Proceedings of The National Conference
On Undergraduate Research (NCUR) 2018
University of Central Oklahoma
Edmond, Oklahoma
April 5-7, 2018

How is your Voice Perceived? A Perceptual Study

Emily Roper
Department of Communication Sciences & Disorders
Texas Woman's University
P.O. Box 425737
Denton, Texas, 76204-5737, U.S.A.

Faculty Advisor: June Levitt, Ph.D., CCC-SLP

Abstract

By listening to the voice, people can approximate the speaker's age, sex, and emotional states, as well as the message being conveyed. Listeners also perceive a variety of information from one's speech. The specific acoustic markers of voice that guide people's perception of speech have not yet been fully investigated. The present investigation was designed to establish acoustic markers that influence the listener's perception of speech produced by various individuals. The goal of the present perceptual study was to identify key acoustic markers of speech signals that were perceived as (1) sex-appropriate, (2) age-appropriate, (3) pleasantness, and (4) smartness. A growing body of research has investigated the features of sex difference and age difference in the field of speech-language pathology. These data serve the needs of particular client populations such as transgender and aging individuals. The general characteristics of pitch, rate of speech, and breathiness of these speakers are reported without specific measurements. Speech samples of five males and five females ranging in age from 21 to 62 years were recorded and used as stimuli. Thirty young college students completed rating the stimuli utilizing a 10-step Likert scale. The perceptual rating scores were contrasted with the acoustic measurements of fundamental frequency, intensity, and duration. They are perceived as pitch, loudness, and length of speech, respectively. Also, the spectral elements were measured and contrasted with the perceptual ratings. As expected, listeners were able to determine sex-appropriateness based on the individual's pitch lower than 150 Hz as males, and higher than 168 Hz for females. Age appropriateness was expected to be related to the rate of speech. However, the results did not show any relationship between the two aspects. The results showed a significant correlation between the perception of the pleasantness and smartness.

Keywords: Voice, Perceived, Listening

1. Introduction

According to Kent,[10] verbal communication carries multiple types of information, including personal, phonetic, emotional, and transmittal information. By listening to the voice of the speaker, people can approximate the speaker's age and sex (personal information). A listener can understand the speaker's message by focusing on the contents (phonetic information). The overall characteristics of one's voice represent his/her emotional state (emotional information), and the surrounding noises let us know the environment in which the verbal message is produced (transmittal information). Voices contribute to the first impression of others, including approximate age, gender, competence, and relationships.[6] According to McAleer, Todorov, and Belin,[9] perceived voice characteristics could influence job selection, friend selection, and other social interactions. Klofstad[2] conducted a research project to see how voice perception affects the choice of a candidate for leadership positions and found that voters prefer lower-pitched voices due to the fact that they are perceived as stronger and being more competent.

A growing body of research has explored different patterns of speech perception by various attributes. It is important for the students studying communication science and disorders to understand the relationship between acoustic markers and speech perception. With this knowledge, therapists can guide clients with communication disorders or differences to work on the restoration of functional verbal communication in the clients' respective communities. Therefore, voice improvement is the goal of various disciplines, such as singers, transgender voice modification, voice over, politicians, and public speakers. Two of the critical evaluation methods of speech/voice are acoustical analyses and perceptual judgment. First, acoustical analyses are important to Speech-Language Pathologists (SLPs) working on voice because they provide objective information about the pre- and post-therapy changes. Second, perceptual studies are critical to validate the effects of the information identified by the acoustical analyses.[7]

The goal of the present perceptual study was to identify key acoustic markers of speech signals that are perceived as (1) sex-appropriate, (2) age-appropriate, (3) pleasantness, and (4) smartness. The author predicted that the speakers' pitch would be the main cue to indicate sex. The speaker's rate of speech measured by the number of words per minute would indicate age. Finally, it was expected that the speaker's perceived pleasantness and smartness would hinge on the intensity of the speaker's voice.

2. Method

2.1. Participants

2.1.1. participant speakers

Ten adult residents of Texas (M_{age} = 40.5 years; *SD* = 19.1; *range* = 21-68 years) served as participant speakers for the current study: five were males and five were females. They are monolingual speakers of General American English who do not have any major health problems. Their vision and hearing are within normal limits with or without corrections as self-reported by participants. All of them do not have history of speech or language disorders.

2.1.2. participant listeners

Thirty students (3 males and 27 females) of Texas Woman's University participated in the present study. The age range was 21 to 39, and the detailed information about the specific age was not collected. The hearing acuity of the participants was within normal limits, as reported by the participants. Because this study focused on the perception of the individuals who carry on functional communication in college, the information on history and background of speech and language disorders or differences were not collected.

2.2. Materials

Speech stimuli included recordings of participant speakers reading the first half of the reading of the "North Wind and The Sun" passage from the Aesop's fables. Following the Handbook of the International Phonetic Association[4], this passage was used in the present study.

Table 1: Stimuli

Sex	Age	Voice Identifier
Female	26	Voice 1
	27	Voice 2
	28	Voice 3
	61	Voice 4
	68	Voice 5
Male	21	Voice 6
	23	Voice 7
	31	Voice 8
	58	Voice 9
	62	Voice 10

2.3. Procedures

The data collection was carried out in a quiet room on the TWU Denton campus. A response sheet was distributed to each participant. Participants were requested to rate the recorded voices in 10-point Likert scale. There were four rating categories, including the sex indicator, age-indicator, pleasantness, and smartness. No specific definition of "pleasantness" or "smartness" was provided to the listeners. This is because the researcher planned to conduct acoustic analyses later to figure out the contributing elements to the listener's perception of these four aspects. Before rating the experimental stimuli, participants practiced with two recorded voice materials that were not included in the experimental stimuli. After familiarizing with the rating procedures through the practice, the participant listeners rated the experimental stimuli one at a time. The order of presentation of the ten recorded voices was randomized for each session to prevent systematic bias.[6] Each stimulus was played two times.

2.4. Data Analyses

Participant's responses were organized in an Excel file for each of the four rating categories (i.e., sex, age, pleasantness, and smartness). Each category included 300 ratings (30 listeners x 10 perceptual ratings). The rating scores out of 2-standard deviation were identified as outliers. These outliers were excluded from the data analyses. Twenty-one ratings (7.0%) were eliminated from the sex rating category, 10 ratings (3.3%) were eliminated from each of the age and pleasantness ratings, and nine ratings (3.0%) were eliminated from the smartness rating category.

After eliminating the outliers, data analyses were carried out with IBM SPSS Statistics 24 (v.24.0.0.0). The significance level was set at $p < .05$. First, the correlations between the perceptual categories were analyzed with all possible combinations to examine the relationship between the perceptual categories of sex, age, pleasantness, and smartness. Second, the rating in each category was contrasted with acoustic characteristics of the stimuli (i.e., the pre-recorded voices from 10 speakers) to identify the relationship between speech perception and production. Pearson's correlation analyses were used in both analyses. PRAAT (v.6.0.08)[8] was used for all acoustic analyses.

3. Results

3.1. Relationship Between Four Perceptual Ratings

Before performing statistical analyses, the relationship between four perceptual measures was visually inspected as a graph (see Figure 1). On the response form, the age figures indicated 20 years old (represented as "0") to 70 years or older (represented as "10") to fit the 10-step Likert scale. The obtained ratings of age were converted to the scale of 10 to contrast with three other 10-scale rating categories. The normality of data was examined using the Shapiro-Wilks test, which showed the "sex" rating data were not normally distributed. The transformation of data was determined not to be applicable because the sex data show a binary distribution (i.e., male or female). Therefore, Pearson correlation coefficients were used to examine the relationships between three perceptual rating categories, including age, pleasantness, and smartness.

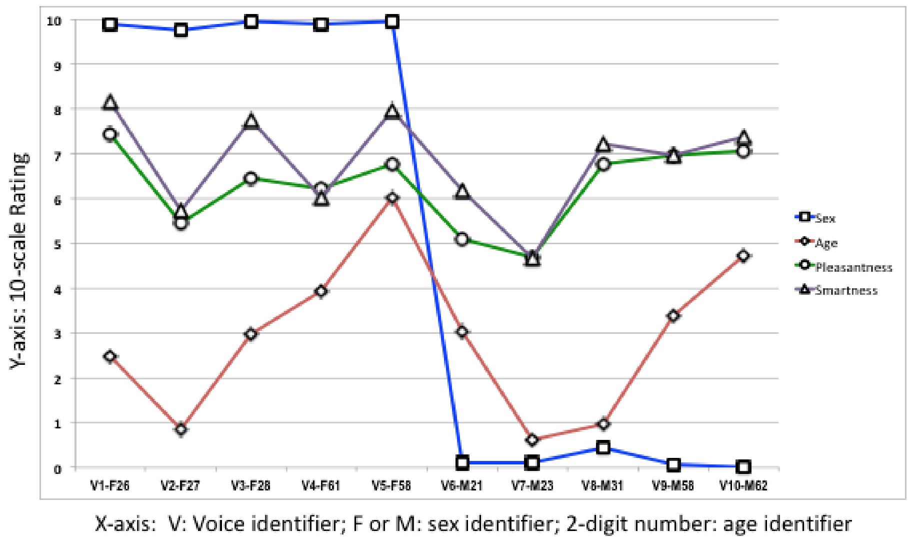

Figure 1. Relationship between four perceptual ratings

3.1.1. "age" vs. "pleasantness" ratings

The Pearson correlation coefficient analysis identified a moderate positive correlation between the ratings of age and pleasantness, $r = .484$, $p = .156$, *ns*. The coefficient of determination (r^2) was computed to explore the proportion of variance in one variable that is explained by the other variable. The r^2 was .234. Thus, the perception of age can explain 23.4% of the variability in the perception ratings of the pleasantness.

3.1.2. "age" vs. "smartness" ratings

The Pearson correlation coefficient analysis showed a medium to moderate positive correlation between the ratings of age and smartness, $r = .543$, $p = .104$, *ns*. The coefficient of determination (r^2) was .295, suggesting that the perception of age explained 29.5% of the variability in the perception ratings of the smartness.

3.1.3. "pleasantness" vs. "smartness" ratings

The Pearson correlation coefficient analysis revealed a strong positive correlation between the ratings of pleasantness and smartness, $r = .881$, $p<.01$. The correlation of these two perceptual ratings is statistically significant ($p<.01$). The coefficient of determination (r^2) was .776. Thus, the perception of pleasantness explained 77.6% of the variability in the perception ratings of the smartness.

3.2. Relationship Between Perceptual Ratings And Acoustic Measurements

Each perceptual rating was contrasted with three acoustic measurements, including fundamental frequency (F0), intensity, and duration. The acoustic measurement of F0 represents the rate at which vocal folds vibrate every second. Approximately, the mean F0 of adult males can range from 100 to 150 Hertz (Hz), and for adult females it can range from 180 to 250 Hz.[3] The perceptual correlate to the F0 is pitch. As such, a higher pitch tends to sound more 'female' (or, feminine), and a lower pitch tends to sound more 'male' (or, masculine). The intensity of speech refers to the power carried by sound waves, and it is related to the amplitude of the vibration. Its perceptual correlate is loudness. The duration refers to the rate of speech, and its perceptual correlate is length. The normality of the acoustic measurements was examined prior to the statistical analyses. A series of Shapiro-Wilks tests confirmed that all of the acoustic measurements were normally distributed (p>.05).

Table 2: Acoustic features of the voice recordings

Voice ID	Age	Speaker's Sex	F0 (Hz)	Intensity (dB SPL)	Duration (ms)
V1	26	F	202.79	69.45	788.27
V2	27	F	168.14	66.76	830.00
V3	28	F	206.76	69.89	1450.00
V4	61	F	216.83	73.88	750.00
V5	68	F	179.69	71.77	870.00
V6	21	M	110.34	73.39	760.00
V7	23	M	94.02	70.14	930.00
V8	31	M	128.30	69.61	720.00
V9	58	M	131.20	67.23	650.00
V10	62	M	113.16	65.62	870.00

3.2.1. sex rating and acoustic measurements

The sex rating data show different perceptions of male and female voices within the given F0 ranges. Specifically, the perceptual rating data identified voices lower than 150 Hz as 'male' and voices higher than 168 Hz as 'female'. This is closer to the reported F0 ranges of males and female by Roseberry-McKibbin and Hedge.[3] The ratings showed a non-linear relationship between the sex indicator of voice without ambiguity. Therefore, a Spearman's rank-order correlation test was conducted to assess the relationship between the perception of sex ratings and the mean F0. The result of the test revealed a statistically significant positive correlation between the perception of sex ratings and F0 average ($r_s(8) = .774$, $p = .01$).

The sex perception data as well as the intensity and duration of speech measurement show noticeable overlaps between males and females, suggesting no systematic relationship between these data sets. Therefore, the statistical analyses were not performed with these combinations.

3.2.2. age rating and acoustic measurements

Published data explain that the female F0 tends to decrease with age due to hormonal changes and swelling of the vocal folds, whereas the male F0 tends to increase with age due to the vocal folds thinning.[5] Thus, the differences of F0 become smaller as males and females age. Before performing further analyses, the age related differences between males and females were inspected. With respect to the F0 changing patterns over age between males and females, no distinct difference was identified with the present data. Therefore, the male and female data were analyzed together.

The Pearson correlation coefficient revealed a small positive correlation between age and F0 data, $r = .273$, $p = .445$, *ns*. The coefficient of determination (r^2) was 0.074. A similar result was obtained with the correlation of the age perception and vocal intensity, $r = .184$, $p = .612$, n*s*. The coefficients of determination (r^2) was 0.034. Taken together, the analyses suggest that the perception of age explained 7.4% of the variability in the F0 and 3.4% of the variability in the vocal intensity. There were noticeable overlaps between the age perception and duration of speech data. Therefore, correlation analyses were not rendered in this combination.

3.2.3. pleasantness rating and acoustic measurements

The pleasantness rating data and acoustic measurements of F0 and intensity were sparse, suggesting no systematic relationship between these two data sets. Therefore, the statistical analyses were not performed. The correlation between the pleasantness and duration of speech was analyzed. The Pearson correlation coefficient showed a medium to moderate negative correlation between the ratings of pleasantness and duration, $r = -.366$, $p =.333$, *ns.* The pleasantness rating was higher for slower speakers, however, the relationship was not significant. The coefficient of determination (r^2 = .134) suggests that the perception of pleasantness explained 13.4% of the variability in the rate of speech.

3.2.4. smartness rating and acoustic measurements

The intensity of speech measurement did not yield any relationship with the perception of smartness by inspecting the data visually. Therefore, the smartness ratings were analyzed with respect to the relationship between F0 and duration of speech, respectively. The Pearson correlation coefficient showed a medium to moderate correlation between the smartness perception and F0, r = .439, p = .204, *ns*, suggesting that higher pitch voices are perceived as 'smarter' than those with a lower pitch. However, the degree of correlation was not statistically significant. The coefficient of determination (r^2= .193) implies that the perception of smartness statistically explained only 19.3% of the variability in the F0. The correlation analysis revealed a small negative correlation between the ratings of smartness and duration, r = -.253, p =.512, *ns*, suggesting that the smartness rating was higher for the faster speakers than the slower ones. The correlationship is not statistically significant. The coefficient of determination was .064, which implies that the perception of smartness statistically explained 6.4% of the variability in the rate of speech.

4. Discussion

The goal of the study was to identify what acoustical features impact a listener's perception of recorded voice. The most intriguing finding of the study was the relationship between the perceptions of pleasantness and smartness. The smarter a voice sounded, the more pleasant it was rated. These findings are consistent with McAleer's research.[9] The perceptions of pleasantness and smartness do not have a statistically significant relationship with any of the other factors, such as the perception of age, sex, or any acoustical marker. This relationship may suggest that a listener judges pleasantness and smartness in conjunction with, but separate from, other characteristics of the voice. However, Pearson's correlation analyses tend to yield statistical significance with relatively moderate correlation, with large sample sizes (n =100+), while they do not reach significant levels with small sample sizes.[12] Thus, the present study needs to be replicated with a greater number of participants. Future research must seek to identify characteristics, such as confidence levels or attractiveness, that potentially impact the observed relationship.

As predicted, the F0 was the main cue for sex identification. The other factors did not contribute to the identification of sex. It is intuitive, however, that higher F0 requires a faster rate of vocal folds vibration as well as vocal fold lengthening. Thus, a higher level of tension is required to increase F0. Transgender individuals who transition from male to female often seek speech-language pathology services to develop their desired voices. If one focuses only on the increase of F0, it could produce excessive tension on the vocal folds, resulting the development of Muscle Tension Dysphonia (MTD). To prevent such functional voice disorders, further analyses are necessary to identify other reliable indicators of male and female voices, such as resonance, pragmatics, and various suprasegmental features.

The present findings suggest that duration of speech does not statistically impact the listener's age perception, which is not consistent with Duchin and Mysak's research.[11] They found that the rate of speech computed from the duration is affected by the perception of age. In the present study, the ages of the Voice 1 (26 years of age) and Voice 7 (23 years of age) were most accurately identified. The rates of speech of these participants were .79 and .93 words per second, respectively. It suggests that the rate of speech alone is not an important cue for age. With the growing number of aging population and increasing number of undesirable scams by telemarketers targeting older individuals, it is helpful if the voice is manipulated to mask the advanced age. Future studies with a greater number of participants should be conducted to further identify the cues of aging voice.

5. Conclusions and future directions

Future research will seek to replicate the present study with a more diverse group of individuals (i.e., young vs. old raters and male vs. female raters). A more diverse group of individuals would be useful to see if different sexes or ages perceive voices differently. Other research by Dilley[5] has suggested that children and adults perceive voices differently. Therefore, research should be conducted to examine whether raters aged 20-30 and raters aged 50-60 groups show different ratings. Further investigation should be completed to evaluate whether perceived "pleasantness" and "smartness" are correlated equally by the listeners of all ages.

The data procured from the present study contributes to the growing body of research in voice perception analysis. These data have applications in other fields beyond speech and language pathology, such as teaching, politics, and business. The understanding of how one's voice may be perceived by others can impact voice therapy, speeches, and advertisements.

6. Acknowledgment:

The present investigation was supported by the Texas Woman's University Quality Enhancement Plan (QEP) of 2017. The author wishes to express her appreciation to Dr. Kim Miloch and Ms. Kaye Garrison of the QEP office for their guidance, and Dr. Donna Tilley for her encouragement to make this project possible.

7. Financial and Non-financial Disclosures:

The present research was funded by the Texas Woman's University Quality Enhancement Plan (QEP): QEP Student Scholar program. The authors have no non-financial relationships to disclose.

8. References:

1. Alejandro R. Jadad, Andrew Moore, Dawn Carroll, Crispin Jenkinson, John M. Reynolds, David J. Gavaghan, and Henry J. McQuay. "Assessing the quality of reports of randomized clinical trials: is blinding necessary?." *Controlled clinical trials* 17, no. 1 (1996): 1-12.
2. Casey A. Klofstad, Rindy C. Anderson, and Stephen Nowicki. "Perceptions of competence, strength, and age influence voters to select leaders with lower-pitched voices." *PloS one*10, no. 8 (2015): e0133779.
3. Celeste Roseberry-McKibbin, and Mahabalagiri N. Hegde. *An advanced review of speech-language pathology: Preparation for praxis and comprehensive examination*. PRO-ED, Inc. 8700 Shoal Creek Blvd, Austin, TX 78757, 2006.
4. Donald, M. Decker Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University Press, 1999, pp. 41-156.
5. Elaine T. Stathopoulos, Jessica E. Huber, and Joan E. Sussman. "Changes in acoustic characteristics of the voice across the life span: measures from individuals 4–93 years of age." *Journal of Speech, Language, and Hearing Research*54, no. 4 (2011): 1011-1021.
6. Huiwen, M. Goy and Kathleen Pichora-Fuller, and Pascal van Lieshout. "Effects of age on speech and voice quality ratings." *The Journal of the Acoustical Society of America* 139, no. 4 (2016): 1648-1659.
7. Laura C. Dilley, Elizabeth A. Wieland, Jessica L. Gamache, J. Devin McAuley, and Melissa A. Redford. "Age-related changes to spectral voice characteristics affect judgments of prosodic, segmental, and talker attributes for child and adult speech." *Journal of Speech, Language, and Hearing Research* 56, no. 1 (2013): 159-177.
8. Paul Boersma. "Praat: doing phonetics by computer." *http://www. praat. org/* (2006).
9. Phil McAleer, Alexander Todorov, and Pascal Belin. "How do you say 'Hello'? Personality impressions from brief novel voices." *PloS one* 9, no. 3 (2014): e90779.
10. Raymond D. Kent. *The speech sciences*. Singular Publishing Group, (1997).
11. Sandra W. Duchin, and Edward D. Mysak. "Disfluency and rate characteristics of young adult, middle-aged, and older males." *Journal of communication disorders* 20, no. 3 (1987): 245-257.
12. Julie Pallant. SPSS Survival Manual 4th edition. McGraw Hill (2010).