LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands — Source link

Astrid Van Wieringen, Jan Wouters

Institutions: Katholieke Universiteit Leuven

Published on: 01 Jan 2008 - International Journal of Audiology (Taylor & Francis)

Topics: Speech Reception Threshold Test and Noise

International Journal of Audiology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713721994

LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands

Astrid van Wieringen [a]; Jan Wouters [a]
[a] ExpORL, Dept. Neurosciences, K.U. Leuven, Belgium

First Published: June 2008

To cite this Article: van Wieringen, Astrid and Wouters, Jan (2008) 'LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands', International Journal of Audiology, 47:6, 348 — 355

To link to this article: DOI: 10.1080/14992020801895144
URL: http://dx.doi.org/10.1080/14992020801895144

informa
healthcare

Research Article

International Journal of Audiology 2008; 47:348–355

Astrid van Wieringen
Jan Wouters

ExpORL, Dept. Neurosciences, K.U. Leuven, Belgium

Key Words
Speech intelligibility
Sentences
Numbers

LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands

Abstract

A Dutch sentence test (LIST) and a Dutch number test (LINT) have been developed and validated for the accurate measurement of speech reception thresholds (SRT) in quiet and in noise with severely hearing-impaired individuals and cochlear implant recipients in Flanders and the Netherlands. The LIST consists of 35 lists of 10 sentences of equal known difficulty uttered by a female speaker, while the LINT consists of 400 numbers (1–100) by two male and two female speakers. Normative values were determined at fixed S/N ratios and using the adaptive method (Plomp & Mimpen, 1979), yielding identical results for SRT and slope. For the LIST, average fitted SRTs were 27.1 (0.9) dB SPL in quiet and −7.8 dB (0.2) SNR in noise. In addition, the LIST in noise displayed a steep discrimination function (17%/dB) and good reliability (within-subject standard deviation = 1.2 dB). For the LINT average fitted SRTs in quiet were 20.7 (0.9) dB SPL and about −9.0 dB SNR in noise. Again, the slopes of the performance intensity functions were relatively steep, i.e. 8.5%/dB in quiet and 15.2%/dB in noise, suggesting that the LINT is accurate and efficient and thus capable of reflecting subtle changes in performance. First data with cochlear implanted subjects show that both LIST and LINT are feasible and are capable of mapping a large range of hearing disabilities.

Sumario

Una prueba holandesa de frases (LIST) y una prueba holandesa de números (LINT) han sido desarrolladas y validadas para la medición exacta de los umbrales de recepción del lenguaje (SRT) en silencio y en ruido, en individuos con alteración auditiva severa y en receptores de implante coclear, en Flandes y en los Países Bajos. El LIST consiste de 35 listas de 10 frases con igual dificultad conocida presentadas por una hablante femenina, mientras que el LINT consiste de 400 números (1–100) y es presentado por dos hablantes masculinos y dos femeninas. Se determinaron los valores normativos a tasas S/R fijas y utilizando el métodos adaptativo (Plomp & Mimpen, 1979), rindiendo resultados idénticos en cuanto al SRT y la pendiente. Para el LIST, los SRT amplificados promedio fueron 27.1 (0.9) dB SPL en silencio y −7.8 dB (0.2) SNR en ruido. Además, el LIST en ruido mostró una función de discriminación empinada (17 %/dB) y una buena confiabilidad (desviación estándar intra-sujeto = 1.2 dB). Para el LINT, los SRT promedio en silencio con amplificación fueron de 20.7 (0.9) dB SPL y alrededor de −9.0 dB de SNR en ruido. De nuevo, las pendientes de las funciones de desempeño/intensidad fueron relativamente empinada, p.e., 8.5 %/dB en silencio y 15.2 %/dB en ruido, sugiriendo que el LINT es exacto y eficiente, y por lo tanto, capaz de cambios sutiles en el desempeño. Los primeros datos con sujetos con implante coclear muestran que tanto el LIST como el LINT son factibles y capaces de mapear una amplio rango de discapacidades auditivas.

Introduction

For many years sentence intelligibility of hearing-impaired persons in Flanders, the Dutch speaking part of Belgium, has been evaluated with Dutch sentence materials developed in the Netherlands. The most well-known Dutch sentence material is that of Plomp and Mimpen (1979). Because of its extensive use and because several words were outdated, Versfeld et al (2000) developed a much larger set of sentence materials (39 lists of 13 sentences for one male and one female speaker), also known as the 'VU-sentences'. Although these materials are also used in Flanders for evaluation of signal processing algorithms and/or evaluation of mildly hearing-impaired persons, they are not suitable for measuring sentence intelligibility with cochlear implantees or other severely hearing-impaired persons. With the VU-sentences intelligibility in quiet, let alone in noise, is very difficult for cochlear implantees, possibly because of the conversational speaking rate or other suprasegmental aspects. Moreover, some words or sentence structures are typical of Dutch spoken in the Netherlands, and therefore more difficult to understand for the Dutch speaking population in Flanders. Therefore, based on a growing demand, a new set of sentences has been recorded and validated in Flanders with the purpose of developing a test that accurately quantifies speech intelligibility for a large range of hearing abilities. In order to be able to reflect subtle changes in performance at different signal-to-noise ratios the speech materials were optimized with the steepest possible slope at the intercept (50%) for speech intelligibility in noise and in quiet. In addition, the lists of sentences were made equivalent to demonstrate good test-retest reliability (Plomp & Mimpen, 1979) and to be efficient. Testing efficiency increases with increasing slope of the psychometric function, and depends on the redundancy of the speech materials. Since speech materials become less difficult when they are repeated or reused, this requirement also implies that the speech materials must be different for each trial.

ISSN 1499-2027 print/ISSN 1708-8186 online
DOI: 10.1080/14992020801895144

Received:
January 6, 2008
Accepted:
January 6, 2008

Astrid van Wieringen
ExpORL Dept., K.U. Leuven, O & N 2, Herestraat 49/721, B-3000 Leuven, Belgium.
E-mail: astrid.vanwieringen@med.kuleuven.be

Together with the Leuven intelligibility sentences test (LIST), the Leuven intelligibility number test (LINT) was developed. A number test is suitable for intelligibility studies with severely hearing-impaired users because of its simplicity, closed response format, and lack of learning effect. In addition, limited linguistic competence is required, making the test suitable for speakers of a foreign language, people with limited education, or those with a mild mental handicap.

Both the LIST and the LINT were evaluated using the fixed method (fixed signal-to-noise ratios) and the more efficient adaptive method with varying signal-to-noise ratios. The adaptive method accurately measures the speech reception threshold (SRT), the presentation level necessary for a listener to recognize speech materials correctly 50% of the time. It is a very suitable measure for rapid speech intelligibility testing, since, unlike the fixed method, it is not subject to floor and ceiling effects. An important aim of the present study was to determine whether both methods reflect the same level of performance and are equally reliable under adverse listening conditions.

Methods

Selection criteria of speech materials

The selection criteria for the sentences were the same as described by Plomp and Mimpen (1979). Sentences represented conversational speech, were short enough to repeat, contained a verb and noun, were not too redundant (no proverbs), and were not too difficult or confusing (no questions or exclamations). An additional criterion for the present materials was that the vocabulary had to be common in both Flanders and in the Netherlands. A set of 730 sentences was recorded for the LIST, while the numbers 1 to 100 were chosen for the LINT.

Speakers and recordings

Four experienced speakers with background in speech therapy, two males (JW, MD) and two females (WD, AG), each produced the 730 sentences. They were instructed to articulate well, as if speaking to a severely hearing-impaired person, and to avoid placing undue emphasis on any of the words.

Recordings of sentences and numbers were made in a double-walled sound-proof booth in the Laboratory of Acoustics and Thermal Physics of the Department of Physics (KU Leuven). The speaker was seated in front of a B&K 4165 microphone. Recordings were routed through a B&K 2639 pre-amplifier and a B&K 2610 amplifier to a digital studio DAT-recorder TASCAM DA-30. They were sampled at 44 100 Hz (16 bit A/D converter). Lists of numbers (1–100) were produced by the same four experienced speakers (WD, AG, JW, MD) and under the same circumstances as the sentences. The numbers were recorded in random order, to avoid possible order effects.

Editing, speech rate, and noise

For the LIST the sentence speech material of only one female speaker (WD) was evaluated perceptually, while for the LINT the numbers of all four speakers were taken into account. For both tests the best token of each sentence was selected and edited in Praat (Boersma, 2001), and 30-ms cosine windows were applied at the beginning and end of the utterance to avoid clicks. Sentences were not equated by level adjustment, because the average RMS of the recorded sentences only varied within 3 dB, and because the perceptual equalization was considered more important. However, numbers were scaled to their average RMS before the first perceptual evaluation.

All speech sounds were stored as '.wav' files on the hard disk of a computer. Two speech-weighted stationary noises, one based on the long-term average speech spectrum of 730 sentences of speaker WD, and one based on the average spectrum of all the numbers produced by the four speakers, were developed. The use of a spectrally matched masker ensures that, on average, the S/N ratio will be approximately equal at all frequencies. The final set of LIST sentences consisted of 350 of the 730 sentences, but the noises generated for these two different sets were spectrally identical.

Subjects, test set-up, and calibration

Most subjects were students between 20 and 25 years of age. All were screened for normal hearing (<20 dB HL for octave frequencies from 125 Hz to 8000 Hz). The subject was seated in a quiet room and heard the sentences monaurally through TDK headphones. The sentences were played directly from a computer using the software interface APEX (Laneau et al, 2005) and passed through an audiometer. Sentence levels were adjusted manually during adaptive testing. The noise was routed from a CD player via the audiometer to the same earpiece of the headphone. The levels of the speech and noise were calibrated with a Type 4152 artificial ear. For the LIST subjects were instructed to listen and repeat aloud whatever was heard or understood. For the LINT subjects were asked to identify the numbers from a closed-set of response alternatives. No feedback was provided, and subjects were paid for their participation. Answers were recorded automatically by APEX, and SRTs and slopes were determined by means of non-linear regression fits to a logistic function (SPSS, 2004).

LIST

First evaluation and construction of new corpus for LIST

The aim of the study was to select a subset of sentences that are equally intelligible under the same adverse conditions. In the first phase of the study half of the 730 sentences were administered together with the speech-weighted noise at −8 dB, and half at −10 dB signal-to-noise ratios to 26 normally hearing persons. The noise level was presented continuously at 65 dB SPL. The values of −8 dB and −10 dB were based on a pilot experiment. The total testing time was 4–5 hours per subject.

Based on this pilot test a homogenous set of sentences was constructed by selecting sentences with a similar slope and a similar intelligibility score (those that straddled the 50% point between the −10 and −8 dB signal-to-noise ratios). The resulting 355 sentences were reduced to 350 sentences and cast into 35 lists of 10 sentences each. Figure 1 shows the distribution of the number of syllables per sentence of the 350 sentences. Care was taken that the total number of syllables per list was the same (90) and that the distribution of keywords (2–3 per sentence) was more or less equal per list (to enable syllable or keyword scoring). The final set of sentences consists of 23 lists containing 32 keywords each, and 12 lists containing 33 keywords each (a

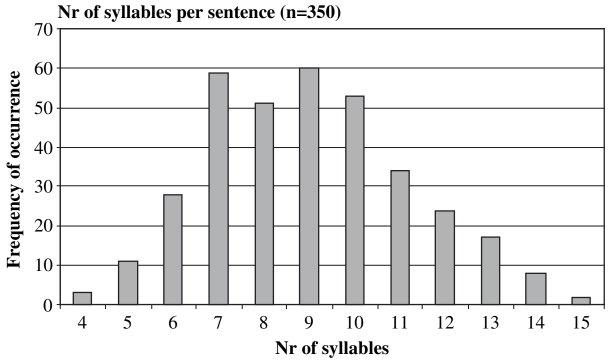

Figure 1. Distribution of number of syllables per sentences in the LIST.

total of 1132 keywords). The number of syllables per sentence in each list varied from low to high (to avoid abrupt differences in number of syllables per sentence). In addition, the first (or second, or third, etc.) sentence of each list of 10 sentences could be interchanged to avoid learning effects.

Phonemic transcription The remaining 350 sentences were transcribed phonemically to verify that the different Dutch phonemes were distributed equally over the 35 lists. The frequencies of occurrence of 8520 phonemes are listed as percentages in Table 1 in descending order. The ranking of the phonemes of this distribution compares well with previous phoneme counts of Dutch (van den Broecke, 1988). Moreover, none of the 35 pairwise comparisons between the phoneme distributions of an individual list of sentences were statistically significantly different from the total percentage of occurrence (as listed in Table 1).

Second evaluation of LIST

The remaining corpus was administered to 24 new subjects using both the fixed and the adaptive procedures in quiet and in noise.

Fixed method: Sentences in quiet and in noise A pilot experiment with sentences in quiet showed intelligibility scores to range between 0% and 100% at the following intensity levels: 35, 33, 31, 29, 27, 25, and 23 dB SPL. The 35 lists of sentences were divided over the seven levels, and each of seven normal-hearing subjects identified each list of sentences at one of the seven levels. For speech in noise the speech-weighted noise was presented at 65 dB SPL and the signal-to-noise ratios were −5 dB, −6 dB, −7 dB, −8 dB, −9 dB, −10 dB, and −11 dB. Once again the 35 lists were divided among the seven levels and presented to seven different normal-hearing subjects.

Adaptive method: Sentences in noise With the adaptive method the level of the noise was held constant at 65 dB SPL. Starting at 55 dB SPL, the level of the first sentence of each list was increased in steps of 2 dB until the sentence was identified correctly. Subsequently, the intensity level within a list of sentences varied in steps of 2-dB adaptively in a one-down, one-up procedure to target the 50% intercept (cf. Plomp & Mimpen, 1979). After determining the level of the (imaginary)

Table 1. Frequency and percent frequency of occurrence of the phonemes of the LIST, in descending order.

	8520	*100%*
Schwa	967	11.35
t	773	9.07
n	743	8.72
r	641	7.52
d	466	5.47
s	420	4.93
x	370	4.34
e	320	3.76
k	300	3.52
i	298	3.50
l	290	3.40
a	261	3.06
h	226	2.65
aa	224	2.63
ee	193	2.27
v	187	2.19
m	184	2.16
o	180	2.11
ei	172	2.02
w	170	2.00
oo	161	1.89
p	146	1.71
b	140	1.64
i	139	1.63
z	137	1.61
f	72	0.85
oe	63	0.74
ng	57	0.67
ui	43	0.50
u	40	0.47
uu	37	0.43
j	35	0.41
ou	19	0.22
eu	17	0.20
au	8	0.09
ai	5	0.06
g	5	0.06
ieu	5	0.06
oi	5	0.06
oei	1	0.01

11th item, the SRT was calculated on the basis of the last six levels. All 35 lists were presented to 10 new normal-hearing subjects.

Manner of scoring In each sentence the keywords are underlined (van Wieringen & Wouters, 2005). The sentence score is 1 if all of the keywords are identified correctly, otherwise 0. Errors of non-keywords are not taken into account, but incomplete keywords or minor variations of verb tenses of keywords are. For subjects with very poor sentence scores it is possible to determine either a keyword score (max = 32 or 33) or a syllable score (max = 90).

Table 2. Average SRTs (dB) and slopes of LIST, together with their precision values.

	Average SRT	*Precision*	*Average*	*Precision*
Quiet, fixed, fitted	27.1 dB SPL	±0.9 dB SPL	12.5%/dB	±1.7%/dB
Noise, fixed, fitted	−7.8 dB	±0.2 dB	17.5%/dB	±2.0%/dB
Noise, adaptive, fitted	−8.0 dB	±0.2 dB	17.8%/dB	±1.9%/dB
Noise, adaptive	−7.8 dB	±1.2 dB		

LINT

Contrary to the sentence test, where half of the original sentences were omitted, the number test has to contain all of the 100 recorded numbers per speaker. Therefore, intelligibility of numbers was equated by adjusting the RMS levels of different tokens of the four speakers in 4 successive experiments with normal-hearing subjects. Once equal intelligibility had been achieved, the 400 numbers were presented in quiet and in noise to 20 new normal-hearing subjects. For the perceptual evaluation the numbers were divided into 10 lists per speaker. Each ten occurred only once in a list (e.g. 15 and 17 did not occur in the same list). In addition, the same units (e.g. 45 and 55) did not occur more than once in a list. After evaluation the numbers were recorded in the same order on the audio CD. However, any order of presentation is possible if presented with computer software such as APEX (Laneau et al, 2005).

Numbers in quiet and in noise using the fixed and adaptive methods

Using the fixed method the forty lists of 10 numbers were presented in quiet at 30, 28, 26, 24, 23, 22, 21, 20, 19, 18 dB SPL to 10 normally hearing listeners. For evaluation of the LINT in noise, the numbers were presented at 10 different signal-to-noise ratios between −14 and −5 dB SNR in steps of 1 dB to 10 new normal-hearing subjects (the level of the speech-weighted noise was held constant at 65 dB SPL). With the adaptive method the same procedure was followed as with the LIST. Again, all 40 lists were presented to 10 new normal-hearing subjects.

Results

LIST

SRTs and slopes at 50% scores are based on non-linear regression fits to a logistic function (SPSS 12.0) of the performance intensity curve of each individual subject obtained at fixed levels in quiet and in noise with data averaged over lists. Moreover, as the adaptive procedure only yields a SRT value, the performance intensity function of each subject was also tracked on the basis of the scores at the different presentation levels. After determining fitted SRTs and slopes of each subject, the resulting SRTs and slope are the arithmetic average of the individually fitted SRTs and slopes of the different subjects. The precision (error bar) on both parameters is deduced from the quadratically averaged error bars of the fit to the data of each individual subject.

Table 2 shows that the SRTs for sentences in noise, expressed in dB signal-to-noise ratio are very similar for both the fixed and the adaptive procedures. There were no outliers with the fitted sentence scores. The fitted SRT from the measurements in noise at fixed levels and from the adaptive test procedure also correspond well with the average SRT derived from the adaptive test. The slopes at the 50% point are in the order of 17.5%/dB for sentences in noise and somewhat shallower for sentences in quiet (12.5%/dB). Performance intensity functions of sentences in quiet and in noise, using the fixed and adaptive methods are given in Figure 2.

List equivalency

Figure 3 illustrates the variation in SRT of the 35 lists determined with the adaptive method (not the fitted values). The values are plotted in terms of a deviation score from the overall mean, together with their respective standard deviations. A repeated measures ANOVA on the measured SRTs of the adaptive speech in noise task (35 lists × 10 listeners) showed no significant differences between lists [$F(1,9) = 2.1$, $p = 0.18$].

Reliability

The reliability of the SRT measurements was determined by considering the within-subject standard deviation of repeated measurements (Plomp & Mimpen, 1979). The within-subjects standard deviation is 1.17 dB for sentences in noise using the adaptive method. This value for a list of 10 sentences is in the same order of magnitude as within-subject standard deviations of other speech materials: 1.07 dB for 13 VU sentences per list (Versfeld et al, 2000); 0.9 dB for 13 Plomp sentences per list; 1.13 dB for 12 HINT sentences per list (Nilsson et al, 1994); 1.1 dB for 20 Canadian French sentences per list (Vaillancourt et al, 2005); and <1 dB for 10 Swedish sentences per list (Hällgren et al, 2006).

Comparison with literature

The fitted SRTs and slopes of the LIST in noise are of the same order of magnitude as those obtained by Plomp and Mimpen (1979) (i.e. −5.9 dB SNR, SD = 0.9 dB). The SRTs of Versfeld et al (2000) are somewhat higher (−3.1 dB SNR) and the corresponding slopes somewhat shallower (i.e. 11.6%/dB and 11.8%/dB). The SRTs of the VU sentences in noise are in the same order of magnitude as the American-English HINT sentences (Nilsson et al, 1994), i.e. −2.9 dB SNR (SD = 0.78 dB). The SRT of the HINT in quiet is 23.9 dBA. Meanwhile, the HINT has been modified for different languages using the adaptive procedure. For the Canadian-French version of the HINT the mean SRT is 16.4 dBA in quiet, −3.3 dBA SNR (SD = 0.5 dB) (Vaillancourt et al, 2005). The average SRT in noise for the Swedish version of the HINT sentences is −3.0 dB SNR (SD = 1.1dB, Hällgren et al, 2006). For German a set of 200 sentences has been standardized (Kollmeier & Wesselkamp, 1997) by means of the fixed method, resulting in a SRT of −6.1 dB SNR with a steep slope of 20%/dB.

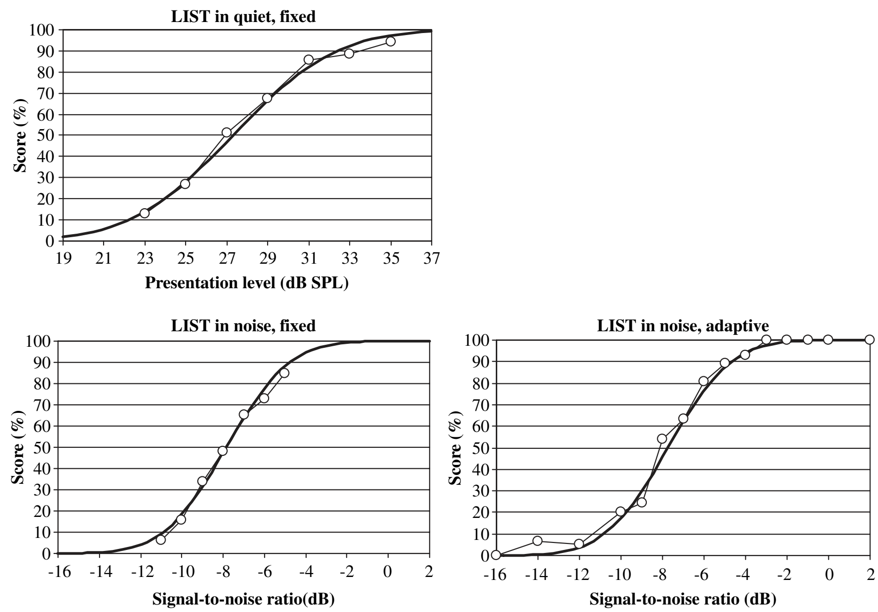

Figure 2. Measured and fitted performance intensity functions of the LIST in quiet and in noise with the adaptive and fixed methods

LINT

The average fitted SRTs of the LINT in quiet and in noise, expressed in dB SPL or in dB signal-to-noise ratio, are listed in Table 3, together with their respective slopes (%/dB) and precision values. Again, the precision (error bar) on both parameters is deduced from the quadratically averaged error bars of the fit to the data of each individual subject. In three of the six parameters, two of the 40 values were discarded from the average value because the fitted values showed extreme error bars due to bad fits to the data (>2 SD). Note that the average SRT of the adaptive analysis corresponds well with the fitted values. Performance intensity functions of numbers in quiet and in noise using the fixed and adaptive methods are given in Figure 4.

One-way analyses of variance (SPSS, 2004) were conducted on the fitted SRTs to investigate possible statistical differences per speaker. These analyses showed a significant effect of speaker for LINT in quiet [$F(3,36 = 12.1$, $p < 0.001$], but no effect of speaker for LINT in noise (adaptive) [$F(3,36 = 1.3$, $p = 0.29$], nor for LINT in noise (fixed) [$F(3,36 = 1.0$, $p = 0.4$]. The fitted SRTs in quiet were 21.7 dB SPL for speaker AG (female), 20.9 dB SPL for speaker WD (female), 19.8 dB SPL for speaker MD (male), and 20.4 dB SPL for speaker JW (male).

The SRTs in noise correspond well with those of the Dutch automatic speech-in-noise screening test developed in the Netherlands (Smits et al, 2004). With headphones average SRTs in noise of this digit triplet test are −11.2 dB with a slope of 16%/dB.

LIST EQUIVALENCY

Figure 5 illustrates the SRTs determined in noise with the adaptive procedure (not the fitted values) for the 10 lists of digits. Data of the four speakers are given in terms of a deviation score from the overall mean, together with their respective standard deviations. A two-way repeated measures ANOVA showed that neither the factor ‘speaker’ [$F(3,27) = 1.4$, $p = 0.26$], nor ‘list’[$F(9,81) = 2.03$, $p = 0.19$], nor the interaction of list ×speaker was statistically significant. The fitted SRTs in both noise conditions fall within ±1.2 dB of the overall mean. The data of the LINT in quiet display more variability (between 1.5 and −2.5 dB of the overall mean).

RELIABILITY

The standard deviation of the SRT between lists of numbers, averaged over the different normal-hearing listeners is 1.44 dB for numbers in noise using the adaptive method. As a comparison the standard deviation of difference in repeated measurements of the Dutch triplet-in-noise screening test is 1.3 dB under headphones (Smits et al, 2004).

Validation of speech materials with cochlear implantees

The speech materials of the LIST and the LINT have been developed for evaluation of speech in noise with severely hearing-impaired persons. Speech intelligibility results in quiet

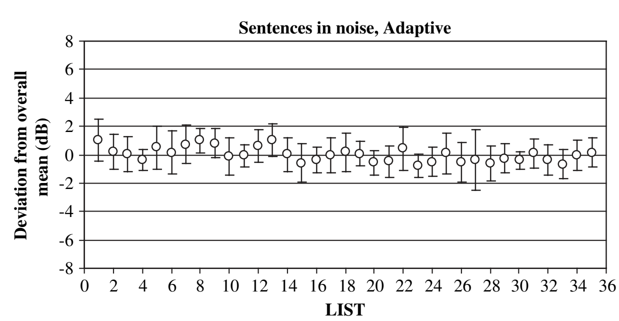

Figure 3. SRTs of the 35 lists (+standard deviations) of the LIST determined with the adaptive procedure. Data are averaged over subjects and plotted in terms of a deviation score from the overall mean.

Table 3. Average SRTs (dB SNR) and slopes of LINT.

	Average SRT	*Precision*	*Average slope*	*Precision*
Quiet, fixed, fitted	20.7 dB SPL	±0.9	8.5%/dB	±2.5%/dB
Noise, fixed, fitted	−9.9 dB	±0.5	15.2%/dB	±3.6%/dB
Noise, adaptive, fitted	−9.2 dB	±0.4	15.2%/dB	±2.6%/dB
Noise, adaptive	−8.9 dB	±1.4		

and in noise for 16 adult cochlear implantees implanted in University Hospital UZLeuven are shown in Figure 6. Sentences scores in quiet were obtained by means of the open response format in APEX at 65 dB SPL, while speech intelligibility in noise was assessed by means of the adaptive procedure. Each value is the average of the score of 5 lists. An important aim of the current research is to map the hearing abilities of both good and poor performers using the same test materials. First data collected with the LIST and LINT show that this aim has been met. Intelligibility scores in quiet for the 16 cochlear implantees ranged between 42% and 100%. Figure 6 also shows that some implantees perform very well in quiet and in noise (above 90% in quiet, and SRT between 0 and 5 dB SNR for LIST in noise), while other good performers in quiet perform relatively poorly in noise. Two subjects found the sentence in noise task too difficult (indicated at 16 dB SNR). Note that the SRT in noise of the best CI performer was approximately 9 dB higher than that of normal-hearing subjects.

Number intelligibility in quiet and in noise are shown in Figure 7 (n = 24). The same cochlear implant subjects listened to the female speaker WD (n = 11, unfilled circles), and the male speaker JW (n = 13, crosses). They were asked to type the digit they heard. Again, each value is the average of the score on five lists of numbers. While all subjects are able to achieve high scores for numbers in quiet (80% and higher), number intelligibility in noise ranges from −6 dB to +6 dB SNR. Note that the SRT in noise of the best CI performer is approximately 4 dB higher than that of normal-hearing subjects.

Discussion

An important objective of the sentence test is its ability to document the hearing abilities of severely hearing-impaired persons who are poor and good performers. Therefore, the speakers were instructed to speak clearly, as if to a hearing-impaired person. Previous studies have shown intelligibility to be enhanced by producing clear speech (e.g. Liu et al, 2004). One of the acoustic consequences is that the overall rate is slower for clear speech than for conversational speech. The overall speaking rate of the LIST is 2.5 syllables/second, about half the speaking rate of the Dutch VU sentences (4.7 syllables/second). As mentioned above, the VU sentences in quiet are too difficult for the severely hearing-impaired. The LIST in quiet proved to be feasible for both poor and good performers. The task can be made more difficult by adding noise. Despite the relatively low speaking rate, a high homogeneity of sentence difficulty both within each test list and across all test lists was achieved. Due to the relatively steep slope, the SRTs can be determined accurately and efficiently.

SRTs in noise and slopes of the LINT are very similar to those of a Dutch digit triplets test developed and validated in the Netherlands (Smits et al, 2004). Similarly, Ramkissoon et al

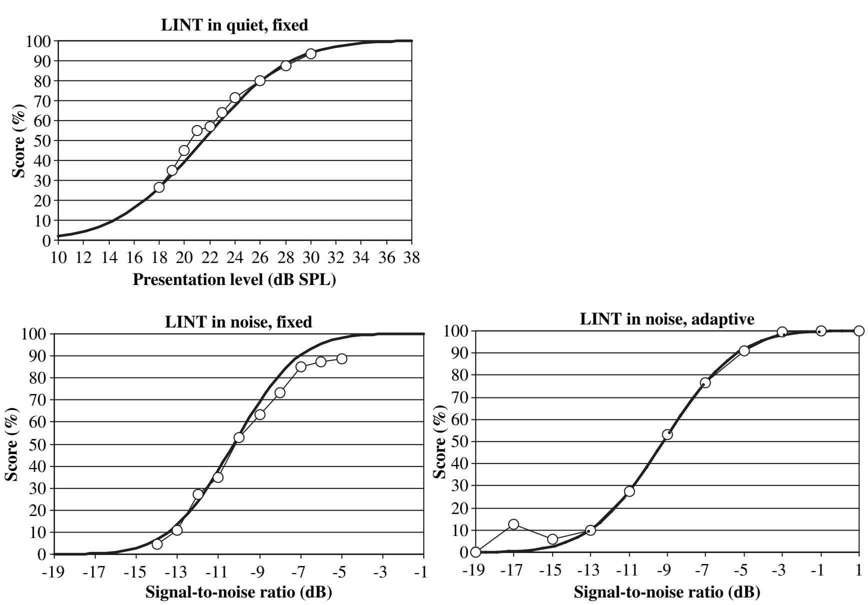

Figure 4. Measured and fitted performance intensity functions of the LINT in quiet and in noise with the adaptive and fixed methods

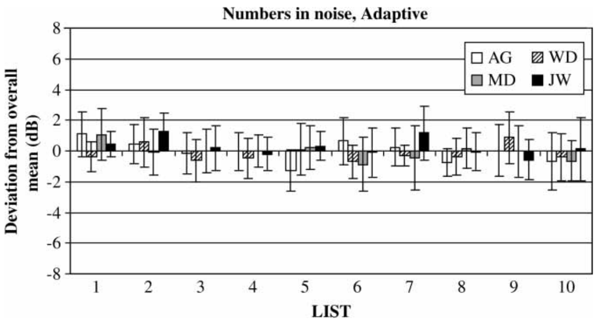

Figure 5. SRTs of the 10 lists (+standard deviations) for the four different speakers of the LINT determined with the adaptive procedure. Data are averaged over subjects and plotted in terms of a deviation score from the overall mean.

(2002) reported SRTs in noise of a Finnish digit test of −10.9 dB SNR for native English speakers and −11.9 dB SNR for non-native English speakers (not validated). It is expected that SRTs of digits tests are lower than those of sentence tests due to the difference in response format, i.e. closed versus open (e.g. Miller et al, 1951).

Conclusions

The LIST consists of 35 lists of 10 sentences that are representative of daily communication and that are of equivalent and known difficulty. Normative data of speech intelligibility parameters have been obtained with normal-hearing subjects by using both the fixed and adaptive methods. Data are similar, allowing either method to be chosen for future research. The advantage of using the adaptive method is that the speech reception threshold (SRT) is not subject to floor and ceiling effects. It is possible to record only keywords (32 or 33 per list) or syllables (total of 90), but this is only recommended when sentence scoring is not possible.

The LINT consists of 400 numbers (1–100) by four speakers. All lists of numbers by the four speakers are equated for intelligibility in quiet and noise. Again, data obtained with the fixed and adaptive methods are similar. Recordings of the LIST and LINT (van Wieringen & Wouters, 2005) are available.

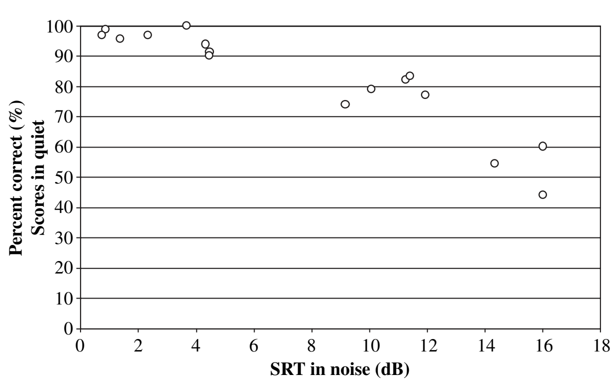

Figure 6. LIST intelligibility in quiet (% correct) and SRT in noise (dB) of 16 cochlear implantees. Each data point is based on five lists of sentences.

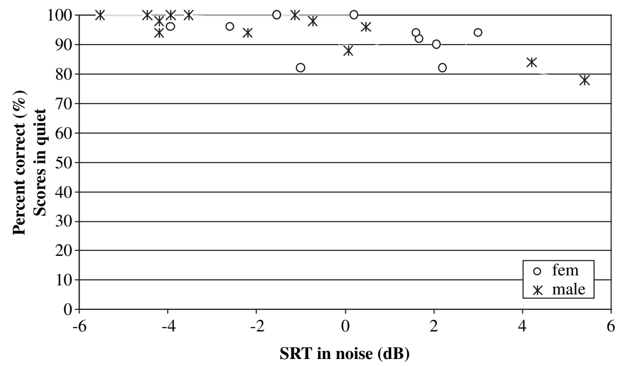

Figure 7. LINT intelligibility in quiet (% correct), and SRT in noise (dB), of 24 cochlear implantees: Eleven subjects listened to the female speaker (WD, unfilled circles), 13 to the male speaker (JW, crosses). Each value is the average of the score of five lists of numbers. Two subjects yield identical results for the female speaker (100% in quiet, −1.53 dB in noise), which is why only 10 circles are visible.

First data collection with cochlear implantees shows that a large range of hearing abilities can be mapped using these speech materials. While 100% scores can be obtained, test conditions can always be made more difficult. These speech materials open avenues for precise and efficient testing of speech reception with severely hearing-impaired subjects, e.g. to optimize new speech processing strategies. In the near future the speech materials will be validated for different types of noises and a subset of sentences that are suitable for children will be selected and validated.

Acknowledgements

The authors would like to thank all those who participated in the development of the LIST and LINT: Wivine Decoster, Ann Goeleven, Marc De Bodt, and Jan Wolf, Nele Corstjens, Sofie Debussche, Mieke Goossens, Veerle Hufkens, Koen Eneman, Tom Francart, Johan Laneau, Jean-Baptiste Maj, Jeff Vanden Berghe, Leen Vermeulen, Lot Van Deun, Valerie Vandenbroeck, and Jessie Vandeput. Ellen Boon is gratefully acknowledged for administering the sentence and number tests to the cochlear implantees. This research was funded by the Fonds voor Wetenschappelijk Onderzoek FWO-Vlaanderen (FWO G.0233.01), and the Research Council of the Katholieke Universiteit Leuven (OT/99/33, OT/03/58).

References

Boersma, P. 2001. Praat: a system for doing phonetics by computer. *Glot International*, 5, 9/10, 341–345.

Hällgren, M., Larsby, B. & Arlinger, S. 2006. A Swedish version of the hearing in noise test (HINT) for measurement of speech recognition. *Int J Audiol*, 45(4), 227–237.

Laneau, J., Boets, B., Moonen, M., Wieringen, A. van. & Wouters, J. 2005. A flexible auditory research platform using acoustic or electric stimuli for adults and young children. *J Neurosc Meth*, 142, 131–136.

Liu, S., Del Rio, E., Bradlow, A.R. & Zeng, F-G. 2004. Clear speech perception in acoustic and electric hearing. *J Acoust Soc Am*, 116(4), 2374–2383.

Kollmeier, B. & Wesselkamp, M. 1997. Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J Acoust Soc Am*, 102, 2412–2421.

Miller, G.A., Heise, G.A. & Lichten, W. 1951. The intelligibility of speech as a function of the context of the test materials. *J Exp Psych*, 41, 329–340.

Nilsson, M, Soli, S.D. & Sullivan, J.A. 1994. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*, 95, 1085–1099.

Plomp, R. & Mimpen, A.M 1979. Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18, 43–52.

Ramkissoon, I., Proctor, A., Lansing, C.R. & Bilger, R.C. 2002. Digit speech recognition thresholds (SRT) for non-native speakers of English. *Am J Audiol*, 11, 23–28.

Smits, C., Kapteyn, Th.S. & Houtgast, T. 2004. Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol*, 43, 15–28.

SPSS 12.0, 2004. SPSS Inc, Chicago, USA.

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., et al. 2005. Adaptation of the HINT (hearing in noise test) for adult Canadian Francophone populations. *Int J Audiol*, 44, 358–369.

Van den Broecke, M. 1988. Frequenties van letters, lettergrepen, woorden en fonemen in het Nederlands. In M.P.R. van den Broecke (ed.), *Ter Sprake: spraak als betekenisvol geluid in 36 thematische hoofdstukken*. Fortis Publications.

Van Wieringen, A. & Wouters, J. 2005. LIST en LINT, Nederlandstalige spraakaudiometrielijsten met zinnen en getallen, realisatie Lab.Exp.ORL-NKO, K.U. Leuven, CD SIG0501/2.

Versfeld, N.J., Daalder, L., Festen, J.M. & Houtgast, T. 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, 107, 1671–1684.