



Published in final edited form as:

J Am Acad Audiol. 2012 May ; 23(5): 313–331. doi:10.3766/jaaa.23.5.3.

Evaluation of TIMIT Sentence List Equivalency with Adult Cochlear Implant Recipients

Sarah E. King¹, Jill B. Firszt^{1,2}, Ruth M. Reeder², Laura K. Holden², and Michael Strube³

¹Program in Audiology and Communication Sciences, Washington University School of Medicine, St. Louis, Missouri

²Department of Otolaryngology-Head and Neck Surgery, Washington University School of Medicine St. Louis, Missouri

³Department of Psychology, Washington University, St. Louis, Missouri

Abstract

Background—Current measures used to determine sentence recognition abilities in cochlear implant recipients often include tests with one talker and one rate of speech. Performance with these measures may not accurately represent the speech recognition abilities of the listeners. Evaluation of cochlear implant performance should include measures that reflect realistic listening conditions. For example, the use of multiple talkers who vary in gender, rate of speech and regional dialects represent varied communication interactions that people encounter daily. The TIMIT sentences, which use multiple talkers and incorporate these variations, provide additional test material for evaluating speech recognition. Dorman and colleagues created thirty-four lists of TIMIT sentences that were normalized for equal intelligibility using simulations of cochlear implant processing with normal-hearing listeners. Adults with sensorineural hearing loss who listen with cochlear implants represent a different population. Further study is needed to determine if these lists are equivalent for adult cochlear implant recipients, and if not, to identify a subset of lists that may be used with this population.

Purpose—To evaluate the speech recognition equivalence of 34 TIMIT sentence lists with adult cochlear implant recipients.

Research Design—A prospective study comparing test-retest results within the same group of listeners.

Study Sample—Twenty-two adult cochlear implant recipients who met the inclusion criteria of at least three months device use and a monosyllabic word score of 30% or greater participated in the study.

Data Collection and Analysis—Participants were administered 34 TIMIT sentence lists (20 sentences per list) at each of two test sessions several months apart. List order was randomized and results scored as percent of words correct. Test-retest correlations and 95% confidence intervals for the means were used to identify equivalent lists with high test-retest reliability.

Results—Mean list scores across participants ranged from 66% to 81% with an overall mean of 73%. Twenty-nine lists had high test-retest reliability. Using the overall mean as a benchmark, the

Corresponding Author: Jill B. Firszt, Ph.D., Department of Otolaryngology, Washington University School of Medicine, 660 S Euclid Avenue, Campus Box 8115, St. Louis, MO 63110, Phone: 314.362.4760, Fax: 314.362.4436, firsztj@ent.wustl.edu. Sarah King is currently at Midwest Ear Institute, Kansas City, MO

Portions of this manuscript were presented by the first author at the American Auditory Society Annual Meeting, Phoenix, Arizona, March 3-6, 2007 and the American Academy of Audiology Annual Conference, Denver, Colorado, April 18-21, 2007.

95% confidence intervals indicated that 25 of the remaining 29 lists were equivalent (e.g. the benchmark of 73% fell within the 95% confidence interval for both test and re-test).

Conclusions—Twenty-five of the TIMIT lists evaluated are equivalent when used with adult cochlear implant recipients who have open-set word recognition abilities. These lists may prove valuable for monitoring progress, comparing listening conditions or treatments, and developing aural rehabilitation plans for cochlear implant recipients.

Keywords

speech perception; speech recognition; cochlear implant; TIMIT; speaker variations

INTRODUCTION

Advancements in cochlear implant (CI) technology have resulted in improved performance for recipients. Studies have shown that with current test measures, CI recipients can obtain high levels of auditory-only speech recognition (Skinner et al, 1997; Firszt et al, 2004; Spahr and Dorman, 2004). As cochlear implant outcomes continue to improve, individuals with more residual hearing are being implanted (Alkaf and Firszt, 2007; Gifford et al, 2007; Dorman et al, 2008; Vermeire and Van de Heyning, 2008; Gifford et al, 2010). Additionally, bilateral cochlear implantation is becoming more prevalent and provides distinct benefits compared to those of unilateral implantation (Buss et al, 2008; Dunn et al, 2008; Firszt et al, 2008; Noble et al, 2008).

Common sentence measures used to evaluate CI performance have included Central Institute for the Deaf (CID; Davis and Silverman, 1978) sentences, City University of New York (CUNY; Boothroyd et al, 1985) sentences and Hearing In Noise Test (HINT; Nilsson et al, 1994) sentences. The HINT sentences presented in quiet have been the most common measure since the introduction of the Minimal Speech Test Battery for Adult Cochlear Implant Patients (MSTB; Nilsson et al, 1996; Luxford et al, 2001). HINT sentences were selected for the MSTB to provide a reasonable number of lists (25 lists to be given two at a time) and to avoid floor and ceiling effects. The developers of the MSTB envisioned the HINT sentences would eventually be given adaptively in noise as CI technology improved and ceiling effects were present in quiet (Luxford et al, 2001). However, the HINT sentences have rarely been administered in this manner for clinical CI evaluations. Presentation of sentence material in quiet is necessary to determine implant candidacy and has become a standard part of the test battery used to monitor performance over time. For research purposes, testing in quiet may be used, for example, to assess the effects of new programming parameters or speech processing strategies; however, ceiling effects in quiet with current sentence tests can negate their usefulness. Gifford and colleagues (2008) administered a battery of tests to 156 adult CI recipients and 50 adult hearing aid users. Ceiling effects were especially notable for HINT sentences in quiet regardless of listening mode (unilateral, bilateral, and bimodal) for implant recipients. The HINT sentences, as well as the CID and CUNY sentences, use a single male talker with an unnaturally clear speaking style which likely contributes to many postlingually deaf CI recipients scoring at ceiling when these sentences are presented in quiet. Measures such as these create an unrealistic listening task, do not represent everyday communication (Loizou et al, 1999; Koch et al, 2004) and may not be appropriate for the growing number of CI recipients who have considerable auditory-only speech recognition.

The amount of variability in acoustic characteristics, resulting from diverse speaking styles, speaker gender, and speaker rate can affect accuracy in speech recognition. Research has shown that performance is significantly reduced when multiple talkers are used across trials (Mullennix et al, 1989; Sommers, 1997). Word recognition scores are poorer when speech is

presented with multiple speaking styles (Kirk et al, 1997; Mullenix et al, 2002; Sommers and Barcroft, 2006). Likewise, including several speaking rates significantly reduces word identification (Hosoi et al, 1992; Uchanski et al, 1996; Sommers and Barcroft, 2006). As is seen in these studies, the use of a single well articulated speaker has been shown to improve speech recognition; however, since it does not represent the varied conversational partners one comes across in daily life, it can falsely inflate the perception of an individual's abilities. In the study by Gifford et al (2008), AzBio sentences (Spahr and Dorman, 2005) were also presented to unilateral, bilateral, and bimodal CI users as well as to hearing aid users. In contrast to HINT sentences, no ceiling effect was seen for AzBio sentences which have both male and female talkers and a speaking rate that is more common in daily conversation. Ideally, speech recognition measures should reflect the individual's abilities in the real world and need to be sensitive to variations in test conditions (Mackersie, 2002). In order to more accurately reflect skills needed for success in daily communication, speech recognition measures should include common speaker variations such as varied regional dialects, gender, and speaking rates.

The TIMIT acoustic-phonetic speech database (Lamel et al, 1986) was developed in 1986 as a joint effort between researchers at Texas Instruments (TI), the Massachusetts Institute of Technology (MIT) and the Stanford Research Institute to evaluate automatic speech recognition systems. The sentences were selected to be phonetically diverse and to include dialectal variants present in American English. The TIMIT acoustic-phonetic speech database consists of 6,300 different low context sentences spoken by 630 individuals (10 sentences per speaker). The speakers represent eight different American English regional dialects with both female and male speakers and a variety of speaking rates (Byrd, 1992). The database provides the sentence text, wave files and a phonetically-based transcription for each utterance (Linguistic Data Consortium, 1992).

The TIMIT sentences have been used as stimuli in several CI research studies (Loizou et al, 2000; Fu et al, 2002; Shannon et al, 2002). In two studies by Dorman et al (2003, 2005) the authors created lists of TIMIT sentences in order to evaluate simulations of combined acoustic and electric stimulation (EAS). For each study, TIMIT sentences were processed through a five-channel cochlear implant simulation and presented to normal-hearing listeners. Sentences were then combined into lists of 20 sentences each having equal intelligibility. Details given in the 2005 publication indicate that 950 sentences were processed and presented to 10 normal-hearing listeners, and that the average level of intelligibility for the sentence lists was 75% correct, $\pm 1\%$ standard deviation (SD).

Since TIMIT sentences include various speakers and speaking patterns, the lists developed by Dorman et al (2005) are appealing for use with CI recipients as a sentence recognition measure that better reflects real-life listening. However, these lists have not been evaluated for equivalency with actual CI recipients. Since hearing deficits result in a disruption to the auditory system (Kopra et al, 1968; Beattie, 1989; Wilson and Carter, 2001), hearing impaired listeners using an implant may differ greatly from normal-hearing individuals using CI simulations. The present study was designed to evaluate the test-retest reliability and equivalence of intelligibility of the 34 TIMIT sentence lists used by Dorman et al (2005) with adult CI recipients.

MATERIALS AND METHODS

The research protocol and informed consent for this study were reviewed and approved by the Institutional Review Board at the Human Research Protection Office of Washington University School of Medicine.

Participants

Twenty-two adult postlingually deaf CI recipients participated in the study. The number of participants included was based on the number of subjects used in a similar study by Skinner et al, 2006 that examined the equivalency of 30 lists of monosyllabic words. Individuals invited to participate had at least three months of device use and a score greater than 30% on their most recent Consonant-Vowel Nucleus-Consonant (CNC) Monosyllabic Word Test (Peterson and Lehiste, 1962). This minimum auditory-only speech recognition level was chosen to avoid floor effects and better allow comparison of variations in results between lists. All participants were recruited from the Washington University School of Medicine (WUSM) Adult Cochlear Implant Program.

Demographic information for the individual CI participants is shown in Table 1. The mean age of participants was 58 years (SD = 13.4) with a range of 25 to 78 years. Length of severe to profound hearing loss before implantation ranged from 3 months to almost 34 years with a mean of 10 years (SD = 8.3). The mean length of CI use for the participants was 3.9 years (SD = 2.5) with a range of 10 months to 10.6 years. Information about the CI device and processing strategy used by each participant can be found in Table 2. Eleven participants continued to use a hearing aid in the contralateral ear; however, the hearing aids were removed for all testing.

The participants' CNC word scores ranged from 44% to 92% with a mean of 73% (SD = 13%). Based on these word scores, this group of CI recipients performs above the average postlingually deafened CI recipient. Holden et al (2011) reported on a group of 114 postlingually deaf CI users whose CNC word scores were evaluated longitudinally from two weeks to two years post initial activation of the CI. The mean word score for the final test interval for this group of CI users was 62% with a range of 2% to 95%. This result is in close agreement with Gifford et al (2008). In that study, 162 CNC word scores were collected from both unilateral and bilateral CI recipients with the test administered in the unilateral CI condition to all participants. A mean CNC word score of 56% was reported.

Stimuli

The 34 TIMIT sentence lists used by Dorman et al (2005) were evaluated in the current study. The make-up of each list is summarized in Table 3 and the details needed to recreate the lists are in Appendix A. Each list had 20 sentences with an average of 128 words per list (range = 113 to 142 words). There were a total of 680 unique sentences spoken by a total of 301 unique talkers (94 female and 207 male). Each list included 18-20 unique talkers with at least six female and six male talkers per list (average of 10.4 female and 9.6 male). The talkers across all lists were from four dialectical regions (New England, Northern, North Midland and Western) with three lists (3, 12 and 17) having talkers from three of the dialectical regions and the remaining 31 lists having talkers from all four dialectical regions. The age range of talkers across lists was 20 – 67 years with the average age for each list being 29.4 years. The average reading level across lists was approximately sixth-grade, with a range from fourth to eighth grade, based on the Flesch-Kincaid formula (Kincaid et al, 1975; Flesch, 1979; Microsoft, 2007) compared to the average adult reading level in the United States of eighth to ninth grade (Kirsch et al, 2002).

Test Environment

The 34 TIMIT sentence lists were stored as wave files on the hard disk of a Dell workstation with dual Xeon CPUs and a 24-bit sound card for attenuation and mixing of audio stimuli. Sentences were presented through an amplifier (Crown model D-150) and loudspeaker (JBL LSR32 Linear Spatial Reference) in a double-walled sound booth (IAC model 1204-A) at 60 dB SPL which represents an average level of conversational speech (Pearsons et al, 1976).

The stimuli were calibrated using a sound level meter (Bruel and Kjaer model 2230) with a linear frequency weighting from 20 Hz to 20 kHz and the microphone (Bruel and Kjaer model 4155) placed at the approximate location of the participant's head. The participants were positioned at one meter and at 0° azimuth from the loudspeaker.

Procedure

Participants were tested with the speech processor program, volume and sensitivity settings they used daily. Detection thresholds were obtained in the sound-field at 250 to 6000 Hz with frequency-modulated (FM) tones using the standard Hughson-Westlake procedure and 2 dB increments. Thresholds were confirmed to be 30 dB HL or better from 250 – 6000 Hz prior to presentation of the sentences. These are expected levels for CI recipients with well programmed processors and properly functioning equipment (Holden et al, 2007; Holden et al, 2011).

List order was determined using a randomized-block (within subjects) design with a different random order of the 34 lists presented to each participant. Testing occurred during two sessions with breaks as needed to avoid fatigue. After the first session (Test 1), participants returned several months later and were retested on the same lists presented in a different order (Test 2). The length of time to administer a single TIMIT list was typically 4-5 minutes. Participants were asked to repeat each sentence and encouraged to guess if unsure. Lists were scored as percentage of words correct. During Test 2, a single list of 50 CNC words was administered at 60 dB SPL.

RESULTS

Group mean FM-tone sound-field threshold levels and standard deviations for frequencies .25, .5, 1, 2, 3, 4, and 6 kHz were 21.0 (5.1), 23.8 (5.3), 24.9 (4.9), 20.7 (5.1), 24.5 (6.0), 24.2 (6.3) and 20.8 (6.3), respectively. The group mean threshold across all frequencies and across participants was 23 dB HL (SD 4.26). These levels ensured audibility of the sentences when presented at 60 dB SPL.

Figure 1 shows the average of all 34 TIMIT list scores for each participant at Test 1 (gray bars) and Test 2 (white bars) with the averages across participants and lists shown on the far right. CNC word scores are also shown for each participant (open diamonds). Participants are ordered according to word scores from lowest to highest. Word scores ranged from 52% to 94% with an average across participants of 77.7% (SD 12.6%). The mean TIMIT list scores ranged from 54.2% to 89.1% for Test 1 and from 56.7% to 93.0% for Test 2. The grand average across lists and participants was 73.2% (SD 12.1%) and 73.1% (SD 13.0%) for Test 1 and Test 2, respectively. The participants' average TIMIT scores were significantly correlated with CNC scores, $r = 0.76$, $p < 0.001$ and in good agreement. The mean difference between each participant's CNC score and average TIMIT score was 7.6% (SD 5.8%).

Test-retest correlations for each list are shown in Table 4. All correlations were significant ($p < 0.01$) except for List 28 that had a correlation of 0.30. The average correlation for the remaining lists was 0.76 (ranged from 0.62 to 0.91).

Figure 2 shows the mean score across participants for each list at each of the two test sessions (Test 1, gray squares and Test 2, open diamonds) and the 95% confidence intervals for the means (error bars). Mean participant scores for all 34 lists ranged from 66.0% to 80.7% for Test 1 and from 66.9% to 79.0% for Test 2. Using the overall mean (73%) as a benchmark (shown with the solid black line), 29 of the lists (all except 1, 3, 6, 10 and 27) were considered equivalent. That is, for 29 of the lists, the 95% confidence intervals

included 73% for both Test 1 and 2. The benchmark falls outside the 95% confidence intervals for both Test 1 and Test 2 for Lists 1 and 6 (below for List 1 and above for List 6). For Lists 3, 10, and 27, the benchmark fell within the confidence interval for Test 2 but below the confidence interval for Test 1.

The similarity of the lists in phonetic structure was examined by calculating the dissimilarity of pairs of lists using the Euclidean distance based on the frequencies of the 40 speech sounds represented in each list. Figure 3 shows the average scaled Euclidean distance of each list from the remaining lists. Distances can range from 0 to 1, with lower values indicating similarity. The most obvious feature of Figure 3 is that List 10 is noticeably different from the other lists, which have fairly uniform average dissimilarities. (Note that List 10 is one of the 5 lists considered not equivalent.)

DISCUSSION

TIMIT sentences use multiple speakers and incorporate gender, dialect and speaking rate variations. These speaker variations represent the unpredictability of speech in everyday communication situations and therefore, have the potential to better assess speech recognition abilities than many current test measures. To date, research with the TIMIT sentences has focused on normal-hearing participants, listening through simulated CI processing or small CI participant samples. The CI recipient population is likely more varied; therefore, the current study was performed to determine if these 34 TIMIT sentence lists had high test-retest reliability and were equivalent when used with CI recipients. If not, the objective was to identify a subset of equivalent lists with sufficient test-retest reliability to be used either clinically or for research purposes. Results indicated that not all lists were equivalent when presented in quiet at a conversational level with mean list scores across subjects ranging from 66% to 81%. The overall average score across participants and lists (73%) was used as a benchmark and 29 of the 34 TIMIT sentence lists were identified as equivalent for this group of participants. One of the 29 lists (List 28) had poor test-retest reliability (correlation = 0.30, $p > 0.5$). Although the remaining 28 lists could arguably be used as equivalent with sufficient test-retest reliability, the authors suggest using a more stringent criterion of test-retest correlations > 0.65 . This results in a set of 25 equivalent lists (Lists 2, 4, 5, 7, 8, 9, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 24, 25, 26, 29, 30, 31, 32, 33, 34) with high test-retest reliability and phonetic balance for use with CI recipients who may reach ceiling on other sentence measures.

For many CI recipients, ceiling effects can limit the effectiveness of the commonly used HINT sentences. Mean HINT sentence scores for three groups of adult CI recipients described by Gifford et al (2008) ranged from 84.8% to 94.1%. Dorman and colleagues (1997) reported mean HINT sentence recognition scores of 94% for normal-hearing individuals listening through four channels of simulated CI processing. In a similar experiment, Loizou et al (1999) used 135 TIMIT sentences from the North Midland American English dialect region (half spoken by males and half by females) as sentence stimuli. Mean sentence recognition scores were 63%, substantially lower than those reported by Dorman et al (1997). In Fu et al (2002), both HINT and TIMIT sentences were presented to two of three Nucleus 22 study participants. Whereas, HINT scores were high for both participants (96.2% and 100% for N3 and N7, respectively), TIMIT scores were approximately 37 percentage points lower (58.9% and 63.2% for N3 and N7, respectively) as shown in Table III of Fu et al (2002). Collectively, these studies illustrate the strong effect that speaker variation has on speech recognition and support the need for such variation in test measures used with CI recipients to obtain a more representative indication of how CI recipients understand speech in everyday life.

There have been recent efforts by clinicians working with CI recipients to move towards more difficult and varied test materials. Fabry et al (2009) recommended a clinical test battery that included the AzBio sentences presented in quiet and noise to CI recipients in order to prevent ceiling effects observed with the HINT sentences and to better represent how recipients perform in daily life. To this end, the three cochlear implant manufacturers, Advanced Bionics, Cochlear, and MED-EL have provided clinics with an updated test battery, the New Minimum Speech Test Battery (MSTB) for Adult Cochlear Implant Users 2011, to be used both pre- and post-operatively. The New MSTB includes the AzBio sentences in place of the HINT sentences.

The AzBio sentences incorporate the use of multiple talkers who were instructed to speak in a conversational style rather than a clearly enunciated style. The sentences range in length from 4 – 10 words, use a total of 4 talkers (2 female and 2 male), all with standard American dialect. In addition, the AzBio sentences approximate a fourth grade reading level compared to an approximate first grade reading level for HINT sentences and sixth grade reading level for TIMIT sentences. The difficulty of the test material should be considered by both clinicians and researchers.

In a study of relatively high performing CI subjects, Spahr and Dorman (2004) reported mean HINT scores of approximately 95% compared to 75% for AzBio sentences. Gifford et al (2008) also found that CI recipients did not have the same ceiling effect with AzBio sentences as with HINT sentences. Given the recent trend toward use of AzBio sentences to evaluate potential CI candidates as well as monitor performance, the authors thought it would be informative to test the current group of participants with AzBio sentences. Twenty-one of the participants in the current study returned and were administered 33 lists of AzBio sentences in quiet at 60 dB SPL. For each participant, the mean TIMIT sentence score across lists (for the 25 lists found to be equivalent) at the first test session (Test 1) was compared to the mean AzBio sentence score across lists. Means ranged from 53% to 89% for TIMIT sentences and from 69% to 98% for AzBio sentences. On average, these participants scored higher on AzBio sentences (mean = 88%, SD = 9%) than TIMIT sentences (mean = 72%, SD = 11%), $t(20) = -10.21$, $p < 0.001$. In fact, half the participants scored higher than 90% on AzBio sentences. See Appendix B for more detail.

The TIMIT and AzBio sentences both offer the benefit of having varied speakers who do not over enunciate the test sentences. Still, ceiling effects may be seen for higher performing CI recipients when AzBio sentences are presented in quiet at an average conversational level. The 25 TIMIT sentence lists found to be equivalent in the present study may offer an alternative to evaluate performance outcomes for CI users with significant auditory-only speech recognition, particularly when testing in quiet is necessary, for example to differentiate effects of programming options such as new processing strategies or pre-processing algorithms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Support:

NIH/NIDCD K23DC 05410 and R01DC009010

Department of Otolaryngology at Washington University School of Medicine

References

- Alkaf FM, Firszt JB. Speech Recognition in Quiet and Noise in Borderline Cochlear Implant Candidates. *J Am Acad Audiol*. 2007; 18:872–882. [PubMed: 18496996]
- Balkany T, Hodges A, Menapace C, Hazard L, Driscoll C, Gantz B, Kelsall D, Luxford W, McMenemy S, Neely JG, Peters B, Pillsbury H, Roberson J, Schramm D, Telian S, Waltzman S, Westerberg B, Payne S. Nucleus Freedom North American clinical trial. *Otolaryngol Head Neck Surg*. 2007; 136:757–762. [PubMed: 17478211]
- Beattie RC. Word recognition functions for the CID W-22 test in multitalker noise for normally hearing and hearing-impaired subjects. *J Speech Hear Disord*. 1989; 54:20–32. [PubMed: 2915523]
- Boothroyd, A.; Hanin, L.; Hnath, T. A sentence test of speech perception: reliability, set equivalence, and short term learning. New York: City University of New York; 1985.
- Buss E, Pillsbury HC, Buchman CA, Pillsbury CH, Clark MS, Haynes DS, Labadie RF, Amberg S, Roland PS, Kruger P, Novak MA, Wirth JA, Black JM, Peters R, Lake J, Wackym PA, Firszt JB, Wilson BS, Lawson DT, Schatzer R, D’Haese PS, Barco AL. Multicenter U.S. Bilateral MED-EL Cochlear Implantation Study: Speech Perception over the First Year of Use. *Ear Hear*. 2008; 29:20–32. [PubMed: 18091099]
- Byrd D. Preliminary results on speaker-dependent variation in the TIMIT database. *J Acoust Soc Am*. 1992; 92:593–596. [PubMed: 1512325]
- Davis, H.; Silverman, SR. Hearing and Deafness. 4. New York: Holt, Reinhart and Winston; 1978.
- Dorman MF, Gifford RH, Spahr AJ, McKarns SA. The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiol Neurotol*. 2008; 13:105–112. [PubMed: 18057874]
- Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J Acoust Soc Am*. 1997; 102:2403–2411. [PubMed: 9348698]
- Dorman, MF.; Loizou, PC.; Spahr, A.; Dana, CJ. Simulations of combined acoustic/electric hearing. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology; 2003. p. 199-2001.
- Dorman MF, Spahr AJ, Loizou PC, Dana CJ, Schmidt JS. Acoustic Simulations of Combined Electric and Acoustic Hearing (EAS). *Ear Hear*. 2005; 26:371–380. [PubMed: 16079632]
- Dunn CC, Tyler RS, Oakley S, Gantz BJ, Noble W. Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. *Ear Hear*. 2008; 29:352–359. [PubMed: 18453885]
- Fabry D, Firszt J, Gifford R, Holden L, Koch D. Evaluation speech perception benefit in adult cochlear implant recipients. *Audiology Today*. 2009 May-Jun;:36–40.
- Firszt JB, Holden LK, Reeder RM, Skinner MW. Speech Recognition in Cochlear Implant Recipients: Comparison of Standard HiRes and HiRes 120 Sound Processing. *Otol Neurotol*. 2009; 30:146–152. [PubMed: 19106769]
- Firszt JB, Holden LK, Skinner MW, Tobey EA, Peterson A, Gaggl W, Runge-Samuelson CL, Wackym PA. Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear Hear*. 2004; 25:375–387. [PubMed: 15292777]
- Firszt JB, Reeder RM, Skinner MW. Restoring hearing symmetry with two cochlear implants or one cochlear implant and a contralateral hearing aid. *J Rehabil Res Dev*. 2008; 45:749–768. [PubMed: 18816424]
- Flesch, R. How to Write Plain English. New York: Haper and Row; 1979.
- Fu QJ, Shannon RV, Galvin JJ III. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *J Acoust Soc Am*. 2002; 112:1664–1674. [PubMed: 12398471]
- Gifford RH, Dorman MF, McKarns SA, Spahr AJ. Combined electric and contralateral acoustic hearing: word and sentence recognition with bimodal hearing. *J Speech Lang Hear Res*. 2007; 50:835–843. [PubMed: 17675589]

- Gifford RH, Shallop JK, Peterson AM. Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol Neurootol*. 2008; 13:193–205. [PubMed: 18212519]
- Gifford RH, Dorman MF, Shallop JK, Sydlowski SA. Evidence for the expansion of adult cochlear implant candidacy. *Ear Hear*. 2010; 31:186–194. [PubMed: 20071994]
- Holden, LK.; Finley, CC.; Holden, TA.; Brenner, CA.; Heydebrand, G.; Firszt, JB. Factors affecting cochlear implant outcomes. Poster presentation at the 2011 Conference on Implantable Auditory Prostheses; July 25, 2011; Pacific Grove, CA. 2011.
- Holden LK, Reeder RM, Firszt JB, Finley CC. Optimizing the perception of soft speech and speech in noise with the Advanced Bionics cochlear implant system. *Int J Audiol*. 2011; 50:255–269. [PubMed: 21275500]
- Holden LK, Skinner MW, Fourakis MS, Holden TA. Effect of increased IIDR in the nucleus freedom cochlear implant system. *J Am Acad Audiol*. 2007; 18:777–793. [PubMed: 18354886]
- Hosoi H, Murata K, Ohta F, Imaizumi S. Effect of the rate of speech flow on speech intelligibility in normal and hearing-impaired subjects. *Nippon Jibiinkoka Gakkai Kaiho*. 1992; 95:517–525. [PubMed: 1602353]
- Kincaid, JP.; Fishburne, RP.; Rogers, RL.; Chissom, BS. Research Branch report. Memphis: Naval Air Station; 1975. Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel; p. 8-75.
- Kirk KI, Pisoni DB, Miyamoto RC. Effects of stimulus variability on speech perception in listeners with hearing impairment. *J Speech Lang Hear Res*. 1997; 40:1395–1405. [PubMed: 9430759]
- Kirsch, IS.; Jungeblut, A.; Jenkins, L.; Kolstad, A. Adult Literacy in America. Washington, D.C: U.S. Department of Education, Office of Educational Research and Improvement; 2002.
- Koch DB, Osberger MJ, Segel P, Kessler D. HiResolution and conventional sound processing in the HiResolution bionic ear: using appropriate outcome measures to assess speech recognition ability. *Audiol Neurootol*. 2004; 9:214–223. [PubMed: 15205549]
- Kopra LL, Blosser D, Waldron DL. Comparison of Fairbanks Rhyme Test and CID auditory test W-22 in normal and hearing-impaired listeners. *J Speech Hear Res*. 1968; 11:735–739. [PubMed: 5719227]
- Lamel FL, Kassel RH, Seneff S. Speech database development: design and analysis of the acoustic-phonetic corpus. Proceedings of DARPA Speech Recognition Workshop, Report No SAIC-86\1546. 1986
- Linguistic Data Consortium. TIMIT Acoustic-Phonetic Continuous Speech Corpus. University of Pennsylvania; 1992. <http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1>
- Loizou PC, Dorman M, Poroy O, Spahr T. Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution. *J Acoust Soc Am*. 2000; 108:2377–2387. [PubMed: 11108378]
- Loizou PC, Dorman M, Tu Z. On the number of channels needed to understand speech. *J Acoust Soc Am*. 1999; 106:2097–2103. [PubMed: 10530032]
- Luxford WM. Ad Hoc Subcommittee. Minimum speech test battery for postlingually deafened adult cochlear implant patients. *Otolaryngol Head Neck Surg*. 2001; 124:125–126. [PubMed: 11226944]
- Mackersie CP. Tests of speech perception abilities. *Curr Opin Otolaryngol and Head Neck Surg*. 2002; 10:392–397.
- Microsoft Word. Bellevue, WA: Microsoft Corporation; 2007.
- Mullennix JW, Bihon T, Brickleyer J, Gaston J, Keener JM. Effects of variation in emotional tone of voice on speech perception. *Lang Speech*. 2002; 45:255–283. [PubMed: 12693687]
- Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. *J Acoust Soc Am*. 1989; 85:365–378. [PubMed: 2921419]
- Nilsson, M.; McCaw, V.; Soli, SD. Minimum Speech Test Battery for Adult Cochlear Implant Users: User Manual. Los Angeles: House Ear Institute; 1996.
- Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*. 1994; 95:1085–1099. [PubMed: 8132902]

- Noble W, Tyler R, Dunn C, Bhullar N. Unilateral and bilateral cochlear implants and the implant-plus-hearing-aid profile: comparing self-assessed and measured abilities. *Int J Audiol*. 2008; 47:505–514. [PubMed: 18608531]
- Pearsons, KS.; Bennett, RL.; Fidell, S. Speech levels in various environments. Canoga Park, CA: 1976.
- Peterson GE, Lehiste I. Revised CNC lists for auditory tests. *J Speech Hear Disord*. 1962; 27:62–70. [PubMed: 14485785]
- Riss D, Arnoldner C, Baumgartner WD, Kaider A, Hamzavi JS. A New Fine Structure Speech Coding Strategy: Speech Perception at a Reduced Number of Channels. *Otol Neurotol*. 2008
- Shannon RV, Galvin JJ, Baskent D. Holes in hearing. *J Assoc Res Otolaryngol*. 2002; 3:185–199. [PubMed: 12162368]
- Skinner MW, Holden LK, Holden TA, Demorest ME, Fourakis MS. Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants. *J Acoust Soc Am*. 1997; 101:3766–3782. [PubMed: 9193063]
- Skinner MW, Holden LK, Fourakis MS, Hawks JW, Holden T, Arcaroli J, Hyde Martyn. Evaluation of equivalency in two recordings of monosyllabic words. *J Am Acad Audiol*. 2006; 17:350–366. [PubMed: 16796301]
- Sommers MS. Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *J Acoust Soc Am*. 1997; 101:2278–2288. [PubMed: 9104029]
- Sommers MS, Barcroft J. Stimulus variability and the phonetic relevance hypothesis: effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification. *J Acoust Soc Am*. 2006; 119:2406–2416. [PubMed: 16642853]
- Spahr AJ, Dorman MF. Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. *Arch Otolaryngol Head Neck Surg*. 2004; 130:624–628. [PubMed: 15148187]
- Spahr AJ, Dorman MF. Effects of minimum stimulation settings for the Med El Tempo+ speech processor on speech understanding. *Ear Hear*. 2005; 26:2S–6S. [PubMed: 16082262]
- Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *J Speech Hear Res*. 1996; 39:494–509. [PubMed: 8783129]
- Vermeire K, Van de Heyning P. Binaural Hearing after Cochlear Implantation in Subjects with Unilateral Sensorineural Deafness and Tinnitus. *Audiol Neurootol*. 2008; 14:163–171. [PubMed: 19005250]
- Wilson BS, Dorman MF. Cochlear implants: A remarkable past and a brilliant future. *Hear Res*. 2008; 242:3–21. [PubMed: 18616994]
- Wilson RH, Carter AS. Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *J Am Acad Audiol*. 2001; 12:7–14. [PubMed: 11214979]

Abbreviations

CI	Cochlear Implant
CID	Central Institute for the Deaf
CNC	Consonant-Vowel Nucleus-Consonant
CUNY	City University of New York
dB	decibels
EAS	acoustic and electric stimulation
FM	frequency-modulated
HINT	Hearing In Noise Test
HL	hearing level

MSTB	Minimal Speech Test Battery
MIT	Massachusetts Institute of Technology
SD	standard deviation
SPL	sound pressure level
TI	Texas Instruments
WUSM	Washington University School of Medicine

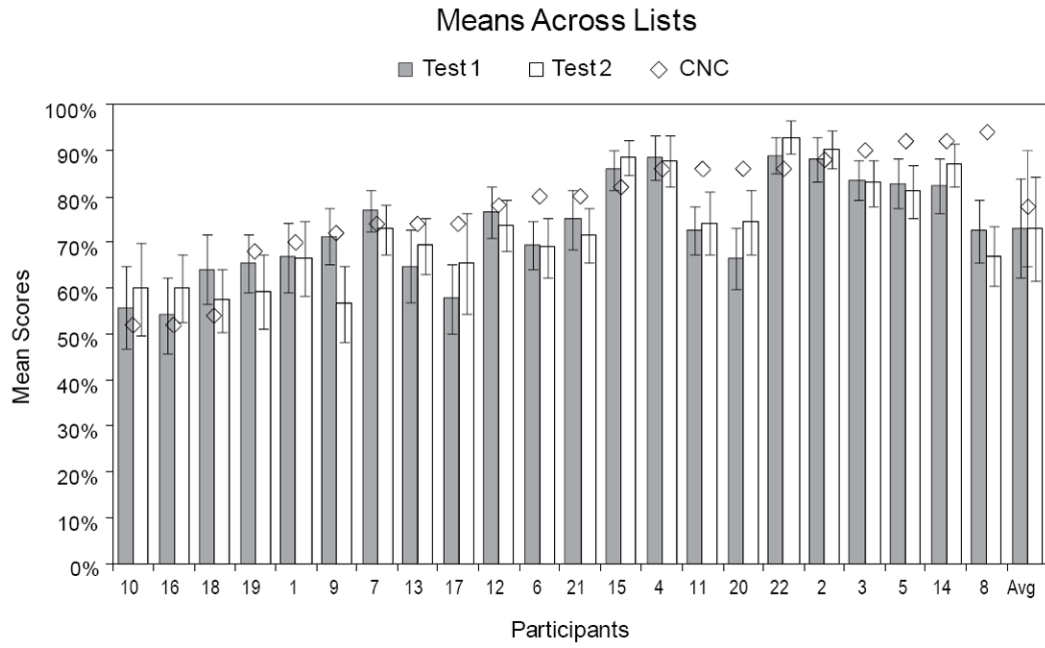


Figure 1. Mean TIMIT sentence scores in percent correct across lists for each of the 22 participants. Mean scores for Test 1 are displayed as gray bars and for Test 2 as white bars. CNC word scores are shown as the open diamonds. Participants are ranked from low to high on CNC word scores. Average scores across participants are shown to the far right. Error bars are ± 1 standard deviation of the mean.

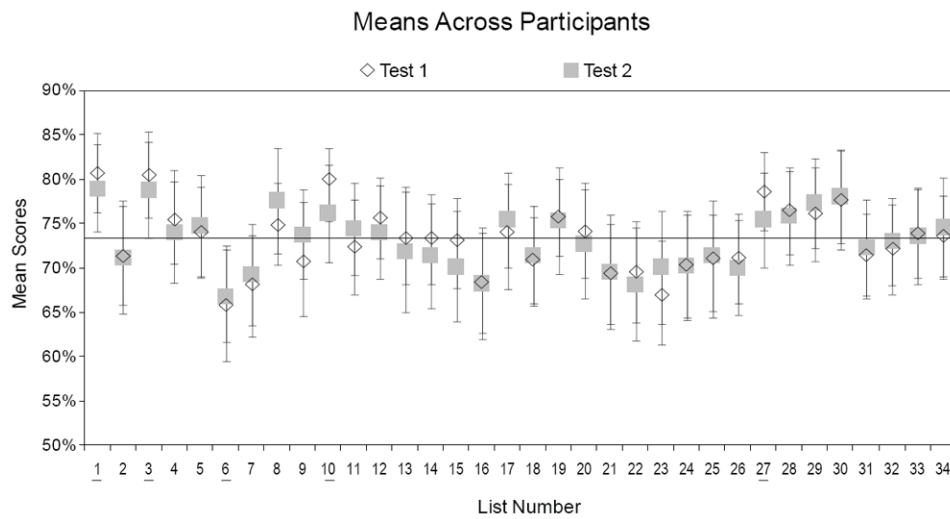


Figure 2. Group mean sentence scores in percent correct for each of the 34 TIMIT lists. Mean scores for Test 1 are displayed as open diamonds and for Test 2 as gray squares. Error bars indicate the 95% confidence intervals. A horizontal line indicates the overall average of 73% that is used as a benchmark for identifying equivalent lists. Lists that did not include this benchmark in the 95% confidence interval for both Test 1 and Test 2 (1, 3, 6, 10 and 27) are underlined.

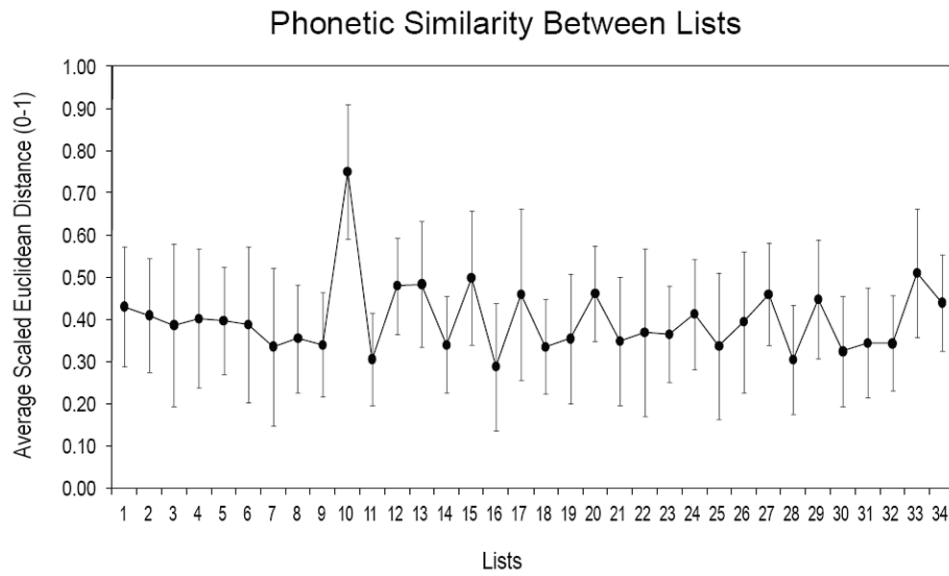


Figure 3. Average scaled Euclidean distance (0 to 1) of each list from all other lists arranged in TIMIT list number order. Lower values indicate increased similarity. Error bars are ± 1 standard deviation.

Table 1

Participant Demographic Information

Participant	Gender	Etiology	AAT	LOIU (yrs)	LOD (yrs)	CNC (%)
1	M	Usher II	51	5.0	33.9	44
2	F	Unknown	46	7.5	13.3	66
3	F	Otosclerosis	58	4.0	10.1	78
4	F	Unknown	46	2.8	13.0	86
5	F	Genetic	52	4.2	5.4	90
6	F	Genetic	57	6.2	12.8	71
7	M	Unknown	72	8.2	4.1	78
8	M	Genetic	55	10.6	5.9	84
9	M	Otosclerosis	74	5.1	14.4	82
10	F	Genetic (AIED)	48	4.5	5.5	48
11	F	Unknown	38	1.2	2.7	74
12	F	Genetic	70	1.1	10.2	83
13	F	Unknown	72	1.6	15.5	81
14	M	MS	49	3.4	3.7	86
15	M	Genetic	71	1.5	4.0	92
16	F	Unknown	54	2.2	28.9	59
17	M	Noise	65	3.0	4.1	58
18	F	Unknown	78	3.7	2.3	62
19	F	Unknown	75	2.9	0.4	58
20	F	Unknown	58	0.8	6.8	68
21	F	Ototoxicity	25	2.2	12.1	79
22	F	Genetic	57	3.9	10.6	80

Note: AAT = age at test; LOD = length of auditory deprivation; LOIU = length of implant use; CNC = most recent word score on Consonant Vowel-Nucleus Consonant test; AIED = Autoimmune Inner Ear Disease; MS = Multiple Sclerosis

Table 2

Cochlear Implant Device Information

Participant	Ear	Internal Device	Processor	Strategy (Maxima)	Rate (pps/ch)
1	L	N24	ESPrit 3G	SPEAK (9)	250
2	R	N24	ESPrit 3G	SPEAK (10)	250
3	L	N24C	ESPrit 3G	ACE (8)	1800
4	L	N24C	ESPrit 3G	ACE (8)	900
5	R	ABCII	Auria	HiRes-S	2175
6	R	N24	ESPrit 3G	ACE (12)	900
7	L	N24	Sprint	ACE (8)	1800
8	R	N22	ESPrit 3G	SPEAK (8)	250
9	L	N24C	Sprint	ACE (8)	1800
10	L	N24C	ESPrit 3G	ACE (8)	1800
11	R	N24CA	Freedom	ACE (10)	1800
12	R	NF	Freedom	ACE (10)	1200
13	R	AB90K	Auria	HiRes-S	1406
14	R	N24C	ESPrit 3G	ACE (8)	900
15	L	N24C	Freedom	ACE (10)	1200
16	L	N24C	ESPrit 3G	ACE (8)	900
17	L	N24C	ESPrit 3G	ACE (8)	1800
18	L	N24C	Freedom	ACE (11)	1200
19	R	NF	Freedom	ACE (10)	1200
20	R	AB90K	Auria	HiRes-S	2855
21	L	N24CA	ESPrit 3G	ACE (12)	1200
22	R	ABCII	PSP	HiRes-S	1024

Note: N24 = Nucleus 24; N24C = Nucleus 24 Contour; ABCII = Advanced Bionics CII; N22 = Nucleus 22; N24CA = Nucleus 24 Contour Advance; NF = Nucleus Freedom; AB90K = Advanced Bionics HiRes 90K;

Table 3

Summary of Descriptive Information for Each TIMIT Sentence List

List	Female Talkers	Male Talkers	Dialectical Regions	Unique Talkers	Total Words	Mean Words	Mean Age	Min Age	Max Age	Grade Level
1	10	10	4	20	142	7.1	30.9	23	64	5.1
2	12	8	4	20	132	6.6	26.5	21	37	6.7
3	11	9	3	19	131	6.5	23.0	64	32	4.3
4	12	8	4	20	134	6.7	31.6	23	67	4.8
5	12	8	4	20	119	5.9	29.4	20	67	5.7
6	6	14	4	20	124	6.2	28.6	23	46	3.9
7	8	12	4	19	131	6.5	29.8	24	47	4.8
8	10	10	4	19	127	6.3	28.5	22	42	5.3
9	6	14	4	19	134	6.7	28.4	22	44	5.6
10	8	12	4	20	129	6.4	30.7	24	51	7.4
11	11	9	4	20	129	6.4	28.8	22	35	4.7
12	11	9	3	20	123	6.1	31.4	23	51	7.3
13	10	10	4	19	136	6.8	28.5	23	57	5.6
14	12	8	4	19	121	6.0	26.5	22	39	8.1
15	11	9	4	20	131	6.5	28.5	23	51	5.9
16	12	8	4	20	120	6.0	29.1	23	52	5.8
17	12	8	3	19	113	5.5	27.5	22	42	4.2
18	7	13	4	20	121	6.0	32.4	23	64	7.3
19	13	7	4	19	118	5.9	29.3	22	67	6.0
20	14	6	4	19	123	6.1	30.3	22	64	7.8
21	9	11	4	20	130	6.5	31.2	22	67	4.4
22	12	8	4	20	129	6.4	28.2	23	45	6.4
23	6	14	4	20	126	6.3	28.1	22	44	5.9
24	8	12	4	19	128	6.4	29.4	22	51	5.6
25	7	13	4	20	129	6.4	31.0	22	64	5.0
26	13	7	4	20	135	6.7	33.5	22	67	4.9
27	14	6	4	20	133	6.6	30.1	23	48	5.0
28	14	6	4	19	130	6.5	29.3	23	47	5.7
29	13	7	4	19	134	6.7	29.6	20	51	6.3

List	Female Talkers	Male Talkers	Dialectical Regions	Unique Talkers	Total Words	Mean Words	Mean Age	Min Age	Max Age	Grade Level
30	10	10	4	18	131	6.5	30.2	22	57	6.6
31	10	10	4	18	126	6.3	30.6	22	67	4.8
32	8	12	4	19	130	6.5	27.3	21	46	4.8
33	13	7	4	19	126	6.3	29.2	23	57	7.3
34	8	12	4	18	133	6.6	31.1	22	57	6.3
Mean	10.38	9.62	3.91	19.41	128.18	6.38	29.37	23.53	52.55	5.74
SD	2.45	2.45	0.29	0.66	5.98	0.31	1.94	7.21	10.49	1.09

Note: Dialectical Regions = the number of dialectical regions represented; Mean Words = the average number of words per sentence; Mean Age = the mean age of the talkers; Min Age = the minimum age of the talkers; Max = the maximum age of the talkers; Grade Level = the Flesch-Kincaid reading grade level as calculated by Microsoft Word 2007; SD = standard deviation.

Table 4

Test-Retest Correlations for Each TIMIT Sentence List

List	Correlation	List	Correlation
1	0.726	18	0.777
2	0.734	19	0.692
3	0.640	20	0.640
4	0.810	21	0.746
5	0.700	22	0.826
6	0.804	23	0.623
7	0.878	24	0.835
8	0.799	25	0.823
9	0.784	26	0.655
10	0.734	27	0.911
11	0.838	28*	0.307
12	0.618	29	0.851
13	0.815	30	0.796
14	0.837	31	0.886
15	0.772	32	0.658
16	0.817	33	0.730
17	0.799	34	0.665

Note:

* The correlation for list 28 was not significant.