



Development of a Danish speech intelligibility test

Nielsen, Jens Bo; Dau, Torsten

Published in:
International Journal of Audiology

Link to article, DOI:
[10.1080/14992020903019312](https://doi.org/10.1080/14992020903019312)

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Nielsen, J. B., & Dau, T. (2009). Development of a Danish speech intelligibility test. *International Journal of Audiology*, 48(10), 729-741. <https://doi.org/10.1080/14992020903019312>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Jens Bo Nielsen
Torsten Dau

Centre for Applied Hearing Research,
Technical University of Denmark,
Lyngby, Denmark

Key Words

Speech intelligibility assessment
Equalization of intelligibility
HINT
Speech perception

Abbreviations

CLUE: Conversational language
understanding evaluation
HINT: Hearing in noise test
HL: Hearing level
RMS: Root-mean-square
SNR: Signal-to-noise ratio
SRT_N: Speech recognition threshold
in noise
SPL: Sound pressure level
SI: Sentence intelligibility
WI: Word intelligibility

Development of a Danish speech intelligibility test

Abstract

A Danish speech intelligibility test for assessing the speech recognition threshold in noise (SRT_N) has been developed. The test consists of 180 sentences distributed in 18 phonetically balanced lists. The sentences are based on an open word-set and represent everyday language. The sentences were equalized with respect to intelligibility to ensure uniform SRT_N assessments with all lists. In contrast to several previously developed tests such as the hearing in noise test (HINT) where the equalization is based on scored (objective) measures of word intelligibility, the present test used an equalization method based on subjective assessments of the sentences. The new equalization method is shown to create lists with less variance between the SRT_{NS} than the traditional method. The number of sentence levels included in the SRT_N calculation was also evaluated and differs from previous tests. The test was verified with 14 normal-hearing listeners; the overall SRT_N lies at a signal-to-noise ratio of -3.15 dB with a standard deviation of 1.0 dB. The list-SRT_{NS} deviate less than 0.5 dB from the overall mean.

Sumario

Se ha desarrollado una prueba danesa de inteligibilidad del lenguaje para evaluar el umbral de reconocimiento del lenguaje en ruido (SRT_N). La prueba consta de 180 frases distribuidas en 18 listas fonéticamente balanceadas. Las frases están basadas en palabras de contexto abierto y representan el lenguaje cotidiano. Las frases fueron ecualizadas con respecto a la inteligibilidad para asegurar una evaluación uniforme del SRT_N en todas las listas. En contraste con varias pruebas previamente desarrolladas, tales como la prueba de audición en ruido (HINT), donde la ecualización está basada en medidas calificadas (objetivas) de inteligibilidad de la palabra, esta prueba usó un método de ecualización basado en evaluaciones subjetivas de las frases. El nuevo método de ecualización parece crear listas con menos variancia entre el SRT_{NS} que el método tradicional. El número de niveles de frase incluidos en el cálculo del SRT_N también fue evaluado y difiere de pruebas previas. La prueba fue verificada con 14 sujetos normo-oyentes; el SRT_N global se encuentra con una relación señal-ruido de -3.15 dB con una desviación estándar de 1.0 dB. El SRT_{NS} de la lista se desvía menos de 0.5 dB de la media global.

Understanding speech is a fundamental human ability and listening to spoken language is probably the most important application of our hearing. Therefore, methods for a reliable assessment of speech perception capabilities are essential, particularly when hearing difficulties are suspected. An assessment must take into account that a hearing loss can affect speech intelligibility through at least two distinctly different sub-effects: (1) attenuation of all sounds entering the ear, and (2) distortion of the perceived sounds (Plomp, 1978). Most hearing impairments are a combination of attenuation and distortion, but the distribution between the two parts varies from individual to individual. The effect of attenuation can be fully compensated by an increase in the overall sound pressure level, whereas the intelligibility loss due to distortion can only be compensated by an increase in the signal-to-noise ratio (SNR) (Middelweerd et al, 1990). Distortion can represent a considerable handicap in everyday situations because much speech communication takes place where the speech-to-noise ratio is low (Plomp, 1978). For a group of hearing-impaired listeners, the pure-tone thresholds at 0.5, 1.0, 2.0, and 4.0 kHz are usually correlated with the speech intelligibility performance. However, a reliable prediction of speech intelligibility cannot be made for the individual listener as only the attenuation part of the hearing impairment is measured directly in a tone audiogram. The audiogram is an inadequate method for predicting speech intelligibility, especially in noise,

where the intelligibility is more affected by distortion than by attenuation (e.g. Glasberg & Moore, 1989; Middelweerd et al, 1990).

Various speech intelligibility tests that take distortion effects into account have been developed over the last decades. The earliest tests were based on short words presented in noise (e.g. Fairbanks, 1958; House et al, 1965), and the intelligibility score was calculated as the percentage of correctly repeated words. While including important features for the assessment of speech intelligibility, notably real speech stimuli and background noise, these tests have not been ideal for assessing a listener's ability to follow a natural conversation. The short, individually recorded words do not include many of the characteristics of natural speech, such as word transitions, reductions, contractions, temporal fluctuations, and intonation (Nilsson et al, 1994). Also, the listener's ability to exploit the redundancy as well as the semantic and syntactic cues in natural speech is not taken into account. Furthermore, word tests are not suited for more advanced testing and fitting of hearing aids, since the compression and noise-reduction algorithms do not take full effect with isolated single words (Nilsson et al, 1994). Examples of intelligibility tests using sentence-length stimuli are the speech perception in noise test (SPIN) (Kalikow et al, 1977), the Hagerman-type test (Hagerman, 1982), and the hearing in noise test (HINT) (Nilsson et al, 1994). The Hagerman-test was

originally developed in Swedish and consists of five-word sentences constructed according to a fixed scheme: a name, a verb, a number, an adjective, and finally a noun. The Hagerman-type test is also available in Danish (DANTALE II; Wagener et al, 2003). This test is suitable for extensive testing because the sentences are semantically unpredictable and difficult to memorize (Wagener et al, 2003). However, the sentences are also unnatural and nonsensical, and significant learning effects have been observed (Nilsson et al, 1994). Nilsson developed the HINT, which is a speech intelligibility test with natural sentences that comprise the pronunciation and content characteristics of conversational speech. The HINT mimics everyday speech communication and the test is sensitive to most of the speech perception problems encountered by the hearing impaired. The sentences are syntactically different and based on an open word set, which reduces the training effect compared to tests with a closed word set and a fixed sentence structure (Nilsson et al, 1994).

The outcome of a HINT measurement is usually the *speech recognition threshold in noise* (SRT_N), which is equal to the SNR at which the listener is able to correctly repeat 50% of the presented speech material. The use of the SRT_N effectively eliminates the risk of floor and ceiling effects where, respectively, 0% or 100% of the material is correctly identified. In the HINT, the SRT_N is measured using a sentence-based *adaptive* procedure where the SNR is decreased when the listener was able to repeat the whole sentence correctly, and increased when only part of the sentence was recognized.

The adaptive sequence is relatively short and the stimulus consists of 10 or 20 sentences taken from a set of pre-compiled lists (e.g. Vaillancourt et al, 2005; Hällgren et al, 2006). These lists must be of equal difficulty to ensure stable SRT_N assessments. Additionally, it is a requirement for an adaptive procedure that the sentences within each list are of equal intelligibility. Otherwise, the SRT_N assessment will be unreliable and the test will be insensitive to small differences in the SRT_N between listeners or conditions (MacLeod & Summerfield, 1990). Equalization of the sentence intelligibilities is therefore an essential part of the test development process.

Sentences with equal overall root-mean-square (RMS) levels cannot be expected to be equally intelligible in noise, since word familiarity, short-term level variations, intonation, etc., will cause deviations (Nilsson et al, 1994). Therefore, a two-step process has been employed in the equalization of sentence intelligibility in several sentence tests (e.g. Plomp & Mimpfen, 1979; Nilsson et al, 1994; Vaillancourt et al, 2005; Wong & Soli, 2005; Hällgren et al, 2006). In this process, the first step is to determine the intelligibility of all sentences that are candidates for the test. The sentences are presented to a number of listeners at various SNRs and their responses are recorded. In the second step, the intelligibility variations found in step one are compensated by an adjustment of the RMS levels of the individual sentences. The RMS level of sentences with low intelligibility is raised and that of sentences with high intelligibility is lowered. This adjustment exploits that the intelligibility of a sentence in noise is very sensitive to the SNR, thereby making it possible to compensate for intelligibility deviations by manipulating the SNR.

In step one of the equalization process, a considerable number of listeners is needed to obtain reasonably precise estimates of

the sentence intelligibilities. In an attempt to increase the efficiency of the equalization procedure, previous HINT development projects have employed *word* scoring, noting the number of correctly repeated words in each sentence, instead of *sentence* scoring, noting only whether the whole sentence was correctly repeated or not. Compared to sentence scoring, word scoring increases the amount of collected data significantly. However, the use of word scoring in the equalization process has a severe side effect. The RMS adjustment of the sentences will no longer be based on the *sentence* intelligibility (SI), but on the average *word* intelligibility (WI) of each sentence. The WI for a given sentence is calculated as the number of correctly repeated words divided by the total number of words. In the studies on HINT (e.g. Nilsson et al, 1994; Vaillancourt et al, 2005; Hällgren et al, 2006), it was assumed that sentences with equal WIs also have equal SIs, hence allowing an equalization of the SI of a corpus of sentences by ensuring that all sentences have the same WI. However, sentences with equal WI can indeed have significantly different SIs (see Appendix A for examples of sentences with equal WI, but SIs of 24% to 59%). This lack of proportionality between the SI and the WI is caused by the fact that the two entities have a different probabilistic relationship to the sentence. The SI is an 'AND' combination of what the listener repeats: all parts must be correctly identified in order to give the sentence a positive SI score. The WI is an 'OR' combination of what the listener correctly repeats: each individually identified word contributes positively to the WI score. The deviation between the SI and the WI will be particularly high when the distribution of the WI between the individual words in the sentence is very uneven. For example, if one word has a low intelligibility, the SI will also be low because it is seldom that all words in the sentence are understood. However, the words with high intelligibility will lead to a relatively high average WI.

It is essential that the SIs of the test developed here are equalized since the adaptive test procedure is based on sentence and not word scoring. This goal cannot be achieved with the use of word scoring in the equalization process, and the use of sentence scoring is extremely time-consuming due to the large number of sentence presentations that are needed to achieve data of sufficient validity. To solve this dilemma, an equalization procedure based on a 'just-follow-conversation' method (Kollmeier & Wesselkamp, 1997) was developed in the present study. The method builds on experiments where listeners are requested to adjust the individual SNRs of a number of sentences until the sentences are perceived as being equally intelligible. This method does not involve intelligibility tests with explicit sentence or word scoring. Instead, it involves subjective assessments of the sentences done by a group of listeners. The experiments of Kollmeier & Wesselkamp (1997) showed a correlation of 0.78 between the SRT_N found with the subjective method and previously conducted scoring experiments with the same sentences.

In the present study, the new equalization procedure was employed in the development of a speech intelligibility test for Danish. The objective was to develop a test with a minimal within-subject, between-list variation in the speech intelligibility assessments, without reducing the sensitivity to between-subject variability. Since the developed test deviates in some methodological aspects from the HINT, it is named 'conversational language understanding evaluation' (CLUE).

Methods

Sentence material

GENERAL SELECTION CRITERIA

The purpose of the present speech test is to evaluate a listener's ability to follow everyday conversational language. The sentence material for the test was created specifically with this purpose in mind. A set of 'objective' and a set of 'subjective' selection criteria were considered for the sentence material. The criteria are partly based on the criteria described in Versfeld et al (2000).

The objective criteria are as follows: (1) The number of words in each sentence is five. (2) The number of syllables in each sentence is 8–9. (3) Words do not contain more than four syllables. (4) Each sentence contains a verb. (5) The sentences are grammatically correct. (6) The sentences do not contain proper names. (7) Proverbs, exclamations, and questions are not allowed.

The subjective criteria are as follows: The sentences should represent conversational speech and should be (1) neutral, (2) meaningful, (3) natural, and not be (4) (too) redundant, (5) illogical, (6) out of a context, (7) characterized by bad sentiments, or (8) humorous.

The Danish Society for Language and Literature (Det Danske Sprog- og Litteraturselskab) has created a database named *Korpus 2000*, which contains sentences from Danish newspapers, magazines, books, etc., from the period 1998 to 2002. There are about 28 million words in the database, corresponding to more than two million sentences. The sentences were tested against the criteria for inclusion in the test. The 'objective criteria' (except for grammatical correctness) were implemented as MATLAB scripts, and 4075 five-word sentences were extracted from the database. A manual examination with respect to the 'subjective criteria' and grammar showed, however, that most of the sentences were not usable. Many were context dependent while others were characterized by (very) bad sentiments, e.g. 'Vores liv ligger i ruiner' (our life lies in ruins) and 'Der brændte bål i gaderne' (bonfires were burning in the streets). It was concluded that it would not be possible to extract enough usable sentences directly from Korpus 2000. A new corpus of sentences that fulfilled the listed objective and subjective criteria was written based on some of the five-word sentences extracted from the database. The sentences were then checked for naturalness and reoccurrences of words by the first author, and some were rewritten or discarded.

SELECTION OF THE TALKER

An initial recording of 25 sentences was made with 11 different talkers, four women and seven men. Their ages were 24–63 years. Five professional acousticians/audiologists reviewed the recordings of all 11 talkers. The objective was to find a talker with a natural pronunciation, close to the conversational speech that most Danes encounter in their everyday life. Seven talkers were rejected for various reasons, leaving four talkers still in consideration. The four recordings were assessed by two professional phoneticians. Based on their feedback, a 38 year old male talker with a background in phonetics, but with no previous experience in speech recordings, was selected.

RECORDING, EDITING, AND TRANSCRIPTION

The sentences were recorded in a double-walled sound-proof booth directly to a PC using a high-quality 24-bit sound card (RME DIGI96/8 PAD) and a sampling frequency of 44.1 kHz.

A 1-inch B&K condenser microphone (type 4179) with pre-amplifier (type 2660) was used to produce a recording with a low background noise level. The microphone was placed at a distance of approximately 30 cm from the mouth of the talker, symmetrically in the horizontal plane, and at an angle of approximately 45° upwards in the vertical plane. The recorded sentences were digitally high-pass filtered at a cut-off frequency of 50 Hz and split into individual waveforms. The recordings were all adjusted to an average RMS level of –26 dB (re: max. digital output) with maximum peak levels of approximately –5 dB. This allowed headroom for moderate level adjustments during the equalization process without the risk of clipping. The waveforms were stored as wav-files. Based on the sound files, the sentences were transcribed in the International Phonetic Alphabet (IPA). The transcription was done by a master's student in phonetics at the University of Copenhagen, who was highly experienced in transcribing spoken Danish.

THE BACKGROUND NOISE

A speech-shaped background noise was created to match the long-term frequency spectrum of the sentence material. Speech-shaped noise maximizes the slope of the psychometric function, hence increasing the accuracy of the SRT_N determination (Prosser et al, 1991). The speech-shaped noise will also lead, on average, to similar SNRs across frequencies, hence not intentionally favouring some speech frequencies over others. However, some variation in the SNR will occur across the frequency spectrum due to the varying duration of different speech sounds. For example, for short and intense high-frequency consonants, the mean level of these (which by definition is the level in the speech-shaped noise) is relatively low due to their short duration, but when actually present in the speech signal, they will have a level well above the mean.

The noise was created using a superimposing approach (Wagener et al, 2003). The sentence sound files were concatenated in random order and stored as an initial noise file. The files were then randomized in a new order and added to the noise file. The final noise file was the result of 150 superpositions. The noise had only little amplitude fluctuation and a frequency spectrum that matched the long-term spectrum of the sentences. The RMS level of the noise was adjusted to the same level as the sentences.

Equalization of sentence intelligibility

SUBJECTS

18 listeners (11 male, 7 female) participated in the equalization of the sentence intelligibility. Their ages were between 20 and 25 years with a mean of 22.8. Before participation, their audiograms were measured. All listeners had hearing thresholds of 15 dB HL or better in the range 0.125 to 8 kHz. All listeners were native Danish speakers and students at the Technical University of Denmark. They were paid on an hourly basis for their participation. All experiments in this study were approved by the ethics committee of Copenhagen County.

STIMULI

The stimuli consisted of a corpus of 322 five-word sentences and the corresponding speech-shaped noise. The noise was turned off between sentences in order to avoid exposure to a constant noise. The noise onset was 1 s before the sentence start and the offset 600 ms after the end of the sentence. The noise was ramped on and off

by a squared sine function with a ramp duration of 400 ms. The onset of the noise 1 s prior to the speech is believed not to create unintended onset effects and this timing has been used when determining normative data for the HINT in various languages.

APPARATUS AND PROCEDURE

The experiment took place in a sound-proof booth and stimuli were presented over Sennheiser HD580 headphones. The noise level was fixed at 65 dB SPL, whereas the speech level varied according to the listener's response. The experiment was controlled by a PC and a MATLAB application written for this specific purpose. The application presented the stimuli via the PC sound card to the headphones. An instructor was present in the booth for initial instruction and a short training session, but not during the experiment itself.

EXPERIMENTAL DESIGN

The equalization procedure of this study was fundamentally different from previous HINT studies. It was not attempted to objectively assess (score) the sentence or word intelligibility. Instead, the equalization was based on subjective judgements by the listeners. After each sentence presentation, the listener had the option to press one of three buttons: 'difficult', 'easy' or 'ok'. The listener was given written and oral instruction to press the buttons according to the following rules (translated from Danish):

- press 'difficult', if you did not understand the *whole* sentence
- press 'ok', if you were *just* able to understand the sentence
- press 'easy', if it was relatively *easy* to understand the sentence

The listeners were also instructed to create their own subjective criterion for how a sentence should sound to be 'just understandable', and only to press 'ok', when a sentence fulfilled this criterion. A test session with 12 sentences was run to help the listeners create the criterion before the actual equalization experiment started.

The sentences were presented in random order in a number of sequences. The first sequence consisted of all sentences, and the subsequent sequences consisted of all the sentences that had

been judged 'difficult' or 'easy' in the previous one. A press of the 'ok' button excluded the corresponding sentence from further presentation and stored the history of presentation levels for the sentence. The initial presentation level for all sentences was 63 dB SPL, corresponding to an SNR of -2 dB, the SNR at which listeners, on average, were expected to perceive the sentences as 'just understandable'. A press of 'difficult' raised the sentence level by 2 dB; a press of 'easy' lowered the level by 2 dB. This adjusted level was then used for the presentation of the same sentence in the following sequence. After one reversal, for example, when a sentence was judged 'easy' in one sequence and 'difficult' in the next, the step size of the level adjustment was halved to 1 dB. At the second reversal, the step size was reduced to 0.5 dB, where it then remained.

The equalization process was split into two experimental series. Ten listeners participated in the first series and determined an RMS level for each sentence that subjectively equalized the intelligibility. All sentences were then adjusted with the mean adjustments done in this series and a second experimental series with eight new listeners was conducted.

OUTCOME OF THE EQUALIZATION

The number of sentence presentations to each listener in the first series ranged from 686 to 1456, with an average of 994, corresponding to 3.1 presentations per sentence. The mean adjustments of the sentences are shown in the left panel of Figure 1, normalized to an overall adjustment of 0 dB. The adjustments to produce 'just understandable' sentences lie between -3.9 dB and 3.7 dB (averaged across listeners). The standard deviations lie between 0.44 and 2.45 dB.

In the second series, the number of presentations to each listener ranged from 701 to 1442; the average was 986, corresponding to 3.1 presentations per sentence as in the first series. The mean adjustments can be seen in the right panel of Figure 1. The adjustments in the second series are, on average, smaller than in the first series, indicating that the level adjustments obtained in the first series had a positive effect on equalizing the intelligibility of the sentences. However, the adjustments obtained in the second series still deviate

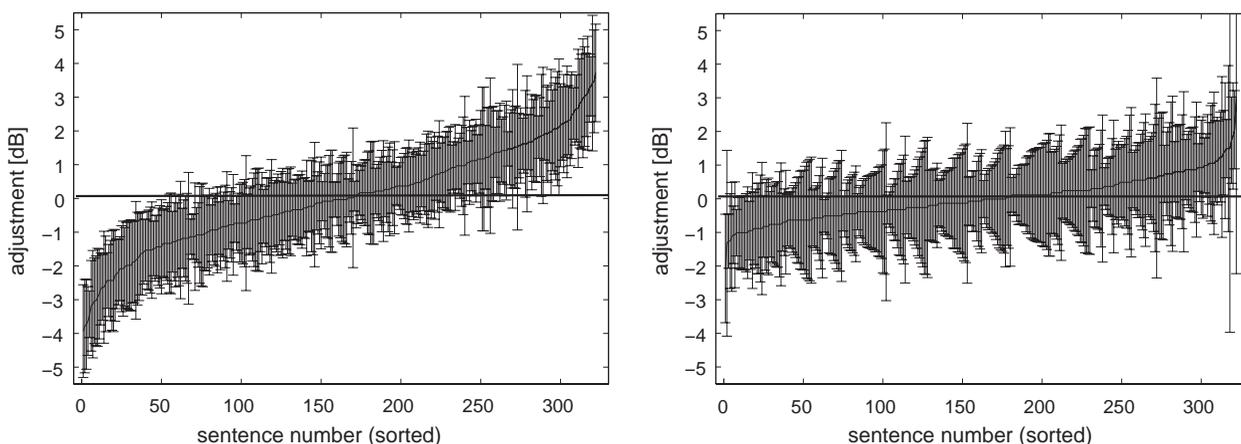


Figure 1. Mean adjustments of the sentence levels in the equalization process in the first experimental series (left) and in the second series (right). The adjustments are normalized to an overall adjustment of 0 dB. Error bars show ± 1 standard deviation. The 322 sentences are sorted with respect to the mean adjustment in each experiment. The adjustments in the first experiment are in the range -3.9 to 3.7 dB.

significantly from the 'baseline' at 0 dB for a considerable number of sentences. Hence, the level adjustments determined in the second series were also imposed on the sentence files. The final sentence levels are thus based on the adjustments done in both series.

The level adjustments of the sentences were the immediate result of the equalization process, but the outcome also led to the omission of several sentences. Sentences with a total level adjustment of more than ± 3 dB were omitted in order to avoid obvious level differences in the final test. Adjustments of more than ± 1 dB were not allowed in the second series because a large adjustment might just be a statistical coincidence and there would be no third series to reveal this. Sentences were also omitted when a listener, during the equalization process, required an SNR of 4 dB or more to comprehend the sentence. Such a high SNR implies an intelligibility flaw in the sentence. A minor, fixed adjustment was done to all sentences to ensure that the average RMS level was -26 dB (re: max. digital output).

List creation

The final sentence lists were created to be as phonetically balanced as possible. The sound inventory for the transcribed sentences consisted of 28 vowels (17 short and 11 long) and 20 consonants. The Danish 'stød' (a short glottal stop) and syllabic consonants were also transcribed and regarded as phones to be balanced. The overall phonetic distribution for all sentences was determined and a trial-and-error procedure distributed the sentences among the lists in order to hit this distribution as closely as possible for each list. 20 lists with 10 sentences each were created.

List verification

The main purpose of the test verification was to document that similar SRT_{NS} are obtained with the different sentence lists. The overall SRT_N for the test and its standard deviation were also determined.

SUBJECTS

The verification of the 20 test lists involved 14 (seven male, seven female) native Danish speaking listeners, between 19 and 32 years (mean 22.9). They all had hearing thresholds of 15 dB HL or better from 0.125 to 8 kHz.

APPARATUS AND PROCEDURE

The experiment took place in a sound-proof booth and the stimuli were presented over Sennheiser HD580 headphones. The noise level was fixed at 65 dB SPL, whereas the speech level varied according to the adaptive test procedure of Nilsson et al (1994). The noise onset and offset were controlled in the same manner as in the equalization procedure. A test leader was present during the whole experiment; he ran the test session using a tailor-made PC-application. The 20 test lists were presented to the listeners in a random order determined by the PC. Before the actual test lists, three training lists were run. This allowed the listeners to get used to the test procedure and the influence of training effects was reduced.

The listeners were asked to repeat all words in the sentences as precisely as possible, but also encouraged to guess the words that

they did not hear. The adaptive presentation procedure was as follows: The first sentence was presented repeatedly, starting at -8 dB SNR and increasing in 2-dB steps until the listener repeated the sentence correctly. The level of the remaining sentences was lowered by 2 dB after a correct repetition of the previous sentence, and raised by 2 dB after an incorrect repetition.

The PC application for running the test controlled the playback of the speech signal and the background noise and adapted the levels according to the test procedure. The test leader scored the sentences by pressing on-screen buttons according to the listener's response. The application created a data log of all sentence presentation levels during the test run.

RESPONSE VARIATIONS

The following general variations of listener responses were accepted during the verification: (1) Change in verb tense, (2) change in article, and (3) change between singular and plural nouns. Sentences were also considered correct if a word was added to the actual sentence. For example, 'Han lagde tasken på bordet' (he put the bag on the table) was accepted in the form 'Han lagde tasken op på bordet' (he put the bag up on the table). The following specific alternatives were also accepted: De/vi (they/we), hun/han (he/she), and min/din (my/your). In some cases these alternatives were mentioned spontaneously by the listeners.

Results

Calculation of the SRT_N

An analysis was conducted in order to determine the number of sentence levels to include in the SRT_N calculation. The difference between the presentation level of each sentence in the verification test ($n = 14$ listeners $\times 20$ lists $\times 11$ levels = 3080) and the average level for each list ($n = 14$ listeners $\times 20$ lists = 280) was computed. (Presentation level 11 results from the response to sentence 10, although the eleventh sentence does not exist.) Figure 2 shows the mean of these level differences (circles) as a function of the position in the list. From the fourth sentence on, the presentation level has stabilized around the average; the level for sentence 4 is slightly closer to the average than levels 5, 9, and 11. The standard deviation (squares) has a minimum for sentences 4 and 5, but varies very little as a function of the sentence position. The result that sentence 4 is closer to the reference level than several of the following sentences was also found in the studies of Nilsson et al (1994) and Hällgren et al (2006). As a result, it was decided here to include the last eight levels (4 to 11) in the SRT_N calculation. This is a change in the calculation compared to Nilsson et al (1994), who only included the levels of sentences 5 to 11.

List verification result

The following results all relate to SRT_N calculations based on sentence levels 4 to 11 in each list. For each of the 20 lists, an estimated list- SRT_N was calculated as the average SRT_N across listeners. In Figure 3, these values are plotted relative to the overall SRT_N , representing the mean SRT_N across all lists and listeners. The 95% confidence intervals of the estimates are also shown, i.e. the interval around the mean that, with a likelihood

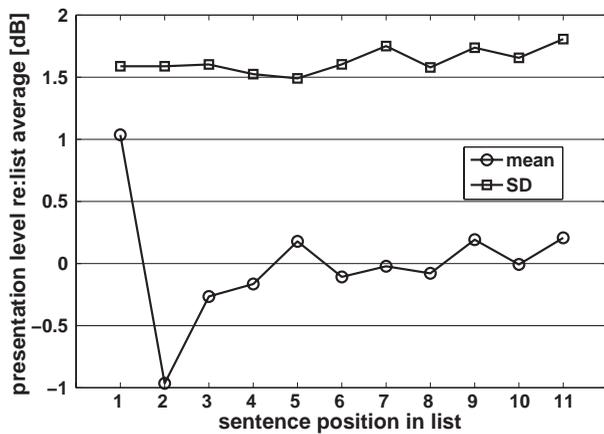


Figure 2. The presentation level mean and standard deviation across all lists and all listeners as a function of the sentence position in the list. The presentation level is given relative to the average level of each sentence list.

of 95%, contains the 'true' list-SRT_N. The confidence intervals show that the SRT_N of list 10 is significantly different from the overall SRT_N, while the SRT_N of list 11 is at the limit of a significant difference. In an attempt to avoid a situation with some test lists being singled out in clinical use as easier or more difficult than the others, lists 10 and 11 were omitted from the test. The final 18 lists can be found in Appendix B. The appendix also contains seven practice lists compiled from sentences that have been excluded during various stages of the development process. The sentences are equalized with respect to intelligibility, but the lists should only be used for practicing the test procedure.

The following results are based on the final 18 lists. The mean SRT_N across all lists and listeners is -3.15 dB with an overall standard deviation of 1.0 dB. A two-way ANOVA shows a significant variation between listeners [$F(13,221) = 3.04$, $p = 0.0004$], but no significant variation between lists [$F(17,221) = 0.64$, $p = 0.86$]. The mean SRT_N for each list relative to the overall SRT_N is shown in Figure 4. Here, lists 1–9 are the

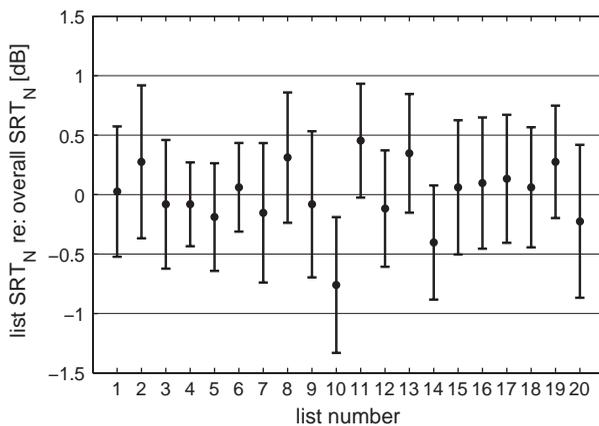


Figure 3. Mean list-SRTs deviations with 95% confidence intervals. When the 0-line is within the interval, the list-SRT does not deviate significantly from the overall SRT.

same as in the original set (Figure 3), while lists 10–18 correspond to lists 12–20 in the original set. All list-SRT_Ns lie within ± 0.5 dB of exact equality.

Test reliability

The reliability of a single SRT_N determination with one test list can be estimated from the repeated SRT_N measurements for the 14 listeners in the verification test. For each listener, an SRT_N was calculated as the mean across the 18 lists. This SRT_N value was then subtracted, listener by listener, from the individual SRT_N determinations with the 18 lists. These differences can be regarded as the deviation of each single SRT_N assessment from the 'true' SRT_N for the listener. Figure 5 shows these deviations ($n = 18 \times 14 = 252$) collected in bins of 0.5 dB. Seventy-one percent of the deviations are within ± 1 dB of the 'true' SRT_N. Ninety-three percent are within ± 1.5 dB and 99% are within ± 2 dB. The within-subject standard deviation in the verification test was 0.92 dB; the dots in the figure indicate a normal distribution curve with this standard deviation. The empirical distribution shows a tendency for more deviations in the interval -1.5 dB to 1.5 dB than predicted by the theoretical distribution, but fewer measurements with deviations above ± 1.5 dB.

Phone distribution

The 20 original lists in the present study were created with respect to phonetic balance. After exclusion of two lists, the overall phonetic distribution was recalculated for the final 180 sentences and the deviation from this optimal distribution was determined for each of the final 18 lists. The sentences consist of 3363 phones in total; 1291 vowels (38%) and 2072 consonants (62%). The distribution of phones is listed in Table 1. Fifty target values were defined for each list: 28 vowel counts, 20 consonant counts, one 'stød' count, and one syllabic consonant count. The distribution of the deviations between the target values and the actual values of the 900 counts ($50 \cdot 18$) is shown in Figure 6. Eighty-one percent of the deviations are within ± 1 , which can

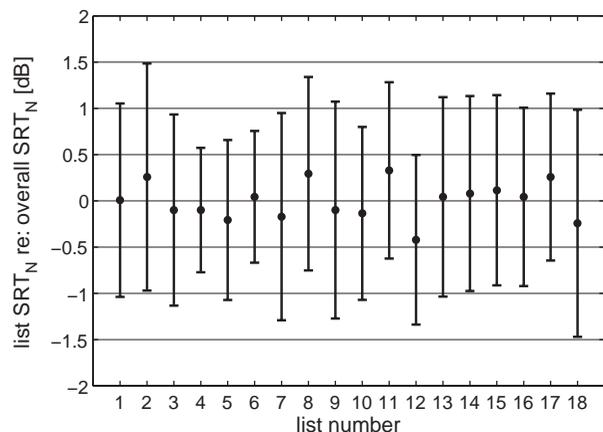


Figure 4. The list-SRTs relative to the overall mean (-3.15 dB). The bars indicate ± 1 standard deviation. The overall standard deviation is 1.0 dB; the standard deviation of the list-SRTs is 0.2 dB. The omission of list 10 and 11 from the original 20 lists means that list 10–18 in the present figure corresponds to list 12–20 in Figure 3.

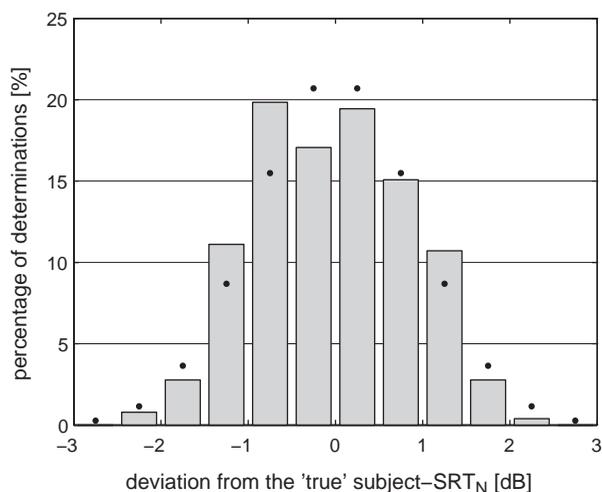


Figure 5. Percentage of SRT determinations with a given deviation from the individual SRT means. Bars show the percentage of measurements with the indicated deviation. Dots show the expected distribution under the assumption that results are normally distributed (std = 0.92 dB). Bin size is 0.5 dB.

be compared to 58% in Nilsson et al (1994), 75% in Vaillancourt et al (2005), and 70% in Hällgren et al (2006).

Psychometric function

The psychometric function of the test, shown in Figure 7, was determined based on the responses across all listeners and lists during the verification test of the original 20 lists. During the test, sentences were presented at levels from -8 to 2 dB SNR in steps of 2 dB. At each level, the percentage of words/sentences that were correctly repeated was calculated. The first sentence in each list was not included in the statistics, as this sentence was presented several times to each listener. The steepest slope of the fitted cumulative normal distribution curve is $18.7\%/dB$ for sentences and $14.9\%/dB$ for the word-based curve.

The slope of the psychometric function is to some extent influenced by the differences in the SRT_N between listeners. To investigate the influence of these differences, a sentence-based psychometric function was fitted to each individual listener. The steepest slope of these functions varied from $15.5\%/dB$ to $26.5\%/dB$. The average slope, $19.8\%/dB$, was only slightly steeper than the slope for the overall psychometric function.

Discussion

Comparison with other sentence tests

Talker, pronunciation, word frequency, language, and sentence redundancy are some of the factors that influence the SRT_N of a speech intelligibility test. Nevertheless, the overall SRT_N of several HINTs falls within a narrow range: -2.9 dB for the American-English HINT (Nilsson et al, 1994), -3.3 dB for the Canadian-French HINT (Vaillancourt et al, 2005), -3.0 dB for the Swedish HINT (Hällgren et al, 2006), and -3.9 dB for the Cantonese HINT (Wong & Soli, 2005). The present test has an

Table 1. Average distribution of phones in the final sentence lists (180 sentences). The notation is in accordance with the international phonetic alphabet (IPA). The goal of the optimization process was to reach a similar distribution for each list of 10 sentences. In addition to the listed phones, the number of ‘stød’ and number of syllabic consonants were also included as counts in the optimization.

Vowel distribution			
i	3.6%	œ	0.1%
y	0.5%	α	3.5%
u	1.5%	ɒ	0.2%
e	3.8%	i:	0.9%
ø	0.4%	y:	0.3%
o	0.7%	u:	0.4%
ə	3.0%	e:	0.4%
ɛ	2.3%	ø:	0.2%
œ	0.3%	o:	0.8%
^	2.3%	ɛ:	0.1%
ɔ	1.1%	ɔ:	0.9%
æ	0.9%	æ:	1.8%
ɐ	3.3%	ɑ:	0.7%
a	4.0%	ɔ:	0.5%

Consonant distribution

p	1.2%	ʃ	0.1%
t	2.5%	R	2.1%
k	1.5%	κ	1.4%
b	3.0%	h	2.9%
d	5.7%	l	5.6%
g	4.4%	j	1.3%
f	1.8%	w	0.7%
v	2.6%	m	4.0%
ð	4.2%	n	8.4%
s	6.9%	ŋ	1.3%

overall SRT_N of -3.15 dB. The standard deviation of the SRT_N across all lists and listeners is also similar for the different tests. The standard deviation amounts to 1.0 dB for the present test, 1.1 dB for the Canadian-French (Vaillancourt et al, 2005), 1.1 dB for the Swedish (Hällgren et al, 2006), and 1.7 dB for the Cantonese version (Wong & Soli, 2005). These standard deviations are dominated by the standard deviations within listeners, but are also influenced by the variation between the list- SRT_N s (means across listeners). This variation was depicted in Figure 4, and the figure can be compared to the corresponding figures in previous studies, e.g. Nilsson et al (1994); Vaillancourt et al (2005); Hällgren et al (2006). The list- SRT_N s in these studies are distributed within the interval ± 1 dB, while in the present study, the SRT_N s are confined to the interval ± 0.5 dB. This low variability is partly caused by the exclusion of the two most diverging lists and the inclusion of eight sentence levels in the SRT_N calculation. It does, however, also imply that the equalization procedure of the present study has led to more homogeneous sentence intelligibilities than the HINT

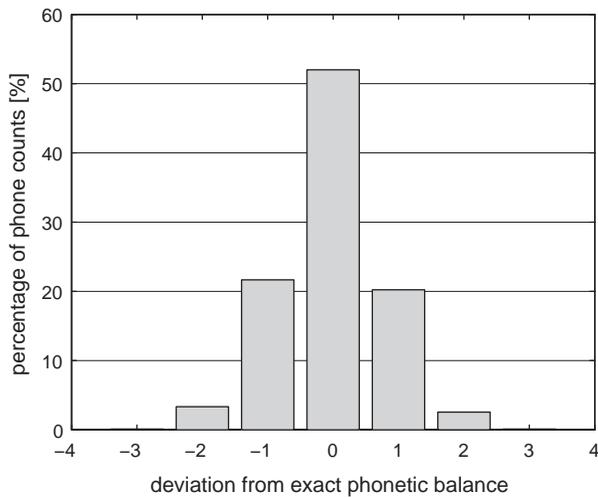


Figure 6. The distribution of the deviation between the target value and the actual value of the phonetic counts. As the target values are (normally) non-integers, the deviations are pooled in bins of size ± 0.5 around the indicated deviation.

equalization procedure. This is presumably due to the fact that the HINT procedure is based on an equalization of the average word intelligibility, which does not necessarily ensure equalized sentence intelligibilities.

The conflict that arises from using word scoring in the equalization procedure but sentence scoring in the adaptive test procedure cannot be resolved by switching to word scoring in the latter. A reliable and sensitive test with word scoring would require an equalization of the individual word intelligibilities—not only the average word intelligibility for each sentence—and this is in general not achievable for natural sentences.

Prediction of SRT_N improvements

The empirical results of the verification showed that 99% of the SRT_N determinations are within ± 2 dB of the listener's 'true' SRT_N . In practice, in a laboratory or a clinic, the SRT_N will often be measured twice for a listener, once for each of two conditions. The purpose can be to compare two different hearing aid settings and to decide whether there is a significant difference. The null hypothesis of such a setup is that the SRT_N is equal in the two conditions, and the rejection of the hypothesis depends on the difference between the two SRT_N determinations. If the within-subject standard deviation of 0.92 dB is also assumed to be valid for hearing-impaired listeners, the deviation of the difference is $\sqrt{2} \times 0.92$ dB = 1.30 dB. In the case of a two-tailed test (there is no expectation about which condition will result in the lowest SRT_N), the 5% critical region will be limited by $\pm 1.96 \times 1.30$ dB = ± 2.5 dB. In the case of a one-tailed test (there is an expectation about which condition will have the lowest SRT_N), the 5% critical region will be limited by 1.64×1.30 dB = 2.1 dB. In practical use, this means that an SRT_N difference of 2.5 dB indicates a significant difference between two conditions for both one-tailed and two-tailed tests. The test must be performed with the same listener in the two conditions.

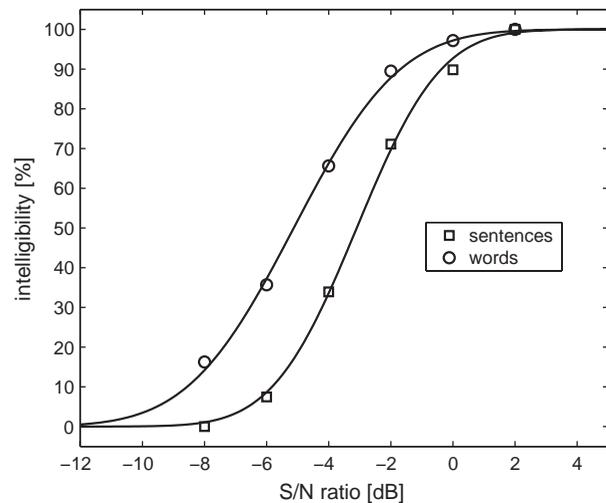


Figure 7. The psychometric function of the test based on correctly repeated sentences (squares) and correctly repeated words (circles). The solid curves are best fit cumulative normal distribution functions.

The HINT versus the CLUE equalization procedure

In the HINT equalization procedure, a linear relationship between SNR and intelligibility is assumed. A common assumption is that a 10% deviation in intelligibility is compensated by a 1-dB change of the SNR (an overview can be found in Soli & Wong, 2008). This method of adjusting the intelligibility is rather coarse and does not take the changing slope of the psychometric function into account. At the 50% point of the function, the intelligibility will be more sensitive to changes in the SNR than when the intelligibility approaches 0% or 100%. The HINT method also makes the implicit assumption that a change in the SNR leads to the same change in intelligibility for all sentences, although this is probably incorrect. In contrast, the CLUE equalization procedure presented in the present study leads directly to an adjustment of the sentence RMS level, without any assumed relationship between intelligibility and SNR.

The sentence equalization process is a time consuming part of the development of a speech intelligibility test. The time efficiency of the equalization procedure can be estimated by how many times each sentence has been presented to a person during the process. During the HINT procedure, each sentence presentation involves a listener and an instructor, who need to go through the sentence two times: the initial presentation and the repetition by the listener. This equals to four 'person-presentations'. In Vaillancourt et al (2005), each sentence was tested 36 times, resulting in 144 person-presentations per sentence. In Hällgren et al (2006), the sentences were tested 32 times, resulting in 128 person-presentations per sentence. During the equalization of the present study, the sentences were, on average, presented 3.1 times to 18 listeners without an instructor present, resulting in 56 person-presentations per sentence. This means that the length of time of the equalization process in the present study was more than halved compared to the procedure of previous HINT studies.

Limitations of the CLUE equalization procedure

One potential problem of the CLUE equalization procedure is that listeners will sometimes press 'ok' for sentences that they have misunderstood or not heard fully. This is possible because there is no instructor to evaluate what the listener has perceived. When this happens repeatedly for the same sentences, it is an indication of an intelligibility flaw. During a pilot test before the final verification, some of the sentences were never correctly repeated. This was regarded as unacceptable and a screening was carried out in order to identify such sentences. These sentences were omitted from the final test lists. The sentences should have been identified earlier in the process and definitely before the equalization. A suggested procedure would be to include a sentence screening at an early stage. All sentences that do not have an intelligibility of 100% at a relatively high SNR (e.g. 0 dB) should then be discarded.

During the equalization process, the sentences were presented in sequences that consisted of fewer and fewer sentences until all had been judged 'ok'. The last sequences inevitably consisted of only a few sentences that became quite well-known to the listeners, who sometimes got locked in an 'easy' judgement of these. Listeners were instructed to avoid pressing 'easy' repeatedly for these sentences, but in practice this was unavoidable to some extent and the level was often lowered substantially. As this 'false' lowering of the level would change randomly among the sentences from listener to listener, the effect has had only a minor influence on the average level adjustment. The problem could have been avoided by imposing a limit of four presentations per sentence. At the fourth presentation, the listener has either adjusted the sentence to a level outside of the acceptable ± 3 dB band or reverted to a level that previously was judged 'easy' or 'difficult'.

Characteristics of the speech material

The speech material in the present test has some characteristics that deviate from comparable sentence tests. The talker has a somewhat less clear pronunciation than a professional talker. This was to some extent the intention of choosing a non-professional talker: The test can thereby reveal the problems that some listeners have in a conversation, because they cannot interpret speech that is not clearly pronounced. The robustness of the test is, however, also affected, as the effect of the pronunciation will tend to occur in an uncontrolled manner at unspecified locations in the stimuli. The reasoning behind the decision to use a non-professional talker may therefore be doubtful. Also, the talker does not keep a constant quality of voice in all sentences. This is an unintended effect caused by a split of the recordings in several takes. The influence of this effect is, however, expected to be minor.

The sentence material contains sentences that can be perceived as more complicated and 'unnatural' compared to other sentence tests. This is based on the fact that the sentences originated from a written source and that they did not undergo an evaluation for naturalness by a group of native speakers. In spite of this, the verification of the test with normal-hearing listeners showed a within-subject deviation between SRT_N assessments that is equal to or lower than the deviation in comparable tests.

Conclusion

A speech intelligibility test with 18 sentence lists has been produced. The test is in many aspects comparable to HINTs developed for other languages and it has a similar overall SRT_N (-3.15 dB) and standard deviation (1.0 dB). The deviations of the list- SRT_N s from the overall mean are less than 0.5 dB, and thus considerably lower than in the HINTs. The CLUE equalization procedure might be the reason for this. In future developments of sentence-based speech tests, it is therefore suggested to consider this method instead of using the traditional HINT equalization procedure.

Acknowledgements

We wish to thank the following colleagues for assistance and advice during the project: Preben Dømler, Nina Grønnum, Birgit Hutter, and Niels Reinholt Petersen (INSS, University of Copenhagen); Claus Lyng Christensen (ODEON); Carl Ludvigsen and Erik Schmidt (Widex); Thomas U. Christiansen, Torben Poulsen, Jørgen Rasmussen, and Eric R. Thompson (Acoustic Technology, DTU). We would also like to thank two anonymous reviewers for their helpful and constructive comments and suggestions. Finally, we would like to thank Oticon's research unit, Eriksholm, for feedback on the final test. The present work was supported by the Oticon Foundation and the H.C. Ørsted Foundation.

References

- Boothroyd, A. & Nittrouer, S. 1988. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am*, 84, 101–114.
- Fairbanks, G. 1958. Test of phonemic differentiation: The rhyme test. *J Acoust Soc Am*, 30, 596–600.
- Glasberg, B.R. & Moore, B.C.J. 1989. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scand Audiol*, Suppl. 32, 1–25.
- Hagerman, B. 1982. Sentences for testing speech intelligibility in noise. *Scand Audiol*, 11, 79–87.
- 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, 227–237.
- House, A.S., Williams, C.E., Hecker, M.H.L. & Kryter, K.D. 1965. Articulation-testing methods: Consonantal differentiation with a closed-response set. *J Acoust Soc Am*, 37, 158–166.
- Kalikow, D.N., Stevens, K.N. & Elliot, L.L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am*, 61, 1337–1351.
- 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.
- MacLeod, A. & Summerfield, Q. 1990. A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. *Br J Audiol*, 24, 29–43.
- Middelweerd, M.J., Festen, J.M. & Plomp, R. 1990. Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology*, 29, 1–7.
- 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. 1978. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am*, 63, 533–549.
- Plomp, R. & Mimpen, A.M. 1979. Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18, 43–52.
- Prosser, S., Turrini, M. & Arslan, E. 1991. Effects of different noises on speech discrimination by the elderly. *Acta Oto-Laryngol (Stockholm)*, Suppl. 476, 136–142.
- Soli, S.D. & Wong, L.L.N. 2008. Assessment of speech intelligibility in noise with the hearing in noise test. *Int J Audiol*, 47, 356–361.
- 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.
- 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.
- Wagener, K., Josvassen, J.L. & Ardenkjaer, R. 2003. Design, optimization, and evaluation of a Danish sentence test in noise. *Int J Audiol*, 42, 10–17.
- Wong, L.L.N. & Soli, S.D. 2005. Development of the Cantonese hearing in noise test (CHINT). *Ear Hear*, 26, 276–289.

Appendix A

Sentence redundancy must be taken into account when comparing the sentence intelligibility (SI) and the average word intelligibility (WI) for a sentence. Boothroyd & Nittrouer (1988) introduced the *j*-factor, which is equal to the number of independently recognized parts in a sentence. Each part will consist of one or more words. In a normal, conversational five-word sentence, *j* will be in the order of two to three (Boothroyd & Nittrouer, 1988). The intelligibility of a sentence can be calculated as the product of the part intelligibilities. The examples below show that although sentences are equalized to have the same WI, they cannot be assumed to have the same SI. The examples are based on a target WI of 70% because this corresponds to an SI of approximately 50% under the assumption of two independent parts ($j=2$) with equal intelligibility ($0.7 \times 0.7 = 0.49$).

Example 1

Independent parts in a sentence do not always contain the same number of words. In a five-word sentence with $j=2$, one part may contain four words, the other the remaining one word. And despite the assumptions that lead to a target WI of 0.7, the intelligibility of these two parts may also differ. The intelligibility may be 0.8 and 0.3, respectively. For such a sentence the WI is equal to the target value 70% ($(4 \times 0.8 + 0.3)/5 = 0.7$), but the SI is only 24% ($0.8 \times 0.3 = 0.24$).

Example 2

The *j*-factor is likely to vary between different sentences, and may even vary for the same sentence, depending on the SNR (Boothroyd & Nittrouer, 1988). In a sentence with $j=1.5$ and a WI of 70% evenly distributed between all words, the SI will be 59% ($0.7^{1.5} = 0.59$). In another sentence with the same WI, but with $j=3$, the SI is 34% ($0.7^3 = 0.34$).

Appendix B

List 1

1. Vinduet vendte ud mod gaden
2. Han/Hun hoppede op på cyklen
3. Den gamle mand smilede stort
4. I regnbuen ses alle farver
5. De/Vi vil hellere male selv
6. Kampen gik godt i begyndelsen
7. Han/Hun har passet sin træning
8. Hver aften spiser de/vi salat
9. Det ringer ud til frikvarter
10. Hans bukser var meget korte

List 2

1. Stuen skal nok blive hyggelig
2. Døren er næsten aldrig åben
3. En ung pige kommer gående
4. De engelske bøffer var møre
5. Han/Hun kunne køre meget stærkt
6. Sofaen står bagerst i rummet
7. Torsdag var han/hun ikke hjemme
8. Begge fodboldhold klarer sig fint
9. Maden blev serveret til tiden
10. Han/Hun havde let ved hovedregning

List 3

1. Skuret er bygget af brædder
2. Hans mor var heldigvis hjemme
3. Under bogen ligger en tegning
4. Han/Hun rensede skærmen for støv
5. De/Vi skal bo på efterskolen
6. Hendes penge var gået tabt
7. Katten kom listende helt stille
8. Blomster og gaver strømmede ind
9. Hun/Han var i strålende humør

10. Vi/De er en fredelig familie

List 4

1. Pigen var køn og velbegavet
2. Vi/De sad ude i køkkenet
3. Flasken var fyldt med æblesaft
4. Katten spinder i hendes arme
5. De/Vi danser på et diskotek
6. Bageren havde tre slags rugbrød
7. Han/Hun kommer mandag med pakken
8. Trøjen er syet af bomuld
9. Hun/Han var en lille solstråle
10. De/Vi kom kørende i hestevogn

List 5

1. Godt håndværk holder i årevis
2. Min/Din kuglepen skriver med rødt
3. Mødet sluttede efter tre timer
4. Han/Hun ønskede sig en jakke
5. Jeg er ikke længere sulten
6. Han/Hun kan lugte hendes parfume
7. Villaen er ikke blevet solgt
8. Hjælpen nåede frem for sent
9. Vi/De spadserede en tur sammen
10. Han/Hun lagde tasken på bordet

List 6

1. Kurven var fyldt med vasketøj
2. Store bølger slog mod stranden
3. Han/Hun lagde brænde på bålet
4. Folk sidder og taler sammen
5. Hun/Han var bedst til matematik
6. Stemningen i klassen er god
7. Hendes mand havde et værksted
8. De/Vi unge gik i biografen
9. Han/Hun trækker gardinet til side
10. Vi/De ligner hinanden ret meget

List 7

1. Børnene sidder i en rundkreds
2. Gæsterne nyder den gode vin
3. Manden ville løbe en tur
4. De/Vi talte lidt om fremtiden
5. Pladsen var spærret af affald
6. Festen varede til over midnat
7. Bakken er halvtreds meter høj
8. Hun/Han havde ingen frakke på
9. De ønsker sig et sommerhus
10. Begge hold scorede otte mål

List 8

1. Forbruget af papir er stort
2. Mandag vågnede vi/de meget sent
3. Hendes far var ikke hjemme
4. I går kom svalerne hertil
5. Jeg havde cyklet i solskin
6. Skoledrengen drikker et glas mælk
7. Butikken holder et stort udsalg
8. Hun/Han lavede en kop kaffe
9. Nu venter landmændene på regn
10. De/Vi kommer sejlene til byen

List 9

1. Vinderen fik en flot pokal
2. Hunden svømmede væk fra kysten
3. De/Vi sidder længe i tavshed
4. Han/Hun læser med stærke briller
5. Pludselig kom der en lastbil
6. Der var altid åbent tirsdag
7. Mine/Dine venner går i gymnasiet
8. Bogen er skrevet på engelsk
9. Der bor mange mennesker her
10. Hun/Han var taget på arbejde

List 10

1. Toget er meget sjældent fuldt
2. Jeg var også utrolig glad
3. Hans datter vil på højskole
4. I går havde filmen premiere
5. Børnene og de voksne sover
6. En taxa kørte langsomt forbi
7. Bilen er ikke længere ny
8. Kaninen sprang ud gennem hullet
9. Næste deltager var smedens søn
10. Jeg sætter mig nede bagved

List 11

1. Reden er bygget af smågrene
2. Nu mangler vi/de blot tallerkner
3. Han/Hun var verdensmester i svømning
4. De/Vi cykler eller tager bilen
5. Huset lå omme bag torvet
6. Jeg spurgte ikke til prisen
7. De/Vi ankom sidst på formiddagen
8. Hun/Han rider på venindens hest

9. Insekter kan flyve meget langt
10. De/Vi har altid boet hjemme

List 12

1. Mødet skal holdes på skolen
2. Udenfor er det fuldstændig mørkt
3. Hun/Han var omgivet af mennesker
4. Børnene kom hjem ved middagstid
5. Bogen var billig på udsalg
6. Cykler kan lejes mange steder
7. Af og til larmer naboerne
8. De/Vi blev hurtigt gode venner
9. Han/Hun afviste det nye forslag
10. Koden til låsen passer ikke

List 13

1. Blomsterne vokser i små skåle
2. Høsten var allerede i hus
3. Vi/De havde en festlig aften
4. Man/Han skal holde korte pauser
5. De/Vi to venner deler arbejdet
6. Hendes kontor ligger langt væk
7. Din/Min bror er meget utålmodig
8. Bogen er fuld af eksempler
9. Manden skal ringe til hende
10. Jeg går ud på dansegulvet

List 14

1. Lakken skal fjernes fra gulvet
2. Han/Hun købte ikke mange blomster
3. Værelset lå ud til baggården
4. Naboerne var med til middagen
5. Lyskrydset skifter snart til rødt
6. Han/Hun er en flittig musiker
7. Vi/De havde en dejlig weekend
8. Lågen bag dem smækkede i
9. Hendes øjne så trætte ud
10. Vi/De får boller og chokolade

List 15

1. Nu skal maskinerne skiftes ud
2. Snart fylder rapporten ti sider
3. Jeg tager fat i dørhåndtaget
4. Tøjet var gået af mode
5. Her går alle med solbriller
6. Kassedamen så venligt på ham
7. Han/Hun ligger stadig i sengen
8. Eleven skriver en lang rapport
9. Hele byen kom til brylluppet
10. Vi/De så lidt af vejrudsigten

List 16

1. Skuffen kunne ikke lukkes helt
2. Vi/De byggede husene af træ
3. I morgen bliver vejret bedre
4. De/Vi sejlede med en husbåd
5. Han/Hun har aldrig lavet middagsmad

6. Udsigten til skoven var god
7. Motorløb kan være ret farligt
8. Vi/De rister pølser over bålet
9. Manden kom til en benzintank
10. Han/Hun kender alle byens gader

List 17

1. Pigen strikker en rød trøje
2. Vi/De ventede længe i køen
3. Om aftenen var der lejrball
4. Det kilder lidt i fingeren
5. Hun/Han gik hen til telefonen
6. Vi/De skal bare blive siddende
7. Suppen smagte godt af tomat
8. Huset her er hans barndomshjem
9. Redskaber skal sættes på plads
10. Vejrudsigten lover regn og slud

List 18

1. Om morgenen lagde stormen sig
2. Lyden kommer oppe fra loftet
3. Hun/Han har købt en vinterfrakke
4. I spisestuen var lyset tændt
5. Han/Hun talte til en kollega
6. Bagefter skal vi/de have jordbær
7. Musik giver en god stemning
8. Spillerne troede på sig selv
9. Tapetet var faldet af væggen
10. Hun/Han havde de smukkeste øjne

Practice list 1

1. Pigerne går rundt i haven
2. Alle skal betale samme pris
3. Hendes ansigt er stadig solbrændt
4. Filmen blev straks en succes
5. Jeg kan godt lide jazzmusik
6. Vi/De siger tillykke og skåler
7. Chaufføren ser ind i spejlet
8. Snakken ved bordet var livlig
9. Drys retten med hakket persille
10. De mørke pletter skyldes maling

Practice list 2

1. Alle foredrag er på engelsk
2. Drengen stikker hånden langt frem
3. Han/hun stiller mange svære spørgsmål
4. Kagen skal bages i ovnen
5. Båndet blev revet i stykker
6. Klokkeren var blevet over midnat
7. Han/hun blev en god skolelærer
8. De/vi fik jordbærkage til dessert
9. Jeg skulle ringe til formanden
10. Hatten passer til min/din tøjstil

Practice list 3

1. De to mænd kender hinanden
2. Båden sejler lidt over elleve

3. Fabrikens port var ikke lukket
4. Hans søster var blevet klippet
5. Jeg ønsker mig et kæledyr
6. Han/hun taler om sit arbejde
7. Natten bliver klar og kølig
8. Tårnet er ikke særlig højt
9. Jeg glemmer aldrig den musik
10. Hendes tøj var helt gennemblødt

Practice list 4

1. Strømperne var gået i stykker
2. Nu begynder en ny sæson
3. Rejsen varer mindst en uge
4. Lad os bare køre igen
5. Første stop er ved svømmehallen
6. Bussen kan ikke komme frem
7. Udsigten er bedst om sommeren
8. Han/hun er tilfreds med artiklen
9. Flyrejsen varer mindst fem timer
10. Jeg tager solbad på stranden

Practice list 5

1. Kunden er tilfreds med svaret
2. Gymnastik gør mig meget stærk
3. Grisene løber frit på marken
4. Holdet er klar til kampen
5. Du skal børste alle tænder
6. Hendes bror vil være brandmand
7. Nu blomstrer roserne på marken
8. Jeg var glad for bryllupsfesten
9. Drengen blev medlem af klubben
10. Renten var kun fire procent

Practice list 6

1. Det var en god fastelavnsfest
2. Kampen skal spilles på onsdag
3. Filmen er rigtig godt lavet
4. Derhjemme spiser vi/de ikke kød
5. Børnene løber rundt og leger
6. Hun/Han kommer meget i teatret
7. Familien går tur i parken
8. Statuen har ikke noget hoved
9. Hun/Han tog en hurtig beslutning
10. Vi/De snakkede med vores venner

Practice list 7

1. Billetterne bliver sendt til os
2. Ikke langt væk ligger rådhuset
3. Posen her er til grøntsager
4. Han/Hun sluttede som nummer fire
5. Chokoladen var dyr og god
6. Byen ser fantastisk dejlig ud
7. Flyttemænd har tit ømme muskler
8. Vi/De sagde farvel til gæsterne
9. Manden kløede sig på armen
10. Arbejdet er hårdt og krævende