The Effect of Experience on Response Time When Judging Synthesized Voice Quality

*Jessica Sofranko Kisenwether and †Robert A. Prosek, *Dallas and †University Park, Pennsylvania

Summary: Objectives/hypothesis. The purpose of this study was to determine the effect of level and type of experience on response time and the number of replays needed when judging voice quality.

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

Methods. Speech-language pathologists, singing voice teachers, 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 (n = 60) 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. Response time (in seconds) and number of replays were recorded during the experiment.

Results. Results showed that experienced listeners took the most time when rating the stimuli. Stimuli with two altered acoustical components also yielded longer response times compared with the stimuli with one altered acoustical component. Finally, level and type of experience had some effect on the number of replays for each stimulus during the rating task.

Conclusions. In conclusion, experience does affect response time when judging voice quality and the number of replays during voice quality rating tasks. Continued research is needed regarding the reasons for extended time and replays as per experience so as to enhance future training protocols.

Key Words: Voice perception–Experienced listener–Response time–Synthesized stimuli.

INTRODUCTION

When controlling for stimulus length and type as well as rating scale, experience has been shown to affect judgments of voice quality.[1,2] A large body of research exists discussing accuracy and agreement among listeners when perceiving voice quality.[3–16] Although experience can affect those judgments, agreement remains moderate[3–6,17] which is often said to be a result of a multidimensional signal in which listeners are using many underlying variables to make their decisions. Although there is detailed information regarding the voice signal, with evidence of relation to specific acoustical measures as well as a list of the factors possibly affecting those perceptions, there are currently no studies that examine the time it takes to make said judgments. One may assume that because experience has been found to affect perceptions of voice quality, those listeners with extensive training or exposure may be more able to focus on the underlying variables of the signal. The question remains as to whether this results in longer response times to judge voice quality? Or whether increased levels of experience result in a faster response time? Finally, does experience have an effect on the number of times listeners need to replay a signal before making a judgment? The answers to these questions are unknown and when obtained may assist in developing appropriate training protocols for judgments of voice quality during assessment and treatment. For instance, if experienced listeners take less time than inexperienced listeners (IEs) when judging voice quality, student clinicians taking a lengthy amount of time during the task may indicate a need for continued training before a level of independence could be considered.

The purpose of this study was to determine the effect of level and type of experience on the response time and the number of replays needed when judging voice quality.

METHODS

Stimuli

The same stimuli used in Sofranko and Prosek (2013) were used for this study. One sample of sustained vowel /ɑ/ with normal voice quality obtained from a woman, age 23, was synthesized using the University of California, Los Angeles (UCLA) synthesizer.[18] The sample was judged to be "normal" by speech-language pathologists (SLPs) who have experience in the area of voice and voice disorders on the basis of quality, pitch, and loudness.[1,2,19–22] The sample was also used in many previous studies as an example of "normal" voice quality.[1,2,19–22]

This voice sample was synthesized using the UCLA voice synthesizer,[18] with a duration of 1 second and a constant fundamental frequency and amplitude. This sample was then systematically altered by changing measurements of jitter, shimmer, and NHR creating 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). 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. Aperiodicity and additive noise components were

Accepted for publication May 29, 2015.

From the *Department of Speech Language Pathology, Misericordia University, Dallas, Pennsylvania; and the †Department of Communication Sciences and Disorders, The Pennsylvania State University, University Park, Pennsylvania 16802.

Address correspondence and reprint requests to Jessica Sofranko Kisenwether, Department of Speech Language Pathology, Misericordia University, 301 Lake St., Dallas, PA 18612. E-mail: jkisenwe@misericordia.edu

Journal of Voice, Vol. 30, No. 4, pp. 394-397
0892-1997/$36.00

© 2016 The Voice Foundation

http://dx.doi.org/10.1016/j.jvoice.2015.05.017

altered separately to significantly reduce the number of stimuli from 250 samples to 35 samples.

Listeners

The same listeners used in Sofranko and Prosek (2013) were used for this study. There were six groups with 10 listeners in each group (n = 60). Groups consisted of SLPs, singing voice teachers (SVTs), speech-language pathology graduate students who had completed a voice disorder course and had not had a voice client (SLPGRADs), speech-language pathology graduate students who had completed a voice disorder course and had treated one or more voice clients (SLPGRADVs), graduate students in the music department who had completed a voice pedagogy course (SVTGRADs), and IEs.

Group 1 consisted of seven women and three men who were American Speech Language Hearing Association certified and state-licensed SLPs. Ages ranged from 29 to 67 years (mean [M], 45.7; standard deviation [SD], 12.92). They had a range of 5–35 years of experience in voice disorders (M, 19; SD, 11.01) and spent 10–40 hours/week treating voice disorders (M, 23.4; SD, 12.21).

Group 2 consisted of eight women and two men, ages ranging from 48 to 69 years (M, 59.6; SD, 6) who were tenured singing voice faculty and full members of the National Association of Teachers of Singing (NATS). Individuals holding a full membership in NATS, with either a Master Degree or Doctor of Musical Arts, teach an average of six or more singing voice students weekly and have ≥2 years of experience.[23] The criterion of tenure implies at least 6 years of full-time faculty work in which the individual mentors undergraduate and graduate students throughout their academic degree of study.

Group 3 consisted of 10 women, ages ranging from 21 to 24 years (M, 22; SD, 0.943), who were current graduate students in a speech-language pathology program and had completed a voice disorder course. Group 4, although similar, consisted of 10 women, ages ranging from 21 to 42 years (M, 26.1; SD, 6.33), who were also current graduate students in a speech-language pathology program, had completed a voice disorder course, but these students had also had one or more voice client(s) in clinic. Students had a range of one to eight voice clients in their clinical experience (M, 2.5; SD, 2.321).

Group 5 consisted of six women and four men with ages ranging from 22 to 46 years (M, 27.9; SD, 7.4). These individuals were current graduate students in either voice pedagogy or vocal performance who had completed a voice pedagogy course in their graduate work. The students taught a range of 1–20 singing voice students weekly (M, 5.6; SD, 5.48).

Finally, group 6 consisted of five women and five men, ages ranging from 24 to 56 years (M, 35; SD, 12.18), with no previous training in voice and/or voice disorders, including singing lessons and voice treatment. This group included individuals from various backgrounds including nursing, real estate, chemistry, culinary arts, fashion, architecture, cosmetology, engineering (mechanical and electrical), and law. All participants in all groups reported no history of a hearing loss, a language disorder, a speech impairment, and/or a neurologic disorder.

Procedures

Approval from the institutional review board at The Pennsylvania State University was obtained before running participants.

Detailed instructions were provided before beginning the experiment defining voice qualities of overall severity, roughness, breathiness, pitch adequacy, and strain according to the Consensus Auditory Perceptual Evaluation—Voice[5] and textbook definitions.[24] Participants listened to the synthesized samples via noise reduction headphones (each presented twice in random order, n = 70) and rated each voice quality on a visual analog scale ranging from mild to severe covering a range from 1 to 1000, for the previously described voice qualities. During this task, a time stamp in seconds was collected as well as the number of times participants replayed each stimulus while making their judgments, using Alvin2.[25] Playback level was adjusted to a comfortable level for each participant individually.

RESULTS

Analysis of variance revealed a significant effect for group and type of stimulus on response time during the rating task, $F(5, 4199) = 5.66$, $P < 0.05$ and $F(3, 4199) = 2.76$, $P < 0.05$, respectively. An interaction effect of group and type of stimulus was not significant, $F(15, 4199) = 0.94$, $P > 0.05$. Tukey honestly significant difference (HSD) criterion indicated that although there was some overlap among groups, individuals with a higher level of experience (SVTs and SLPs) took the longest amount of time when rating stimuli (Table 1).

Tukey HSD criterion also showed some spread regarding stimulus type; however, there was a significant difference in response time between stimuli with simultaneous alterations in jitter and shimmer and stimuli with altered NHR. Samples with simultaneous altered jitter and shimmer took the longest time to rate, whereas NHR stimuli took the shortest time (Table 2).

An additional analysis of variance was conducted to examine the effect of experience on number of replays during the rating task. Each participant was permitted to replay the stimulus as many times as he or she needed during the experiment. Results revealed a significant effect of group type on number of repetitions, $F(5, 4199) = 31.57$, $P < 0.05$. The post hoc Tukey HSD criterion indicated a significant difference between the SLPGRAD group and all other groups. This group used the smallest number of repetitions of each stimulus to make their judgments. Compared to the IE and SVT groups,

TABLE 1.
Group: Tukey HSD Results for Response Time

Group	Mean (in s)			
SVTs	19 196	A		
SLPs	18 323	A	B	
IEs	17 479	A	B	
SVTGRADs	17 438	A	B	
SLPGRADVs	15 169		B	C
SLPGRADs	13 307			C

Note: Means that do not share a letter are significantly different.[26]

Download English Version:

https://daneshyari.com/en/article/1101211

Download Persian Version:

https://daneshyari.com/article/1101211

Daneshyari.com