Development and Evaluation of an English Language Measure of Detection of Word-Final Plurality Markers: The University of Western Ontario Plurals Test

Danielle Glista and Susan Scollie

Purpose: This article describes the development and evaluation of The University of Western Ontario (UWO) Plurals Test, which is an English language measure of detection of the word-final fricative cue for plurality.

Method: Normative data are provided for 26 listeners with normal hearing and 24 listeners with hearing impairment (children and adults), as are evaluations of the acoustical properties of the stimuli, the test’s test–retest reliability, and the test’s sensitivity to changes in hearing aid signal processing (e.g., nonlinear frequency compression).

Results: Results indicate reliable, repeated outcome measurement at the level of the individual. When compared to a global measure of real-world listening preference, the UWO Plurals Test was found to be somewhat sensitive to the effects of changes in hearing aid signal processing.

Conclusion: Findings suggest potential use of the UWO Plurals Test to evaluate aided and unaided ability of listeners between the ages of 6 and 81 years to detect the word-final fricatives /s/ and /z/ as they occur in English plural nouns.

Key Words: hearing loss, hearing aids, aided outcome measurement, speech sound detection

Fricative sounds are produced by creating a narrow opening in the oral cavity and forcing air through; this results in a “noisy” sound referred to as *frication noise* (Stevens, 1998). The fricative sound /s/ is one of the most commonly occurring English consonant sounds. The /s/ and its voiced cognate /z/ play important linguistic functions in the English language, marking the plurality of nouns and word tense in the word-final position. The sound /s/ is composed of predominantly high-frequency energy that varies in spectral shape according to the talker. The spectral peak of /s/ resides between ~2.9 and 8.9 kHz, with the peak of female speech residing closer to 9 kHz (Boothroyd & Medwetsky, 1992; Nittrouer, 1995). Furthermore, the spectral peak of /s/ for child speech has been measured to be higher in frequency and lower in intensity than that for adult speech (Pittman, Stelmachowicz, Lewis, & Hoover, 2003).

Research by Moeller et al. (2007) suggested that children with hearing impairment (HI) who are fitted with amplification at an early age have delayed fricative production compared to their normal hearing (NH) counterparts. Moeller et al. discussed the possible relationship between restricted hearing aid bandwidth and delayed speech and language development. Bandwidth required for accurate fricative recognition in listeners with moderate to moderately severe HI may be different for adults than for children (Pittman & Stelmachowicz, 2000; Stelmachowicz, Pittman, Hoover, & Lewis, 2001; Stelmachowicz, Pittman, Hoover, Lewis, & Moeller, 2004).

There are several signal processing options in current hearing aid technology that are designed to enhance high-frequency

Disclosure Statement
A portion of test development and data collection for the work presented in this paper was supported by a Collaborative Health Research Project grant that was cofunded by the Natural Sciences and Engineering Research Council and the Canadian Institutes of Health Research. Matching funds and in-kind support (hearing aids) for this grant were provided by Phonak AG.

Phonak AG has since licensed the test described in this paper for distribution to clinical audiologists. Licensing fees are received by the University of Western Ontario and are applied to further research and technology transfer. Drs. Scollie and Glista have provided continuing education lectures under Phonak sponsorship on various topics broadly related to pediatric hearing aid fitting and technology-related outcomes. Dr. Scollie is a member of the Phonak Pediatric Advisory Board. This test is made available by Phonak at no charge to clinicians. Licensing fees for the distribution of this test are received by the University of Western Ontario and are applied to research efforts.
audibility. These include extended bandwidth hearing and frequency-lowering hearing aids. Despite improved hearing aid bandwidth, current hearing aids are still limited in providing audibility of high-frequency speech sounds for individuals with a severe high-frequency hearing loss. Frequency-lowering hearing aids provide an alternative for these individuals via extended bandwidth. The goal of frequency-lowering technology is to shift high-frequency components of a signal into a lower frequency region. Current research has shown successful outcomes for some individuals with high-frequency hearing loss (Simpson, 2009). Frequency lowering can be enabled/disabled in various commercially available hearing aids. Two types of currently available frequency-lowering schemes are frequency transposition hearing aid processing (Auriemo et al., 2009) and nonlinear frequency compression (NLFC) hearing aid processing (Simpson, Hersbach, & McDermott, 2005).

One way of evaluating hearing aid benefit is through the use of aided speech perception testing. The properties of the chosen speech materials and psychometric procedure determine the sensitivity of the measure to evaluate hearing aid performance and/or benefit. For listeners receiving new audibility associated with novel hearing aid signal processing, an evaluation of any change to the aided condition may be of interest. For example, an efficacy evaluation of either extended bandwidth or frequency-lowering hearing aids, which are designed to alter the aided response associated with high-frequency signals, could be tested using speech materials and procedures that are sensitive to changes in high-frequency speech perception.

This article describes the development and evaluation of an English language measure of detection of plurality cues: The University of Western Ontario (UWO) Plurals Test. Specifically, this test assesses a listener’s ability to detect the word-final fricatives /s/ and /z/ as they occur in the English plural forms of many nouns (e.g., cats). Stimuli for this task were chosen to be similar to those used in previous research related to hearing aid bandwidth and high-frequency audibility in children (Stelmachowicz, Pittman, Hoover, & Lewis, 2002). Normative data for NH listeners and listeners with HI are provided in this article, as are evaluations of the acoustical properties of the stimuli, the test’s test–retest reliability, and the test’s sensitivity to changes in hearing aid signal processing (specifically, frequency compression).

Method

Development of Stimuli

Stimuli were recorded and digitized using a studio-grade microphone (AKG Acoustics) coupled to a pre-amplifier, an analog-to-digital converter (USB Pre, Sound Devices LLC), and sound recording software (SpectraPlus). Each word was recorded four times; the repetition with the least variation in pitch contour (as judged by an experimenter) was selected for inclusion in the final stimulus set. Stimuli were equalized in level using sound editing software (Goldwave). The final stimulus set included the singular and plural forms of 15 words: ant, balloon, book, butterfly, crab, crayon, cup, dog, fly, flower, frog, pig, skunk, sock, and shoe. Words were spoken in Canadian English by an adult female talker.

To illustrate the amplitude and spectral characteristics of the chosen words, the fricative noise was excised from the plural forms of each word. Excised fricatives were then concatenated, and the spectrum of the resulting file was measured. This fricative-only spectrum was compared to the spectrum of a concatenated stimulus containing all 30 words; the task-specific spectra were compared to a spectrum of the international long-term average speech spectrum (ILTASS, Figure 1) (Byrne et al., 1994). All spectral curves were normalized to a level of 65 dB SPL for display purposes. On average, the word-final fricative sounds had a bandwidth of energy residing at >4 kHz and <~10 kHz, with the spectral peak residing at 5 kHz. These results are consistent with previous data reported on the spectral characteristics of the female word-final /s/ (Pittman & Stelmachowicz, 2000). Comparison to the ILTASS illustrates the atypical boost of high-frequency energy that was present in the test stimuli.

Participants

A total of 49 participants were recruited from the University of Western Ontario (UWO) H.A. Leeper Speech and Hearing Clinic, as well as from local audiology clinics and schools. Participants included 14 NH children (6–18 years old), 12 NH adults (21–27 years old), 11 children with HI (6–17 years old), and 13 adults with HI (50–81 years old). The UWO Research Ethics Board approved this study for health sciences research involving human subjects. Participants with HI were part of a larger project that was carried out at UWO (Glista et al., 2009). Pure-tone air and bone conduction thresholds were measured bilaterally using a Grason-Stadler 61 audiometer. Air conduction thresholds were obtained using Etymotic Research ER-3A insert earphones (NH participants), with coupling to personal earmolds for all participants with HI. Air conduction hearing level was screened at 15 dB HL across all octave and interoctave frequencies for NH child listeners and at 20 dB HL for NH adult listeners. Tympanometry revealed normal middle ear

Figure 1. 1/3-octave band spectra (dB SPL) for all singular and plural words (solid line) and the word-final fricative portion only of the plural words (dashed line). Spectra are referenced to an international long-term average speech spectrum (ILTASS) (dotted line).
function across all NH participants. For participants with HI, thresholds were measured at all octave and interoctave frequencies between 250 Hz and 8000 Hz. Mean audiometric hearing thresholds are reported for the child and adult groups separately (Table 1).

Absolute hearing thresholds ranged from NH in the low frequencies to a profound level of impairment in the high frequencies across both age groups. Mean hearing thresholds for the child group can be characterized as sloping from a mild level of impairment in the low frequencies to a severe level of impairment. Both age groups presented with bilaterally symmetrical hearing loss. Hearing loss symmetry was assessed according to a difference in high-frequency pure-tone average values (HF-PTA), calculated using averaged threshold values across 2000 Hz, 3000 Hz, and 4000 Hz. Individual HF-PTA values ranged from 0 dB to 23 dB, with group mean difference values of 10 dB and 7 dB for adults and children, respectively.

**Testing Procedure**

The testing paradigm included the use of computer-controlled software to automate stimulus presentation and scoring. Stimulus files including two repetitions were automatically generated via a computer in random order, generating a score out of 60 items per measurement. Stimulus presentation was routed to a sound field within a double-walled sound booth via a loudspeaker that was positioned at 0° azimuth. A speech-shaped noise at a +20 dB signal-to-noise ratio (SNR) was presented for the entire duration of each measurement; this was generated from a clinical audiometer and was routed to four loudspeakers encircling each study participant at 72° spacing. During the pilot stage of the project, a listening check revealed a slight stimulus offset cue that was attributable to a change in the noise floor at the end of the stimulus. A low-level masking noise was therefore added to the test to ensure that audibility of the friction cue, specifically, was required to distinguish plural from singular items. Pilot testing was carried out with NH listeners under conditions of 3000-Hz low pass filtering; in this condition, with the inclusion of the masking noise, listeners could no longer perceive stimulus offset cues and could not reliably detect word-final plural markers.

A closed response set was presented on a computer monitor using pictures (from articulation cards, Super Duper Publications) and a corresponding orthographic display. The monitor was positioned slightly behind (and off to one side of) the loudspeaker at 0°. Each response set included the singular and plural form of each target word.

Participants were instructed to use the computer mouse to choose which picture best described what he or she heard.

Presentation level for the NH group was fixed at 50 dB SPL. For the group with HI, presentation level was varied to accommodate the various hearing levels and speech recognition abilities of the participants. The minimum testing level was 50 dB SPL, representing speech at a low vocal effort level (Olsen, 1998). For some listeners with HI, the test level was increased if the participant’s score for a given test level was at or below chance performance. In general, presentation levels ranged from a minimum of 50 dB SPL to a maximum of 65 dB SPL across participants. Once the presentation level for a given participant was determined, the same level was used for all measures obtained from that participant. This procedure is consistent with that reported by Wolfe and colleagues (2010) in a study evaluating the recognition of word-final plurality in an open response format (Wolfe et al., 2010, 2011).

**Hearing Aid Trials**

All of the participants with HI were full-time hearing aid wearers before entering the study, with the exception of two adult participants and one child participant. Bilateral prototype behind-the-ear hearing aids (similar to Phonak Savia 311 or 411) were provided to each participant along with FM-compatible audio shoes. Digital noise reduction features and automatic program selectors were disabled. Device fitting was completed using methods from the DSL v5.0 prescription: Child targets were used when fitting the children, and adults targets were used when fitting the adults (Bagatto et al., 2005). Age-dependent prescriptive targets were matched using simulated real ear measures incorporating individual real ear to coupler difference values. Note that the DSL v5.0 algorithm prescribes more gain for children than for adults (Scollie et al., 2005). The Audioscan Verifit VF-1 hearing instrument fitting system was used to measure aided responses for speech at 55 dB SPL, 65 dB SPL, and 75 dB SPL, and for a 90-dB SPL tone burst test signal, across all hearing aid fittings.

Testing was completed for two aided conditions: with and without NLFC. Testing with NLFC enabled was completed after 11 weeks of real-world experience using the hearing aid, on average (Range = 3 weeks–1.3 years). A follow-up appointment was scheduled for 5 weeks later, on average (Range = 2 weeks–5 months), to measure laboratory performance without NLFC. Participants also completed real-world trials and provided double-blind global preference ratings for the hearing aid program with or without NLFC active. Participants were given the option of choosing “same” if they felt there was no difference between the programs being

---

**Table 1. Average air conduction pure-tone thresholds measured in dB HL for both the right and left ear, by frequency.**

<table>
<thead>
<tr>
<th>Frequency (kHz)</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Children</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>23</td>
<td>26</td>
<td>37</td>
<td>43</td>
<td>53</td>
<td>68</td>
<td>86</td>
<td>92</td>
<td>91</td>
<td>97</td>
</tr>
<tr>
<td>Left</td>
<td>25</td>
<td>35</td>
<td>41</td>
<td>43</td>
<td>60</td>
<td>75</td>
<td>91</td>
<td>96</td>
<td>95</td>
<td>99</td>
</tr>
<tr>
<td><strong>Adults</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>27</td>
<td>25</td>
<td>34</td>
<td>37</td>
<td>47</td>
<td>62</td>
<td>82</td>
<td>87</td>
<td>89</td>
<td>92</td>
</tr>
<tr>
<td>Left</td>
<td>28</td>
<td>27</td>
<td>35</td>
<td>40</td>
<td>49</td>
<td>63</td>
<td>82</td>
<td>90</td>
<td>90</td>
<td>97</td>
</tr>
</tbody>
</table>

---

American Journal of Audiology • Vol. 21 • 76–81 • June 2012
compared. Preference ratings were recorded in a diary format. Retrospective analysis of performance on the UWO Plurals Test and the global measure of participant preference was incorporated into this study to investigate the sensitivity of the UWO Plurals Test to changes in the aided condition.

Testing Conditions and Scoring

Sound-field testing was completed in the unaided condition for all NH participants (new data) and in aided conditions for all participants with HI (data reported in Glista et al., 2009, was re-analyzed for inclusion in this data set). Participants completed two repetitions of each test to yield one test score (60 items); this was then repeated to obtain test–retest data for each testing condition (60 items × 2). Aided testing was completed without NLFC processing to evaluate test–retest scores and with NLFC processing to evaluate performance change across testing conditions. NLFC settings were individually determined and were systematically evaluated according to previously established procedures (Glista & Scollie, 2009).

Results for the participants with HI, previously reported in Glista et al. (2009), were re-analyzed for inclusion in this study. Results were scored using critical difference values according to the binomial theorem (Thornton & Raffin, 1978). Critical difference values were used to determine when two scores were statistically different from each other according to a 60-item word list.

Results

On average, the NH listeners performed at ceiling, with scores ranging from 93% to 100% correct, whereas scores for the listeners with HI ranged from ~45% to near ceiling. Results were analyzed using critical difference values to assess the following: (a) test–retest reliability of the test and (b) the sensitivity of the test in evaluating the effect of NLFC hearing aid processing on word-final plurality recognition ability.

Test–Retest Reliability

Test–retest scores for all NH and HI participants were within the 90% critical difference values, with one HI child’s data point falling outside the 95% critical limit (Figure 2). Another child with HI was excluded from test–retest evaluation because of missing data for the second repetition of the test; this was due to a limitation in available testing time. The intraclass correlation coefficient (ICC) was used to evaluate test–retest reliability, using an ICC type that tests whether repetitions of the same test provide the same score.1 ICCs for test–retest across all participants equaled .95 (F = 42.4, p < .01). However, these values include many scores that were at ceiling for the NH participants, possibly inflating reliability estimates and providing only limited information about test–retest reliability in a clinical population. Therefore, the test–retest performance was re-evaluated including only the participants with HI. The ICC for the participants with HI equaled .85 and was significant (F = 12.2, p < .01). Both ICCs were >.75, indicating good test–retest reliability (Portney & Watkins, 2000).

Sensitivity

To assess the test’s sensitivity, HI results obtained with and without NLFC enabled were analyzed by comparing the change in test scores to the change in global real-world preference ratings (Figure 3) (Stratford & Riddle, 2005). First, the data were dichotomized according to the presence (coded as 1) or absence (coded as 0) of benefit as judged using critical difference values for the Plurals Test (90% limits). Data points falling at ceiling were not included in the analysis (n = 2). Results for 22 participants were included; 11 data points fell outside the critical limits, indicating

---

1ICCs were computed using a two-way random effects model and an absolute agreement definition.

---

Figure 2. Test–retest data for the normal hearing (NH) adults (filled squares), adults with hearing impairment (HI) (open squares), NH children (filled circles), and children with HI (open circles). Aided testing was completed with conventional hearing aid technology (without frequency lowering).

Figure 3. Aided change in performance displayed as a function of hearing aid condition (with nonlinear frequency compression [No-NLFC] versus NLFC active) for adults with HI (open squares) and children with HI (open circles).
significant benefit with frequency lowering on the Plurals Test. No performance decrements were observed. These “change” scores were then compared with participants’ retrospective global ratings of performance change using correlation analysis. This analysis evaluated whether significant changes on the Plurals Test were associated with noticeable changes to the hearing aid wearers. Preference ratings were categorized, with 1 indicating a preference for frequency lowering, 0 indicating no preference, and –1 indicating a preference for conventional hearing aid processing. For the group of children with HI (nine participants were included in this analysis), five children indicated a preference for NLFC enabled, one indicated a preference for NLFC disabled, and three indicated no preference. For the group of adults with HI (13 participants), two participants indicated a preference for NLFC enabled, two indicated a preference for NLFC disabled, and nine indicated no preference.

A two-tailed, nonparametric correlation analysis was completed using the Kendall’s tau algorithm. A significant positive correlation between benefit change and overall subjective preference was measured (α < .05): ICC = .45 (p = .030). This indicates that listeners who performed better on the UWO Plurals Test with NLFC processing were also somewhat less likely to prefer using NLFC on real-world trials, whereas listeners who performed equivocally across conditions were less likely to prefer NLFC (i.e., they may have had no preference or may have preferred NLFC disabled). Evaluation of the relation between preference for hearing aids without NLFC and NLFC decrement on the Plurals Test could not be performed within this data set, as no listeners experienced decrement. The significant relationship between user preference and NLFC benefit on the Plurals Test suggests that this test is somewhat sensitive to changes in the perception of word-final /s/ sounds, and that meaningful changes in scores are somewhat associated with improved hearing aid function as noticed by the hearing aid wearer.

Discussion

The UWO Plurals Test is an English language measure of the detection of plurality markers. Acoustic analyses of the stimuli illustrate the high-frequency emphasis of this task: The bandwidth of energy associated with the word-final fricative sounds resides at >4 and <10 kHz, and the spectral peak resides at 5 kHz, on average. The purpose of this measure is therefore limited to an evaluation of high-frequency audibility of fricatives /s/ and /z/ as bound morphemes. This test was not designed to provide information related to overall consonant recognition or speech discrimination abilities and therefore does not replace aided testing of consonant discrimination. Rather, the focus is on the identification of the presence of a plural marker in the test stimulus. High performance scores on this test are associated with detecting the word-final fricatives in the plural stimuli and do not indicate that the listener correctly perceives an /s/ or /z/ phoneme. Confusion with other phonemes could co-occur with high detection scores.

This study assessed performance across 26 NH listeners in the unaided sound field and 24 HI listeners in the aided sound field. All NH participants achieved high scores regardless of age group, suggesting that the test materials and procedures are feasible for use on individuals between the ages of 6 and 81 years. Performance for the listeners with HI was assessed in two hearing aid conditions: with and without NLFC active. Performance with NLFC’s digital signal processing was deemed an appropriate test condition for use in the sensitivity analyses as past investigations with the UWO Plurals Test have shown significant changes in score between NLFC-on and NLFC-off conditions (Simpson, 2009). This test was not developed to be used exclusively in evaluating the efficacy of NLFC, and may be sensitive to changes in performance with other types of frequency-lowering, hearing aid bandwidth changes and/or gain characteristics. Further research is needed to assess the sensitivity of the test across these or other examples of clinically relevant aided conditions.

Results obtained in NH listeners and listeners with HI (aided conditions) indicate reliable repeated outcome measurement at the level of the individual. When compared to a global measure of hearing aid preference, the Plurals Test was found to be somewhat sensitive to the effects of frequency lowering. Testing was reliably completed with children as young as 6 years of age. Further research is needed to determine if the test can be used with younger children. Findings reported should not be generalized outside the methods used in this study.

Clinical Implications

The UWO Plurals Test has the potential to be used to evaluate detection of high-frequency fricative sounds in a word-final position in clinical environments. The test is available on CD at no charge to clinicians in some countries (refer to the Disclosure Statement). The UWO Plurals Test may offer a way of assessing performance of hearing aid technology designed to enhance audibility of high-frequency sounds beyond what conventional amplification can offer. Specifically, this test may be sensitive to amplification in the range of 5–10 kHz. Methods reported in this article used a closed response set format of administration. For example, pictures of the singular and plural version of the root word appeared on a computer monitor during testing. The listener was responsible for clicking on the picture that best described what he or she heard. The same testing paradigm could also be achieved using paired picture cards, for example, presented in the same order as a prerecorded word list; this type of method would require manual scoring by the tester.

It is possible to administer the Plurals Test in a way that requires the listener to repeat what he or she heard rather than respond by choosing a picture (i.e., an open response set format). If administered in this format, the tester must judge from the listener’s verbal response whether a singular or plural form of a word was heard and mark the response as correct or incorrect on a response form; however, scoring accuracy may depend on the examiner’s hearing ability in the high frequencies. When using open response set formats, it is possible for the listener to correctly identify the absence or presence of plurality without correctly identifying the root word. The impact of this issue on test–retest reliability or test sensitivity has not been evaluated in this
article. However, open-set administration of the UWO Plurals Test was reported in a study by Wolfe et al. (2010, 2011) that evaluated speech perception ability in children and the effects of frequency-lowering hearing aid processing. Results indicate that the test was sensitive to the effects of NLFC.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada and Canadian Institutes of Health Research (Collaborative Health Research Project #313114-2005), The Hearing Foundation of Canada, The Canada Foundation for Innovation, the Masonic Foundation of Ontario Help-2-Hear Project, and Phonak AG. Special thanks to Andrea Dunn, Andreas Seelisch, Julie Seto, Julianne Tenhaaf, Marlene Bagatto, and Melissa Polonenko for their contributions to this work.

References


