The Effect of Levels and Types of Experience on Judgment of Synthesized Voice Quality

***Jessica L. Sofranko and †Robert A. Prosek,** **Albany, New York and †State College, Pennsylvania*

Summary: Objectives/Hypothesis. The purpose of this study was to determine the effect of level and type of experience on the judgment of voice quality and to examine the correlation among acoustical measurements and perceptions of voice quality.

Study Design. This was a within-subjects group design.

Methods. Speech-language pathologists, singing voice teachers (SVTs), speech-language pathology graduate students with and without experience with a voice client, graduate students who have completed a voice pedagogy course, and inexperienced listeners (IEs) rated stimuli with systematically altered measurements of jitter, shimmer, and noise-to-harmonics ratio (NHR) on a visual analog scale ranging from mild to severe for overall severity, roughness, breathiness, strain, and pitch.

Results. Results showed that the type of experience had an impact on judgments of voice quality more than the level of experience. Also, jitter/shimmer combination stimuli and shimmer only stimuli frequently correlated with the ratings of overall severity, roughness, and strain, and NHR stimuli correlated with ratings of breathiness across all groups. Only IEs, SVTs, and their students had significant correlations for ratings of pitch with jitter/shimmer combination stimuli having the highest correlations.

Conclusion. The conclusion was that the level and type of experience affect judgments of voice quality.

Key Words: Voice perception–Experienced listener–Listener agreement–Acoustical measures–Synthesized stimuli.

INTRODUCTION

Perception of communication is not only solely based on the signal itself but also by the knowledge of the listener.[1] When studying sound quality, listeners judge what they hear based on several attributes of interest specific to them.[2] Internal standards, used for perceptual ratings, are defined as an individual's baseline for judgment, which is affected by his or her experience, memory, and/or attention.[3,4] These internal standards, or individual differences, are unstable and can lead to disagreement, a specific sensitivity, or bias among listeners.[5,6]

The effect of internal standards on perceptions of voice quality is well documented.[5,7–11] Specifically, with experience affecting the internal standards of listeners, group selection for perceptual experiments is essential. Expert listeners are said to have access to a larger collection of information related to voice quality perceptions when listening to and judging voice quality as compared with inexperienced listeners (IEs) with a varying educational background and a lack of consistent experience with dysphonia.[12]

Researchers typically create their own operational definitions of an expert or "experienced listener" ranging from job title only (eg, speech-language pathologist [SLP] and phonetician) to someone with 2 years of experience in the area of dysphonia.[13,14] These experienced listeners, or generally those who are thought to have extensive background in the area of voice/voice disorders, not only use different components within the acoustical signal as compared with IEs[14] but also use additional resources within the signal.[14,15]

Most studies examining perceptions of voice quality have used only the following judges: speech-language pathology graduate students; listener groups containing both speech-language pathology graduate students and more experienced listeners such as SLPs experienced with dysphonia and/or otolaryngologists; or experienced listeners with different types of experience such as otolaryngologists, phoneticians, and SLPs.[3,4,9,10,12–14,16–23] Also, many studies did not have a control group. Although speech-language pathology graduate students may be easily accessible, there is evidence that they differ in voice ratings as compared with SLPs with experience in voice disorders. Voice quality ratings of vocal creak, instability, deviation, and hypo/hyperfunctionality were significantly different between SLPs with 2 years of experience and speech-language pathology graduate students.[24] Given the evidence that 2 years of experience leads to differences in judgment, speech-language pathology graduate students should not be considered experienced listeners in the area of voice. Past studies concluded that perceptions of voice are unreliable[8,10,21]; however, perhaps results are due to a true difference in experience, or variability in background, which has affected internal standards.

Results of a recent study indicated that when defining groups using specific definitions of experience, there were significant differences in agreement between experienced listeners and IEs.[25] Experienced listeners included SLPs and singing voice teachers (SVTs). The SLPs were individuals with more than 3 years of experience with voice, spending 10 hours or more in the area of voice each week; and SVTs were individuals who were full members of the National Association of Teachers of Singing (NATS). The IEs were individuals from various backgrounds, with no previous experience in the area of voice including singing training and/or previous voice treatment. Overall, interrater agreement was significantly better for the experienced listener groups as compared with the IEs.

Accepted for publication June 4, 2013.

From the *College of St. Rose, Albany, New York; and the †Pennsylvania State University, State College, Pennsylvania.

Address correspondence and reprint requests to Jessica L. Sofranko, The College of St. Rose, 432 Western Avenue, Albany, NY 12203. E-mail: sofrankj@strose.edu

Journal of Voice, Vol. 28, No. 1, pp. 24-35

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.06.001

In addition to the level and type of experience affecting listener judgments, several decades of research in the area of voice perception show that the type of rating scale, the type of stimuli, and the length of stimuli can also have an effect on the listener.[4,11,12,15,17,20,21,23,24,26–29] In fact, varying internal standards (including level and type of experience), difficulty isolating targeted attribute(s) in a multidimensional signal, the type of scale being used to rate voice quality, and the degree of the characteristic that is being measured account for 84.2% of the variance for listener agreement.[20]

The identified variables not only affect perceptual judgments of voice quality but also correlations with acoustical measurements of voice.[4,11,12,15,17,20,21,23,24,26–31] Acoustical measurements provide more objective information during evaluation and treatment and confirm perceptions of voice quality,[9] refining initial perceptions.[32,33] Experience not only affects one's perceptions of voice quality but also one's capacity to accurately pair subjective and objective measures of voice.[33] Research studies determining which acoustical measurements correlate with which perceptual judgments of voice quality have been contradictory. Overall, there is a well-documented relationship between perturbation measurements and disordered voice[34–40] and a well-documented relationship between noise components and disordered voice.[29,41,42] However, agreement regarding specifics has been limited.[29,34–42]

With little agreement among authors, the exact relationship between acoustical measures and perceptions of voice quality remains unknown.[43] This may be a result of questionable group selection for experienced listeners and an inconsistent use of various types of rating scales. The manner in which listeners are asked to rate the signal can also affect the reliability of judgments.[20,23] Research shows that test-retest agreement was significantly higher for ratings using continuous scales versus equal-appearing interval (EAI) scales.[20] Also, continuous scales, direct magnitude estimation, and visual analog scaling (VAS) do not assume linearity of voice quality perceptions.[11] In fact, when listeners were using both VAS and EAI to rate the same voice qualities, results indicated that judgments made by the listeners using EAI were skewed and both scales were only moderately correlated.[44]

Lastly, the type of stimuli presented during the rating task may also be affecting perceptions of voice quality and their relationship to objective measures. Acoustical measurements, specifically perturbation measurements and noise-to-harmonics ratio (NHR), are highly correlated with one another[27] giving evidence of the limitations of using natural voices as stimuli. These correlations make a relationship between a single acoustical measure and a single perception of voice quality impossible to find because a rise in any one acoustical measure may be due to the presence of another. By systematically controlling one acoustical aspect of the signal at a time, through the use of synthesized stimuli, a researcher can examine possible correlations between perceptions of voice quality and objective measurements of voice.[27,45]

Although there have been many studies in the area of perceptions of voice quality, the reviewed variables can affect perceptual judgments of voice quality, and then in turn, affect correlations with acoustical measurements of voice. Despite our knowledge of these factors affecting the perceptions of voice quality, there are very few studies controlling for all of the above-mentioned variables simultaneously. Most importantly, there are very few studies that address the differences between experienced and IEs for perceptions of voice quality.[12]

The effect of experience on listener perceptions of voice quality needs to be determined for proper group selection before experimentation. All known variables that affect the perceptions of voice quality need to be controlled to determine if the differences in level and type of experience have an impact on listener judgments. After the effect of experience is known, then researchers can appropriately choose listeners for studies aimed to examine relationships among acoustical measures and perceptions of voice quality.

The purpose of this study was to determine the effect of experience on the judgment of voice quality and to examine the correlation among acoustical measurements and perceptions of voice quality among those listeners.

METHODS

Stimuli

One sample of sustained vowel /ɑ/ with normal voice quality obtained from a female, aged 23 years, was synthesized using the University of California, Los Angeles (UCLA) synthesizer.[46] This sample, originally recorded at the University of Utah, was chosen because of its widespread use in other studies as an anchor to control for internal standards. Also, the sample was judged to be "normal" by SLPs who have experience in the area of voice and voice disorders on the basis of quality, pitch, and loudness.[16,47–49]

Using the UCLA voice synthesizer,[46] this voice sample was synthesized with a duration of 1 second and a constant fundamental frequency and amplitude. The newly synthesized file was systematically altered by changing measurements of jitter, shimmer, and NHR to create two sets of stimuli. The first set of stimuli included variations of jitter and shimmer simultaneously in five evenly spaced intervals resulting in 25 stimuli. Jitter was altered in increments of 0.75 μs (0–3 μs) and shimmer was altered in increments of 0.5 dB (0–2 dB). The second set of stimuli included a variation of NHR in 10 evenly spaced intervals resulting in 10 stimuli (−50 to 0 dB). The NHR was altered in increments of 5 dB.

Combining jitter/shimmer stimuli and NHR stimuli resulted in 35 total stimuli. Jitter, shimmer, and NHR combination stimuli were not generated for this study in an effort to control for fatigue during the experiment. Aperiodicity and additive noise components were altered separately to significantly reduce the number of stimuli from 250 samples to 35 samples.

Listeners

There were six groups with 10 listeners in each group ($n = 60$). The groups consisted SLPs, SVTs, speech-language pathology graduate students who had completed a voice disorders course and had not had a voice client (SLPGRADs), speech-language pathology graduate students who had completed a voice