DEVELOPMENT OF AN EFFICIENT ANALYTICAL HIGH FREQUENCY NONSENSE SYLLABLE TEST

Petri Korhonen, M.Sc., Bryan Crose, B.S., Chi Lau, Ph.D.

INTRODUCTION

Several recent advances in hearing aid technologies have resulted in processing methods that affect the way the higher frequencies are delivered to the listener. The signal processing effects related to these features are typically specific to particular frequency regions, and therefore the tools that we use to assess these effects have to be sensitive to the changes in signals at these frequencies.

We had developed a nonsense syllable speech identification test for evaluating the effects of high-frequency processing in hearing aids. All consonants and five vowels of American English were included in the test to allow a comprehensive analysis of phoneme identification in an open-set task. The test allows a comprehensive analysis of phoneme identification performance for all American English consonants. Initially the test consisted of 115 nonsense CVCVC syllables. However, to evaluate the performance using this test requires a total of 15 - 25 minutes of testing. When evaluating several different test conditions, this may not be clinically feasible. Therefore, to facilitate shorter tesing time, the number of items in the original test were reduced. The reduction was carried out based on the normative data collected from normal hearing subjects.

I. ORIGINAL LIST

- The original speech test consisted of a list of 115 CVCVC nonsense items with 25 consonant sounds /p, t, k, b, d, g, m, n, ŋ, f, v, ð, s, z, ʃ, ʒ, l, w, wh, θ, j, h, ʧ, ʤ/ each presented in initial, medial, and final positions.
- Five vowels / I, ɛ, æ, ʌ, u/ were used to provide different contexts. Each consonant was paired with all five vowels in all positions once.
- Only exceptions to these rules were allowed where pronunciation was considered impossible.
- The source materials were produced by one male and one female speaker resulting in two separate implementations of the test.
- The test followed an open-format with item randomization performed prior to each test trial.
- Stimuli can be presented in quiet or in the presence of noise with selectable signal-to-noise ratio (SNR).
- A custom software program was developed for test administration, scoring, and automatic analysis of results.
- The administration time for the test was approximately 15-25 minutes depending on the number of errors the test subject made.

METHODS

Subjects and procedures

- Normative data using the 115 item list was collected from six normal hearing subjects aged 18-54 years.
- Speech was presented at conversational level (68 dB SPL) in quiet and in speech weighted background noise at SNR = +10, +5, 0, -5, and -10 dB.
- Stimuli and the noise was presented at single loudspeaker at 1 meter distance from the subject at 0 degrees in soundfield.

II. REDUCTION OF ORIGINAL TEST

Item reduction criteria

1. Elimination of non-fricative items

The goal was to have a test that is sensitive to the evaluation of high-frequency sounds. Therefore we required each word to include at least one unvoiced or voiced fricative sound.

2. Elimination of items deemed too difficult

Ideally the identification scores in quiet for normal hearing subjects should be close to 100 %. However, normative data revealed items that were difficult to identify. Items with scores more than one standard deviation from the mean score in the quiet condition were eliminated.

3. Elimination based on noise effects

The third criterion was to include items in the test which were affected similarly by noise. The noise effect for each item was defined as the difference in scores obtained in quiet and at -5dB SNR. Items for which this difference was in the range [mean ... mean + 2×standard deviation] were included in the shortened list test.

4. Elimination based on between-subject variation

Subject-to-subject variation was minimized by selecting the items based on the variation for each item at SNR = +5, 0, and -5. Items with a variance more than (one standard deviation) in two of the three conditions were eliminated.

A unified short list for both male and female speakers was generated using the male speaker data. The impact of each elimination criterion is depicted in Figure 1.

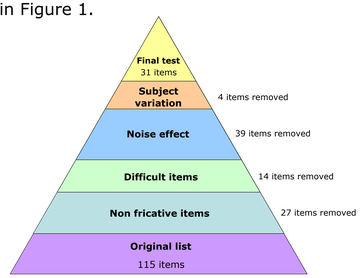


Figure 1: Number of items removed from the original 115 item list under each elimination criterion.

RESULTS

Phoneme distributions

After the elimination, 31 items remained in the short list generated based on male speaker results. Consonant /wh/ was, however, missing. To have a full set of consonants, one item was added back to the list. The ratio of the number of items in the short list to the long list was 1:3.6. This ratio in different phoneme groups between the short and the long lists was in the range of 1:2.7 to 1:4.4. Many groups in the short list kept about one third of the phonemes in the long list (Figure 2).

Table 1: Frequency distribution of consonants in the shortened test. C1, C2, and C3 correspond to the initial, medial, and final positions respectively.

	C1	C2	C3	Total
k	1	2	1	4
g	2	0	0	2
p	3	4	1	8
b	1	1	2	4
t	1	0	1	2
d	1	3	0	4
f	2	2	1	5
v	1	2	3	6
θ	0	2	3	5
ð	3	1	1	5
s	3	1	1	5
z	0	1	1	2
ʃ	2	3	1	6
ʒ	0	0	2	2
ʧ	0	1	1	2
ʤ	0	1	0	1
h	2	0	0	2
wh	1	0	0	1
w	2	2	0	4
j	1	1	0	2
ɹ	2	2	0	4
l	1	1	3	5
m	2	1	1	4
n	1	1	1	3
ŋ	0	0	1	1

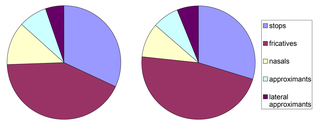


Figure 2: Distribution of phonemes in the long (left) and in the short (right) test.

Comparison of the short and long lists

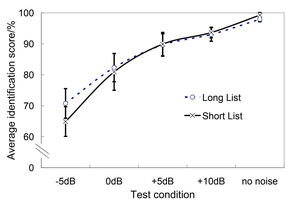


Figure 3: Average performance scores for materials produced by the male speaker for all phonemes.

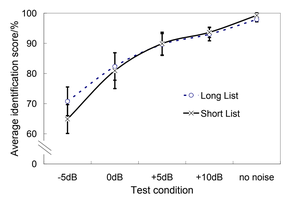
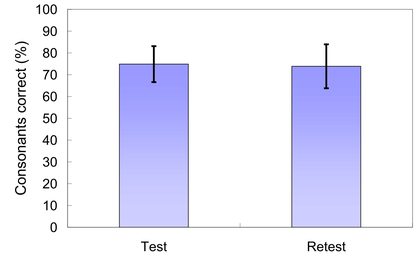


Figure 4: Average performance scores for materials produced by the female speaker for all phonemes.

- Male material (Figure 3): The largest difference (6%) occurred at the -5dB SNR test condition. For other test conditions, both lists resulted in similar scores. The variance of the short list was similar to the variance of the long list. Dentals and lateral approximants had a discrepancy of greater than 10% at more than one test condition. For other categories, the discrepancy between the two lists was less than 10 % at all test conditions.
- Female material (Figure 4): The averaged phoneme scores showed a 13% discrepancy in the –5 dB SNR. For other test conditions, performance with both lists was consistent. Discrepancies of greater than 10% (at more than one test condition) were seen for unvoiced consonants, unvoiced fricatives, and labiodentals. For other phoneme groups the discrepancy was less than 10 %.

RESULTS (CONTINUE)

Test-retest



The SNR = 0dB condition for male speaker was tested twice at different visits to investigate the test-retest reliability. The mean difference in phoneme score between visits was 1.6% (SD=7.3%). The mean difference for vowel sounds was 2.5% (SD = 3.5%), and for consonants was 1.0% (SD=11.0%). Using paired-samples t-test indicated that the test identification score was not significantly different from the retest identification score (t=0.57, p>0.05).

Internal consistency

The Cronbach's alpha was used for each test condition to measure the internal consistency of the selected items in both the male and the female lists. In theory, the higher the Cronbach's alpha, the more internal consistency is found for the items. For our test the Cronbach's alpha was high (0.8 - 1.0) for all test conditions, except for the female-speaker short list tested at +10dB SNR (alpha = 0.6).

DISCUSSION

The item selection reduced the test administration time from 15-25 to 5-7 minutes. This allows us to carry out several test trials during a single test session. This will potentially reduce fatigue effects. The results obtained with the shortened test are comparable to the result obtained using the original 115 item test. The shortened test retains all American English phonemes allowing a comprehensive analysis of phoneme-error patterns. However, in the shortened list the ratio of high frequency sounds is higher, which can make it a practical clinical tool to assess the effects of high frequency processing in hearing aids.

The shortened test has two implementations: one using female and another using male material. The spectrum of the female speech is typically higher in frequency than those for male speakers. Thus, it may be more optimal to study the processing methods that are sensitive to higher frequencies. In addition, since both male and female implementations of the current test consists of exactly the same items, one may be able to use the test to study gender effects.