bases for gender identification from children's voices," J. Acoust. Soc. Am. **109**(6), 2988–2998 (2001). https://doi.org/10.1121/1.1370525

[2]P. Ladefoged and D. E. Broadbent, "Information conveyed by vowels," J. Acoust. Soc. Am. **29**(1), 98–104 (1957). https://doi.org/10.1121/1.1908694

[3]H. K. Vorperian and R. D. Kent, "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," J. of Speech, Lang. Hear. Res. **50**(6), 1510–1545 (2007). https://doi.org/10.1044/1092-4388(2007/104)

[4]S. Bennett, "Vowel formant frequency characteristics of preadolescent males and females," J. Acoust. Soc. Am. **69**(1), 231–238 (1981). https://doi.org/10.1121/1.385343

[5]B. Weinberg and S. Bennett, "Speaker sex recognition of 5- and 6-year-old children's voices," J. Acoust. Soc. Am. **50**(4), 1210–1213 (1971). https://doi.org/10.1121/1.1912757

[6]E. Dossey, C. G. Clopper, and L. Wagner, "The development of sociolinguistic competence across the lifespan: Three domains of regional dialect perception," Lang. Learn. Dev. **16**(4), 330–350 (2020). https://doi.org/10.1080/15475441.2020.1784736

[7]E. Jacewicz and R. A. Fox, "Perception of local and non-local vowels by adults and children in the South," J. Acoust. Soc. Am. **147**(1), 627–642 (2020). https://doi.org/10.1121/10.0000542

[8]E. Jacewicz, R. A. Fox, and J. Salmons, "Regional dialect variation in the vowel systems of typically developing children," J. Speech Lang. Hear. Res. **54**(2), 448–470 (2011). https://doi.org/10.1044/1092-4388(2010/10-0161)

[9]L. Nagels, E. Gaudrain, D. Vickers, P. Hendriks, and D. Başkent, "Development of voice perception is dissociated across gender cues in school-age children," Sci. Rep. **10**(5074), 1–11 (2020). https://doi.org/10.1038/s41598-020-61732-6

[10]O. Amir, M. Engel, E. Shabtai, and N. Amir, "Identification of children's gender and age by listeners," J. Voice. **26**(3), 313–321 (2012). https://doi.org/10.1016/j.jvoice.2011.06.001

[11]P. F. Assmann and S. Barreda, "Perception of gender in children's voices," J. Acoust. Soc. Am. **150**(5), 3949–3963 (2021). https://doi.org/10.1121/10.0006785

[12]W. T. Fitch and J. Giedd, "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**(3), 1511–1522 (1999). https://doi.org/10.1121/1.427148

[13]S. Bennett and B. Weinberg, "Sexual characteristics of preadolescent childrens' voices," J. Acoust. Soc. Am. **65**(1), 179–189 (1979). https://doi.org/10.1121/1.382234

[14]J. M. Hillenbrand and M. J. Clark, "The role of f_0 and formant frequencies in distinguishing the voices of men and women," Atten. Percept. Psychophys. **71**(5), 1150–1166 (2009). https://doi.org/10.3758/app.71.5.1150

[15]M. H. Davis and I. S. Johnsrude, "Hierarchical processing in spoken language comprehension," J. Neurosci. **23**(1), 3423–3431 (2003). https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003

[16]M. R. Schroeder, "Reference signal for signal quality studies", J. Acoust. Soc. Am. **44**(1), 1735–1736 (1968). https://doi.org/10.1121/1.1911323

[17]D. M. Green and J. A. Swets, *Signal detection theory and psychophysics*, (Wiley, 1966).

[18]N. A. Macmillan and C. D. Creelman, *Detection Theory: A user's guide (2nd Ed.)*, (Psychology Press, 2004). https://doi.org/10.4324/9781410611147

[19]J. G. Snodgrass and J. Corwin, "Pragmatics of measuring recognition memory: Applications to dementia and amnesia," J. Exp. Psychol. Gen. **117**(1), 34–50 (1988). https://doi.org/10.1037//0096-3445.117.1.34

[20]J. F. Werker, "Perceptual beginnings to language acquisition," App. Psycholinguist. **39**, 703–728 (2018). https://doi.org/10.1017/S0142716418000152

[21]E. Sederholm, “Perception of gender in ten-year-old children’s voices,” Logoped. Phoniatr. Vocol. **23**(2) 65–68 (1998). https://doi.org/10.1080/140154398434211

[22]H. Goy, M. K. Pichora-Fuller, and P. van Lieshout, “Effects of age on speech and voice quality ratings,” J. Acoust. Soc. Am. **139**(1), 1648–1659 (2016). https://doi.org/10.1121/1.4945094