Validity of self-selected volume adjustments for remote auditory testing in typical-hearers

Hannah Rae Vaden & Jani Johnson

School of Communication Sciences and Disorder, University of Memphis, Memphis, TN

Introduction

Traditionally, whenever audiologists test loudness perception, a procedure called a loudness contour test is used (Cox et al. 1987). In this clinical scenario, the audiologist tests loudness perception by presenting pure tones or speech stimuli at varying intensities through an audiometer while the listener is seated in a sound-treated space. Patients indicate their perception of how loud each sound is using a standardized scale of loudness categories ranging from just audible to uncomfortably loud.

With the occurrence of the COVID-19 pandemic, remote auditory testing has become one of the "new normal" modes of assessment that the public has needed to adjust to. However, research on the reliability of such measures is limited. Some researchers have found that remote hearing tests can be reliable for finding pure-tone threshold measures, in other words, finding the softest sound a person can detect (e.g., Molander et al., 2013; Mosely et al., 2018). However, it is not clear whether supra-threshold measures of loudness could also be reliably administered using remote methods.

This study aimed to establish whether young adults with typical hearing can reliability adjust volume settings to predetermined loudness categories when watching and listening to prerecorded audiovisual speech materials. Specifically, the following questions were explored:

1. **Do participants adjust the background noise to ensure an appropriately quiet test environment?**
2. **Are people able to reliably adjust the volume to match the category of loudness?**
3. **How different were participants amongst themselves?**
4. **Are the norms collected for this study significantly different from formal testing?**

Methods

Design: Non-intervention descriptive design

Recruitment: Participants were recruited through email and recruitment flyers posted in common areas in the school building

Materials: The Connected Speech Test (Cox et. al, 1989) was played through a Bluetooth speaker, tower speaker, or laptop speaker. Output was measured by the NIOSH Sound Level Meter app on an iPhone.

Participants: 35 participants completed the study. Ages of participants ranged from 21-29 years old (31 females).

Procedures: Participants were presented with two blocks of 7 different loudness categories in random order and told to adjust the level of the CST until it represented the category they were given. A sound level meter was used to document output in dB SPL following each adjustment.

Loudness Categories

7. Uncomfortably loud
6. Loud, but o.k.
5. Comfortable, but slightly loud
4. Comfortable
3. Comfortable, but slightly soft
2. Soft
1. Very soft

Results

Q1: Did participants adjust the background noise to ensure an appropriately quiet test environment?

Yes, participants were asked to select an appropriate environment for testing and to adjust the environment if needed. Ambient noise measured at the level of the participants' ears ranged from 25-41 dB SPL (x = 32), which is an adequate noise floor for suprathreshold testing.

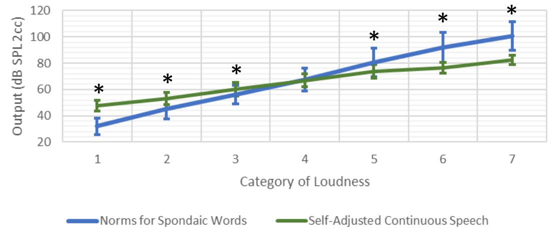
Q2: Were volume adjustments to match loudness categories repeatable?

Yes, self-selected volume adjustments from participants were repeatable between sets. Intensity levels across trials were highly correlated for all loudness categories. Correlations ranged from r = .52 to r = .85, all $p < 0.001$.

Q3: Were volume adjustments to match varying loudness categories reliable across participants?

Yes, the standard deviations of intensity self-adjustments were 5 dB SPL or less for all loudness categories. Error bars in Figure 1 are 1 standard deviation.

Figure 1 Intensities Across Loudness Categories: Norms for Spondees Compared with Self-Adjusted Continuous Speech



Q4: How did average volume self-adjustments to different loudness categories compare to results of traditional loudness perception measures using the Loudness Contour Test procedure in a sound booth?

Using the self-adjustment procedure, participants tended to adjust the volume of the CST to higher intensities for the softest categories, and to lower intensities for the loudest categories. Single sample t-tests supported these observations. * in Fig 1 indicate significant statistical differences.

Post Hoc Analysis and Discussion

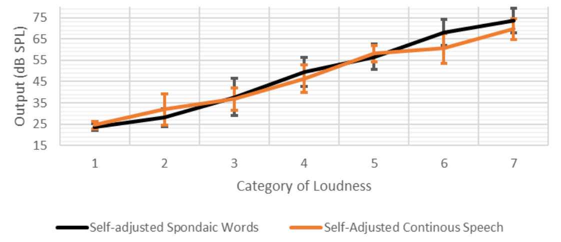
Apples to Oranges? Although transducer differences were accounted for by applying appropriate reference equivalent threshold SPL values (RETSPLs) to the data to estimate equivalent dB SPL values in a 2cc coupler, additional differences between the traditional and experimental test procedures are likely to have impacted loudness perception.

	Traditional Method	Experimental Method
Progression of loudness experience	Ascending	Randomized
Stimulus duration	Spondee	Continuous speech
Transducer	Insert earphone (monaural)	Loudspeaker (binaural)
Mode	Audio-only	Audio-visual
Test environment	Sound-treated room	Quiet room

Differences due to stimulus duration – Post Hoc Experiment:

To explore whether observed differences could be explained primarily by the differences in duration of the speech material, a post hoc experiment was performed. Ten additional participants (all female) completed the self-adjustment procedure in a sound room with spondee words and audio-only CST materials (order of stimulus type was counterbalanced). Results are presented below.

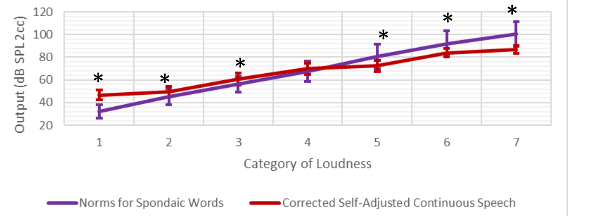
Figure 2 Intensities Across Loudness Categories: Self-Adjusted Spondees Compared with Self-Adjusted Continuous Speech



Discussion. Duration differences likely explain some of the variance observed for the original comparison, especially for louder categories. Differences in intensities obtained for the post-hoc group were applied as a correction factor to the intensities selected for the initial group's self-adjusted continuous speech categories to account for these differences (Displayed in Figure 3). Despite these corrections, small but statistically significant differences remain for most categories (* = 1-sample t-tests, p < .001).

Discussion (con'd)

Figure 3 Intensities Across Loudness Categories: Norms for Spondees Compared with *Corrected* Self-Adjusted Continuous Speech



Our participants adjusted sounds to louder levels for the softest category (Very Soft) compared to the traditional methods (mean difference = 14 dB SPL). This is primarily due to differences in ambient noise in the different environments. When the post-hoc experiment was repeated in a sound room the mean self-adjusted intensities for Category 1 continuous speech were reduced by 10 dB SPL.. For the louder categories our participants adjusted the sound to lower intensities compared to traditional loudness categorization. One explanation for this finding is procedural differences. Randomizing the categories of loudness alters a listener's internal reference differently than using an ascending procedure. The presence of visual cues can also increase the loudness perception of a given stimuli (Fastl, 2004). Although AV speech stimuli presented through a loudspeaker has a relatively small binaural-monaural ratio (x=1.09, Epstien & Florentine, 2012) binaural summation is also a likely contributor to this finding. An additional important observation is the probable psychosocial impact of listening to very loud speech stimuli in a typical environment. In many instances the test location was a room that was adjacent to areas where others might be studying or working. It is likely that participants were hesitant to allow the stimuli to reach levels that would call attention to themselves or be disruptive to others. Future research should explore the different contributions of these factors.

Conclusions

Remote auditory studies requiring super threshold testing with young typical hearers can reliably utilize self-adjustments to different volumes within and across participants. However, only those results obtained at average levels should be compared against data collected in a sound-treated room.

References

- Cox, R. (1989). Comfortable loudness level: stimulus effects, long-term reliability, and predictability. Journal of speech and hearing research, 32 4, 816-28 .
- Fastl, H. (2004). Audio-visual Interactions in Loudness Evaluation. *International Commission to Acoustics*.
- Hawley, M. L., Sherlock, L. P., & Formby, C. (2017). Intra- and Intersubject Variability in Audiometric Measures and Loudness Judgments in Older Listeners with Normal Hearing. Seminars in hearing, 38(1), 3–25. https://doi.org/10.1055/s-0037-1598063
- Molander, P., Nordqvist, P., Oberg, M., Lunner, T., Lyxell, B., & Andersson, G. (2013). Internet-based hearing screening using speech-in-noise: validation and comparisons of self-reported hearing problems, quality of life and phonological representation. BMJ open, 3(9), e003223. https://doi.org/10.1136/bmjopen-2013-003223
- Mosley, C. L., Langley, L. M., Davis, A., McMahon, C. M., & Tremblay, K. L. (2019). Reliability of the Home Hearing Test: Implications for Public Health. Journal of the American Academy of Audiology, 30(3), 208–216. https://doi.org/10.3766/jaaa.17092