Taylor & Francis
Taylor & Francis Group

Orginal Article

International Journal of Audiology 2006; 45:227–237

Mathias Hällgren
Birgitta Larsby
Stig Arlinger

Division of Technical Audiology, Department of Neuroscience and Locomotion, Linköping University, Linköping, Sweden

A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition

Una versión sueca de la Prueba de Audición en Ruido (HINT) para evaluar el reconocimiento del lenguaje

Abstract
A Swedish Hearing In Noise Test (HINT), consisting of everyday sentences to be used in an adaptive procedure to estimate the speech recognition thresholds in noise and quiet, has been developed. The material consists of 250 sentences, with a length of five to nine syllables, normalized for naturalness, difficulty and reliability. The sentences were recorded with a female speaker. From the sentences, 25 phonemically balanced lists were created. All lists fluctuate less than 1 dB of the overall mean. The standard deviation of the test-retest difference is 0.94 dB when testing with one list, and decreases to 0.68 dB and 0.56 dB for two and three lists, respectively. The average speech recognition thresholds in noise for the Swedish sentences were −3.0 dB signal/noise ratio (SD = 1.1 dB). The present study has resulted in a well-defined and internationally comparable set of sentences, which can be used in Swedish audiological rehabilitation and research to measure speech recognition in noise and quiet.

Sumario
Se desarrolló una HINT (Prueba de Audición en Ruido) sueca, constituida por frases cotidianas para usarse en un procedimiento que estime los umbrales de reconocimiento del lenguaje en ruido y en silencio. El material consistió en 250 frases, con una longitud de 5–9 sílabas, normado para naturalidad, dificultad y confiabilidad. Las frases fueron grabadas por un hablante femenino. De estas frases, se crearon 25 listas balanceadas fonéticamente. Todas las listas fluctuaron menos de 1 dB de la media global. La desviación estándar de la diferencia test/retest fue de 0.94 dB cada vez que se evaluaba una lista, bajando a 0.68 dB y a 0.56 dB para dos o tres listas, respectivamente. El umbral promedio de reconocimiento del lenguaje en ruido para las frases suecas fue de −3.0 dB de tasa señal/ruido (DS = 1.1 dB). El presente estudio ha producido un grupo de frases bien definido e internacionalmente comparable, que pueden ser utilizadas para la rehabilitación auditiva en sueco y en la investigación del reconocimiento del lenguaje en ruido y en silencio.

It is of interest to compare speech recognition in noise and quiet across languages. This is of particular importance for small countries, like Sweden for example, where the material available for some groups of patients may be too limited to allow far-reaching conclusions to be drawn. An international study of the average long-term spectrum for speech shows that there are only minor differences between Swedish and many other languages, including English (Byrne et al, 1994). Thus, it should be possible to develop a Swedish set of sentences that can be used for the measurement of speech recognition thresholds in quiet and noise. Measurements done with this speech material can be compared with results from numerous international studies where the original HINT (Hearing In Noise Test) sentences have been used (Nilsson et al, 1994). Sentence materials similar to the American English HINT sentences are also available in other languages such as Dutch (Plomp & Mimpen, 1979), in German (Kollmeier & Wesselkamp, 1997) and in Canadian French (Vaillancourt et al, 2005).

The HINT sentences have been used on many occasions such as recording of speech perception material in noise and quiet, and verifying the benefit from hearing-aid amplification and cochlear implants, especially in noise.

The original HINT material (Nilsson et al, 1994) consists of short everyday sentences in English, which are judged to be natural by native speakers of American English. The sentences are grouped into 25 phonemically balanced lists of ten sentences. The test sentences are designed to be used in an adaptive procedure to establish the speech recognition threshold for sentences, where 50% of the sentences are correctly repeated, a measure that does not produce ceiling or floor effects.

The purpose of this study was to develop and evaluate a new Swedish material of sentences for the measurement of speech recognition. The speech material should, as closely as possible, be comparable to the well-recognized HINT in English (Nilsson et al, 1994).

With a Swedish HINT material it will be possible:

- to compare results from Swedish studies of speech recognition thresholds in noise and quiet with international studies using similar sentence material in other languages;
- for Sweden to contribute to international multicenter studies, which will give a more reliable and valid base for rehabilitative actions, and therefore translate more quickly into benefits for individuals with hearing impairment;
- for Swedish studies to have a greater impact on the international community.

ISSN 1499-2027 print/ISSN 1708-8186 online
DOI: 10.1080/14992020500429583

Accepted:
September 20, 2005

Mathias Hällgren
Department of Neuroscience and Locomotion, Division of Technical Audiology, University Hospital, S-581 85 Linköping, Sweden.
E-mail: matha@inr.liu.se

In addition, a description of the procedure for the development of the Swedish HINT test, as found in this article, can hopefully be useful for researchers in other countries who want to develop their own sentence material.

The study has resulted in a well-defined and internationally comparable tool, which can be used in Swedish audiological rehabilitation and research to measure speech recognition in noise and quiet. The steps in the test development include development and equating of the sentence material and creation and verification of lists, similarly to the procedure used by Nilsson et al (1994).

Methods

Creation of test material

Development of the sentence material

The American English HINT sentences were derived from a set of short sentences that were developed to test children (Bench et al, 1979). The sentences were transcriptions of British children's speech, and incorporated common nouns and verbs. As no similar lists of sentences are available in Swedish, the first step in the development of the Swedish HINT was the creation of new speech material consisting of everyday sentences that sound natural despite differences in dialect, education and background. The sentences available in the HINT material in American English served as a starting point for our development. Some of the sentences were difficult to adjust to the Swedish language and had to be modified according to both content and structure.

The original HINT sentences have a maximum of seven syllables, whereas the Swedish sentences are composed of five to nine syllables. The greater number of syllables in the Swedish set reflects inherent differences in word length between the two languages. Swedish sentences often contain more phonemes and syllables than English sentences with comparable content.

Test of naturalness

The sentences that were derived from the criteria outlined above were evaluated and rated for naturalness using a scale from one to seven, where seven stands for a natural sentence and one represents an artificial sentence. The goal of this evaluation was that all sentences be perceived as natural by native speakers. The procedure followed that outlined in the original HINT study (Nilsson et al, 1994).

The sentences (n = 360) were divided into lists of 40 sentences. Fifteen subjects read and evaluated each list and were asked to provide suggestions that would make artificial sentences sound more natural.

The sentences with a mean score below six were revised using the subjects' suggestions for improvement. The revised sentences (n = 79) were then rated again by fifteen new subjects for approval. After the second round, only four sentences with a rating score below six remained and were excluded from further use.

Processing of the material

Before recording of the sentences another 43 sentences were excluded from the material because of similarities in content with other sentences. Recordings were made of the remaining sentences (n = 313) using a female speech and language therapist with no distinctive dialectical features. She read the sentences at a normal pace and without stressing any words. The sentences were recorded on digital audio tape (DAT) and were digitally transferred to computer soundtrack files. The sampled waveforms were edited into individual sentence files. Silent intervals before and after each waveform were eliminated. The RMS value was computed for each sentence waveform and the sentences were then rescaled to the same RMS level.

After equating the waveform levels, the average long-term spectrum of the sentences was computed and a random noise with the same spectral properties as the speech signal served as

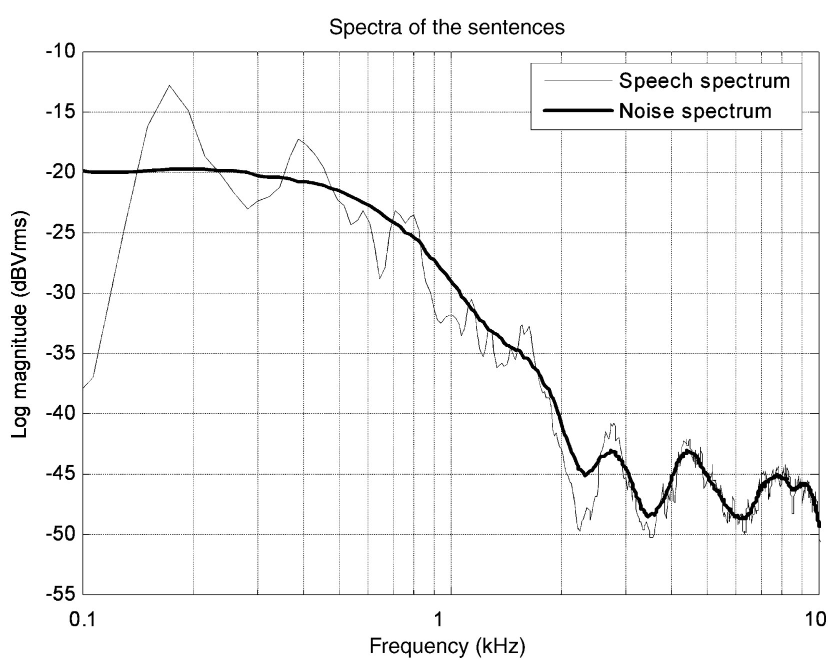

Figure 1. Spectra of the sentences and the masking noise.

target for the background noise. A 128-coefficients finite impulse response (FIR) filter was designed to match the long-term spectrum of the sentences. White noise was filtered through this FIR-filter and scaled to the same RMS amplitude as the sentences. Figure 1 shows the average long-term spectrum of the speech and the measured spectrum of the filtered noise. The general shape of the long-term spectrum agrees with the Swedish long-term spectrum obtained by Byrne et al (1994).

Equating of sentence difficulty

The accuracy and efficiency of an adaptive procedure for determination of the speech recognition threshold is dependent on sentences with approximately equal intelligibility. Phonemic content, word familiarity, intonation, etc. can affect the speech intelligibility in noise despite equal RMS levels (Nilsson et al, 1994). By adjusting the levels of the different sentences, the intelligibility can to some extent be equalized. A total of 313 sentences were tested for equal difficulty.

A total of 96 native Swedish-speaking subjects participated in this part of the evaluation. They all had, according to self-report, normal hearing. The mean age of the subjects was 24 years (range 19–40 years). The test subjects were offered a movie ticket for their participation.

The procedure for equating sentence intelligibility is outlined below. The sentences were presented in noise with the same average long-term spectrum as the sentences, in a sound field. The loudspeaker producing both speech and noise was located one meter in front of the test subject. The noise level was fixed at 65 dB SPL. Fifteen of the sentences, different for each subject, were initially used to individually adjust the S/N ratio in an adaptive procedure with decreasing step-size to reach 50% correct word recognition. Thereafter the S/N ratio was kept constant for each subject (mean S/N = −4.0, range: −7.3 to −1.4).

The adjustment of sound levels of the sentences took place in four rounds. Twenty-four test subjects participated in each round and each subject was presented with 1/3 of the total number of sentences. Each sentence was thus presented to eight subjects.

The difference in percent correctly identified words for each sentence and the average percentage for all sentences was used to adjust the level of each sentence. After the first round, the level of the sentence was adjusted by 1 dB per 10% difference from the grand average, after the second round by 0.7 dB per 10% difference, after the third round by 0.5 dB per 10% difference and finally, by 0.3 dB per 10% difference. As a result of the level adjustments, the standard deviation of the distribution of correct answers for the sentences decreased from 27.1 to 16.9%. The mean intelligibility score varied from 49.1% to 62.7% between rounds.

Sentences with words that were either never, or in less than 5% of the cases, correctly recognized were excluded from the sentence material. Also, sentences adjusted by more than 3 dB compared to the average adjustment for all sentences were excluded. It can be assumed that these sentences were pronounced differently or had a content that differed from the rest of the sentences. After excluding these sentences, a total number of 279 remained. The aim of this study was to create 25 lists with ten sentences each. The standard deviation of the distribution of correct answers decreased even further by the exclusion of another 29 sentences with the highest difference (more than 1.6 SD) from the grand mean. In this way the standard deviation decreased to 14.1% for the remaining 250 sentences (mean percent correct words = 62.1%).

Creation of lists

All 250 sentences were translated to phonemic notation and the phonemic distribution for the entire set of sentences was determined. Table 1 shows the phonemic distribution for the 250 sentences. Other Swedish sample materials (Bertenstam et al, 1995; Hedelin et al 1988; Fant, 1967) where the phonemic distribution has been established show the same general pattern with a few exceptions. The most obvious difference is the frequency of the phoneme ‘ɛ’, which appears more often in the HINT material than in the other materials. The sentences were composed of 5216 phonemes, 3197 (61%) consonants and 2019 vowels (39%).

Twenty-five lists with ten sentences each that matched the phonemic distribution of the entire set of sentences as close as possible, were created using a computer program in which a trial-and-error process minimized the total deviation from the overall distribution. For each phoneme (n = 41) in each list of sentences (n = 25), the difference between the target phonemic content, as predicted from the overall distribution, and the obtained phoneme count was tabulated, resulting in a total of 1025 (25 × 41) counts. A difference of only ±1 phoneme was found in 70% of the counts. Five practice lists were formed from the sentences that had been excluded during earlier steps in the creation of the sentence material. The sentence material is listed in the Appendix.

Interlist reliability

The next step in the development of the Swedish HINT was to establish the test procedure and to determine the repeatability, and thus the reliability, of speech recognition thresholds obtained with the different lists. To perform testing, an adaptive automated computer-controlled procedure was developed. The

Table 1. Phoneme distribution for the 250 sentences in the test material. Values in the table express the rate of occurrence of each phoneme as a percentage of the total phoneme count. The sentences were composed of 5216 phonemes, 3197 (61%) consonants, and 2019 vowels (39%).

Consonant distribution					
p	2.7%	f	2.1%	n	8.3%
b	2.0%	v	2.4%	ɳ	1.0%
t	6.1%	s	5.9%	ŋ	1.0%
ʈ	0.5%	ʂ	0.3%	l	5.1%
d	3.8%	ʃ	0.6%	j	1.2%
ɖ	0.2%	ɕ	0.5%	r	6.6%
k	4.8%	h	1.4%		
g	2.1%	m	2.8%		
Vowel distribution					
i:	0.7%	a	8.9%	ɯ:	0.7%
ɪ	3.4%	ɑ:	1.9%	y:	0.2%
e:	1.0%	ɔ	2.1%	ʏ	0.5%
ɛ	10.6%	o:	2.6%	ø	0.9%
ə	0.2%	ʊ	1.5%	ø:	0.9%
æ	0.1%	u:	0.8%		
æ:	0.7%	ɵ	1.2%		

sentences and the noise were recorded on computer sound track files and mixed before they were played through the output of a sixteen-bits soundcard to a loudspeaker at a distance of one meter in front of the test subject. The sentence and the noise signal were calibrated separately. Once levels were calibrated, the computer controlled the speech and noise signals, and adjusted the levels according to the adaptive procedure. The computer controlled what materials were presented, as well as the levels of the signal and the listeners' responses as indicated by the test leader.

Experimental design

Ten young subjects (four women and six men) participated in this part of the study. Their age ranged from 18 to 30 years, with a mean of 21.3 years. All were native speakers of Swedish. The subjects had hearing thresholds of 20 dB HL or less for the frequencies 0.25, 0.5, 1, 2, 3, 4 and 8 kHz. At 0.125 kHz, all subjects had hearing thresholds of 20 dB HL or less except one, who had a threshold of 25 dB HL (right ear). At 6 kHz, three subjects had a threshold of 25 dB HL in the right ear and two subjects had a threshold of 25 dB HL in the left ear, while the rest of the ears had hearing thresholds of 20 dB HL or less.

Speech recognition thresholds in noise for all lists were obtained for each subject. The list order in which each subject was tested was counterbalanced according to a modified Latin-squares design. Within each list, the order of sentences was randomized. For each subject, threshold measurements were first performed for thirteen lists followed by a ten-minute break, after which thresholds for the remaining twelve lists were obtained. The duration of the measurement session was 60 minutes (including break).

The long-term test-retest variability was assessed by repeating the SRT measurements for each of the ten subjects, approximately one week after the first test. For each subject, the exact same list order as on the test occasion was retained for the retest.

The subjects were asked to listen to each sentence and repeat aloud what they heard. The subjects were encouraged to guess if they were uncertain. The noise level was fixed at 65 dB SPL (C-weighted). The speech level was measured as the equivalent C-weighted RMS-level for all sentences with the silent intervals between sentences excluded. An adaptive up-down procedure determined the sentence presentation levels; the first sentence in each list was presented below threshold (−8 dB S/N) and was increased by 2 dB steps until it was repeated correctly. The following sentences in the list were presented once each, with the presentation level dependent on the result of the preceding response. Sentence presentation levels were reduced by 2 dB after a correct response and increased by 2 dB after an incorrect response. Scoring was based on correctly reported whole sentences, since this is the most common method used during actual testing with the American English HINT sentences. This is in contrast to the method used to test sentence difficulty, where scoring was based on correct repetition of words. Small variations in what was accepted as a correct sentence were allowed. These variations occurred in verb tense (for example 'är' and 'var '), in articles (for example 'katt' and 'katten'), and in singular versus plural nouns.

Results

The pattern of presentation levels on individual trials was examined to determine the number of sentences to be included in the SRT calculation. In order to eliminate the effects of subject differences from these analyses, the presentation levels for each subject were expressed in dB re: the presentation level on the eleventh trial for each list. These reference calculations were performed both for the mean and standard deviation of presentation levels. The subject-normalized results from these calculations are shown in Figure 2. The mean presentation levels seem to stabilize after the fourth or fifth sentence, as does the standard deviation. It can thus be concluded that the

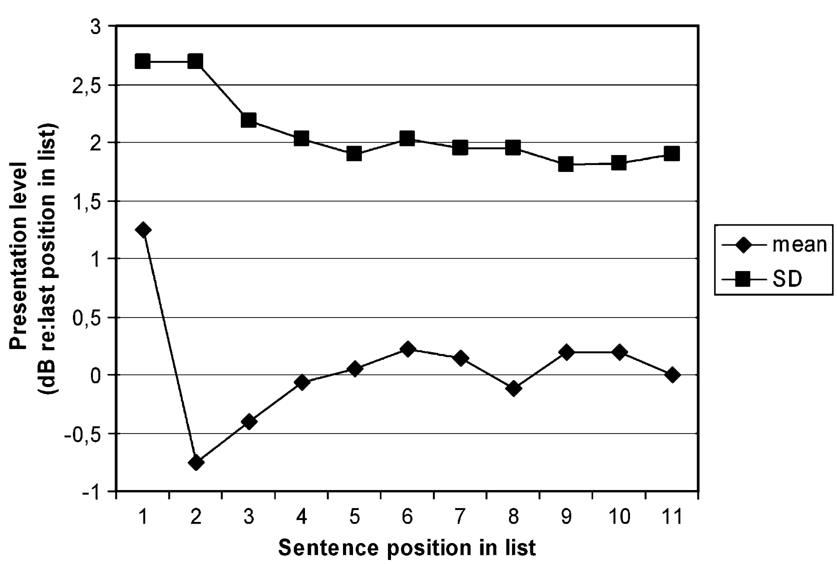

Figure 2. Means and standard deviations of sentence presentation levels for each list position in the sequence of adaptive threshold measurements in noise. All levels are presented as dB re. the predicted level of an eleventh sentence in the sequence based upon the response to the tenth sentence.

listeners are near their speech recognition threshold by the fifth trial, and the threshold was determined from the mean presentation levels of the fifth to the eleventh trials, where the level of the eleventh sentence was predicted from the response of the tenth sentence. The mean S/N ratio at threshold across all subjects and lists was −3.0 dB, with a standard deviation of 1.1 dB.

The ambition when starting the work was to have 25 lists that were as equal as possible, i.e. that yield the same SRT in noise. In order to verify the list equivalence, the mean speech recognition thresholds for each list across all subjects was computed and expressed as a deviation from the mean across all lists and all subjects. Figure 3 shows the mean deviation for each list and the error bars show the standard deviations of the mean. All list means differ by less than 1 dB from the overall mean. An analysis of variance found a significant effect of list at a significance level of 0.05, but not at a 0.01 level ($F(24,225) = 1.71$, $p = 0.024$).

The reliability and repeatability were tested in two ways. Table 2 shows the standard deviation for the test-retest difference in S/N ratio when the first list before the ten-minute break is compared with the first list after the break (row1). The table also shows the reliability when the mean value of two and three lists, respectively, before the break are compared with the mean of the same number of lists after the break. It can be concluded from the table that the standard deviation of the test-retest difference is always less than 1 dB.

For the longer test-retest interval (one week), a systematic change of 0.77 dB in mean S/N ratio was observed from test (S/N = −3.0) to retest (S/N = −3.8). The standard deviation for this test-retest difference was 1.4 dB. Figure 4 shows SRT means as a function of number in the presentation sequence at test and retest. An ANOVA on the SRT values from the first occasion showed no significant effect of presentation order ($F(24,225) = 0.56$, $p > 0.05$). The list performed first by each subject is the only one that shows a result that differs significantly from any other in the test sequence (post hoc analysis). From these data one can conclude that one training list seems sufficient for the test subject to be acquainted with the test procedure.

Table 2. The standard deviations for the short-term test-retest variability when comparing one, two, and three lists respectively. Test is followed by a ten-minute break before retest.

	SD
1 list	0.94 dB
2 lists	0.68 dB
3 lists	0.56 dB

Psychometric function

Data from the test of interlist reliability were used to calculate psychometric functions of the material, both for scoring on words and whole sentences. The first sentence in each list was excluded from the calculations since it was presented several times until it was correctly recognized. By using the data from the test of interlist reliability, different numbers of observations were obtained for different signal-to-noise ratios. However, the highest number of observations were in the most interesting range were the slope is steepest, whereas the tails had fewer observations.

The performance-intensity functions and the calculated psychometric functions are seen in Figure 5. The slope when scoring on whole sentences is 17.9%/dB at its steepest (calculated between −4 and −2 S/N). When scoring on words, the corresponding value is 15.4%/dB (calculated between −6 and −4 S/N).

Discussion and conclusion

For humans, speech is one of the most important acoustic signals. Speech in noise is an essential audiometric test procedure since it reflects the functional hearing and communication ability in individuals with hearing impairment when using a realistic speech signal above threshold and in the presence of interfering noise.

The outcome and the validity of speech-in-noise tests differ depending on the speech signal, the psychoacoustic procedure, and the characteristics of the masking noise used for the test. In Sweden, Hagerman (1982) created lists of sentences to be used to

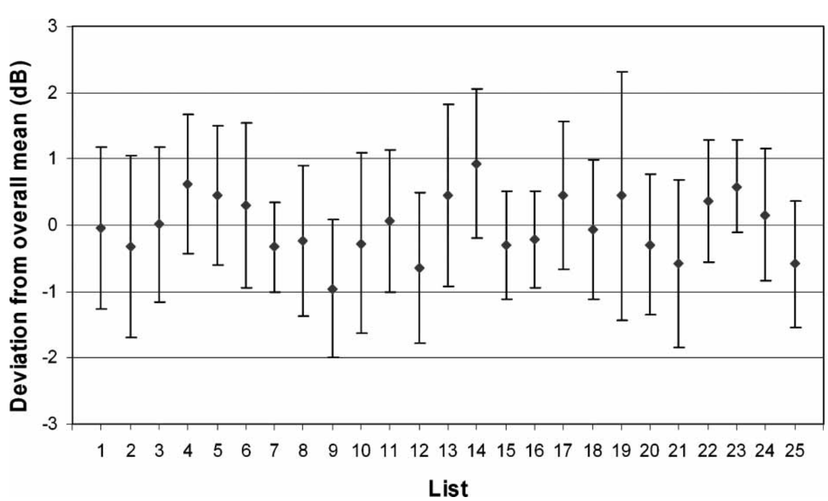

Figure 3. Differences between SRT means for individual lists and the mean SRTs across all lists in noise. For each list n = 10. Error bars indicate ± one standard deviation for each mean.

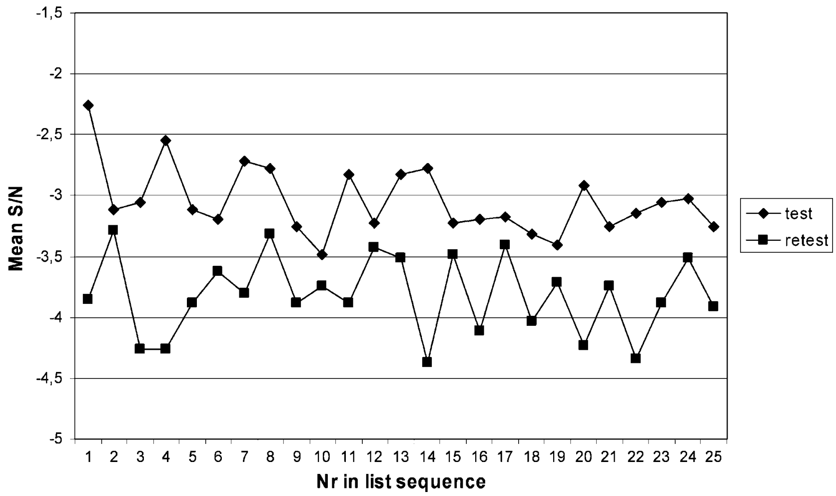

Figure 4. Mean S/N as a function of number in presentation sequence at test (◆) and retest (■) with a one-week interval between test and retest.

measure speech recognition in noise. The speech material consists of eleven lists with ten sentences each. All sentences contain five low-redundancy words with the same structure, and the same 50 words are used in all the lists but in different computer-generated combinations. The structure of the Hagerman sentence material yields very reliable results in the estimation of the SRT but can be experienced as unnatural. It has also been shown, in applications where measurements are repeated frequently (Hällgren 2002, Hagerman & Kinnefors, 1995), that significant learning effects are present. It is worth mentioning that the Hagerman and the HINT sentence materials differ in redundancy, which is likely to affect the influence from the listener's cognitive functions on the test results. It was concluded that a Swedish material with natural everyday sentences would be complementary to the Hagerman lists and provide a base for comparisons of data across languages.

The aim of the present study was to develop a sentence material that matched the original American English HINT sentences as closely as possible. This would allow results with the Swedish HINT material to be compared with those numerous studies where the English HINT material has been used. It was therefore considered essential to follow as closely as possible the same procedure in developing the lists as outlined by Nilsson et al (1994). The Swedish sentences were rated by test subjects to be natural. Sentences that never, not even after modifications, reached a mean rating score of at least six (on a seven point scale) were excluded from the final sentence material. The Swedish material was also equated for sentence difficulty in a

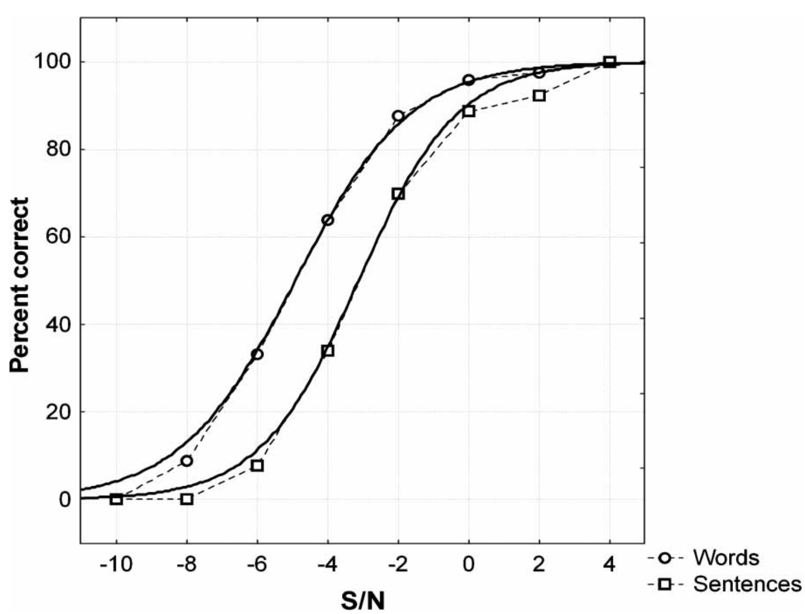

Figure 5. The performance intensity functions (dashed lines) and the calculated psychometric functions (solid lines) for scoring on words and whole sentences, respectively.

similar way as the American English HINT sentences by adjusting the levels of the different sentences towards equal intelligibility. As a result of this process, the standard deviation of the distribution of correct answers for the sentences decreased from 27.1 to 16.9%. In the American English HINT material, the standard deviation of the distribution of sentence intelligibility score decreased from 34.2 to 25.1% by the level adjustments.

In creating lists of sentences, it was important to have the different lists phonemically balanced. From the distribution of the differences between the target and obtained counts, it was calculated that a difference of only ± 1 phoneme was found in 70% of the counts, which should be compared with 68% in the English HINT material.

Scored on whole sentences the mean S/N ratio at threshold across all subjects was -3.0 dB, with a standard deviation of 1.1 dB. The results can be compared with those obtained in other materials scored on whole sentences: in American English, mean $= -2.9$ dB, SD $= 0.78$ dB (Nilsson et al, 1994); in Dutch, mean $= -5.9$ dB, SD $= 0.9$ dB (Plomp and Mimpen, 1979); in Canadian-French, mean $= -3.3$ dB, SD $= 0.5$ dB (Vaillancourt et al, 2005).

The slope of the psychometric function when scoring on whole sentences is for the present material 17.9%/dB at its steepest (Figure 5). The corresponding values for similar materials are 9.7%/dB in American English (Nilsson et al, 1995); 20%/dB (Plomp & Mimpen, 1979) and 11.7%/dB (Versfeld et al, 2000) in Dutch. In the present investigation, the psychometric function when scoring on words has a somewhat shallower slope (15.4%/dB), which was also found by Versfeld et al (2000).

One aim in the process was to have 25 lists with as equal speech reception threshold (SRT) in noise as possible. It was found that all list means differ by less than 1 dB from the overall mean, which is in close agreement with Nilsson et al (1994). The ANOVA of list equivalence showed a slight but significant effect of list ($p = 0.024$). However, a post hoc analysis (Bonferroni) revealed no difference between any pair of lists. Also Nilsson et al (1994) reported a significant effect of list ($p < 0.01$) when measuring SRT in noise.

All subjects were tested with all 25 lists. The test-retest variability was assessed by comparing the result for each subject's list number 1 with his/her list number 14, which was determined after a 10-minute break. Mean values of two and three lists before the break were also compared with the means of the same number of lists after the break. From these data, it can be concluded that the standard deviation for the test-retest difference, measured in this way, is always less than 1 dB (Table 2). Furthermore, the variability decreases from 0.94 dB with one list to 0.68 dB with the mean of two lists to 0.56 dB with the mean of three lists. The fact that the variability decreases with the number of test lists motivates the use of at least two test lists in order to achieve a high accuracy in SRT. Even if a different procedure to verify the variability was used in the American English HINT study, the variability with the present sentence material is of the same order of magnitude.

Repeated SRT measurements are performed for some clinical and research applications, and sometimes the same list(s) are used at different occasions with an interval of days or weeks between tests. The variability for those occasions is also of interest. For the longer test-retest interval (one week), a systematic change of 0.77 dB in mean S/N ratio was observed from test to retest. The standard deviation for the test-retest difference was now slightly higher, 1.4 dB. The fact that the subject is already acquainted with the test procedure at the second occasion made the subjects start from a lower level and reach a plateau sooner (Figure 4). The change in plateau level can at least partly be attributed to the fact that some of the words and sentences were recognized at retest. Subjects recalled having heard some sentences at the first session. The interval between test and retest will certainly affect the magnitude of the test-retest difference as well as the variability.

As outlined above, there are many similarities between the Swedish HINT and the original American English HINT (Nilsson et al, 1994), but there are also differences. The main difference between the two materials is that the American English HINT sentences were recorded with a male speaker while a female speaker was used in the Swedish HINT. This is a fact that has to be considered when comparing data across languages. Male and female speech differs in some aspects. The fundamental frequency of the female voice is approximately one octave higher than for the male voice, and thus the energy of the female voice is somewhat lower at low frequencies and somewhat higher at higher frequencies (Byrne et al, 1994). It is at present unknown how much this shift affects the results in SRT measurements. Hagerman's lists (Hagerman, 1982), which is the most commonly used sentence test material in Sweden, uses a female voice, while Magnusson's speech-in-noise test (Magnusson, 1995), which is also frequently used in Sweden, uses phonemically balanced wordlists recorded with a male speaker. The reason for using a female speaker in the Swedish HINT was to have a female speaker both in the existing Hagerman test and in HINT, since this comparison was of great interest for the research group. In the future, comparable lists of sentences with both a male and a female speaker are desirable, and development of a version with a male speaker is in progress. It can be noted that the Dutch sentences, which are similar to the HINT sentences (Plomp & Mimpen, 1979), use recordings with a female speaker.

The sentences in the American English HINT test are designed to be scored based on recognition of whole sentences. The Bamford-Kowal-Bench sentences, which were the basis for the development of the American English HINT sentences, were designed to be scored on recognition of key words, on the other hand. In the Swedish HINT material, as it is presented here, we have mainly used scoring on whole sentences. When scoring based on correctly repeated sentences is used, some information that can be obtained with scoring on keywords is lost, i.e. the recognition of each single word and phoneme. It is also likely that, especially in noise, some individuals with hearing impairment never hear the complete sentences even when the signal level is increased. Further development of the present sentence material will also allow the possibility to score on key words. A normative study is planned to investigate speech recognition in young adult listeners with normal hearing, comparing keyword and sentence scoring, and female and male speech.

In conclusion, a Swedish Hearing In Noise Test has been developed to be used to estimate the speech recognition threshold for natural sentences, especially in noise. The material consists of 250 sentences divided into 25 phonemically balanced lists and has been shown to be comparable, in most aspects, to the original HINT material in American English.

Acknowledgements

Thanks are due to The Swedish Council for Working Life and Social Research (FAS) and the EU-project HEARCOM for generous support. We would also like to thank the KVTill-grupp 2, 2003, who developed the first set of the Swedish HINT sentences; as well as Hanna Björk and Anna Hilertz who continued and extended the work towards the present sentence material. Finally, thanks are due to all test subjects who willingly helped us to evaluate the sentences at different stages in the process.

References

Bench, J., Kowal, A. & Bamford, J. 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol*, 13, 108–112.

Bertenstam, J., Blomberg, M., Carlson, R., Elenius, K., Granström, B., et al. 1995. Spoken dialogue data collected in the Waxholm project. *Speech, Music and Hearing Quarterly Progress and Status Report*, STL-QPSR 1. KTH, Stockholm, Sweden.

Byrne, D., Dillon, H., Khanh, T., Arlinger, S., Wilbraham, K., et al. 1994. An international comparison of long-term average speech spectra. *J Acoust Soc Am*, 96, 2108–2120.

Fant, G. 1967. Kompendium i talöverföring. Dept. of speech communication, KTH, Stockholm, Sweden. (In Swedish.)

Hagerman, B. 1982. Sentences for testing speech intelligibility in noise. *Scand Audiol*, 11, 79–87.

Hagerman, B. & Kinnefors, C. 1995. Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scand Audiol*, 24, 71–77.

Hedelin, P., Huber, D. & Leijon, A. 1988. Probability distributions of allophones, diphones and triphones in phonetic transcription of Swedish newspaper text. In: *Technical report no 8*, Dept. of Information Theory, School of Electrical and Computing Engineering, Chalmers University of Technology, Göteborg, Sweden.

Hällgren, M. 2002. Listening span and speech recognition. Internal report. Department of Technical Audiology, Linköping University.

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.

Magnusson, L. 1995. Reliable clinical determination of speech recognition scores using Swedish PB words in speech-weighted noise. *Scand Audiol*, 24, 217–223.

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.

Nilsson, M., Soli, S. D. & Sumida, A. 1995. A definition of normal binaural sentence recognition in quiet and noise. *Unpublished data*.

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

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, 358–69.

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

Appendix

List 1

1. Farfar ska vaxa bilen
2. Pojken stod på händer
3. Låset frös under natten
4. Den gamle mannen läste en bok
5. De röda stövlarna var för små
6. Hunden kom hem till slut
7. Bebisen sover i skuggan
8. Hon klättrar upp på taket
9. Killen satt tyst i hörnet
10. En lärare sjöng i aulan

List 2

1. Cykeln ligger slängd i gräset
2. Tanten handlar en gång i veckan
3. Frukten packades i sex lådor
4. Plånboken låg kvar på isen
5. Mamma lär sig hoppa fallskärm
6. Datorn blev överhettad
7. Rektorn tog fram kastrullen
8. Kvinnan arbetade på posten
9. Nästa år börjar de sjuan
10. Bonden skjutsade sina barn

List 3

1. Kvinnan ringde sin syster
2. Isen lade sig över sjön
3. Träden gav svalkande skugga
4. Personalen kommer imorgon
5. Korvgubben tittar på stjärnorna
6. Sovrumsmattan är blå
7. Killen halkade i badrummet
8. Mannen fotograferar slottet
9. Föräldrarna satt på balkongen
10. Tjejen tappade boken

List 4

1. I tallen satt en hackspett
2. Kvinnan missade barnens pjäs
3. Han försökte lyfta sju kilo
4. Mormor ska klippa sig på tisdag
5. Flickan brände sig på ryggen
6. Det gamla bordet står på gården
7. Killen visslade från balkongen
8. Turisten cyklade över bron
9. Pojken startade en förening
10. Rummet har tre gula väggar

List 5

1. Flickan ska åka skridskor idag
2. Mamma slog upp tältet på ängen
3. Eleven glömde sin väska
4. Lillebror förfrös vänstra stortån
5. Tåget stannade vid perrong sju
6. Hon fick schampo i ögonen

7. Pojken föll från trädet
8. Damen handlade på torget
9. Flickorna åkte karusell
10. Pappa bakar torra bullar

List 6

1. De svarta fåren går på ängen
2. Clownen tappar en ballong
3. Pappa byggde kojor med oss
4. Den stora svarta hästen frös
5. Flickan valde mellan tre par jeans
6. Chefen hittade en nyckel
7. Pennan ligger på skrivbordet
8. Sovrumsdörren var målad i rött
9. Hon satte sig vid elden
10. Dagisbarnen gick till lekparken

List 7

1. Den gröna hinken står på marken
2. Golvet täcktes av en vit matta
3. Soppan sjöd i kastrullen
4. Pappa ska laga min fåtölj
5. Bagaren knäppte sin stora rock
6. Båten hade två blå segel
7. Tjuven halkade i leran
8. Morfar provade för stora skor
9. De gula glasögonen är fina
10. Busschauffören äter kex

List 8

1. Doktorn duschade på kvällen
2. Fotbollen försvann bakom huset
3. Flickan svalde ett tuggummi
4. Farbrorn svettades när han sprang
5. Läraren snubblade i trappan
6. I källaren stod en blå byrå
7. Mannen kramar sina söner
8. Grusvägen leder till ett rött hus
9. Pojkarna sjöng vackert igår
10. Mannen skar sig i pekfingret

List 9

1. Min syster fyller snart tjugofyra
2. Barnvakten kom klockan tio
3. Grannens pojke har svart hår
4. En blå duk låg på pianot
5. Golfspelaren skadade axeln
6. Lillebror försov sig i morse
7. Killarna retades med fröken
8. Tuppen står mitt på vägen
9. Mattan i köket är handvävd
10. Mannen planterade rädisor

List 10

1. Kocken skaffade glasögon
2. Hon målar hjärtat blått
3. Pojkarna lekte i parken
4. Det står tjugo bänkar längs kajen
5. Svärmor bjöd på middag igår
6. Gräset var vitt av frost
7. Mannen slog sönder en tand
8. Grannen låste sin cykel
9. Smöret smälter i stekpannan
10. Han förlorade partiet

List 11

1. Tanten rullade bollen
2. Skivan har ett rött fodral
3. Mamma skickar fjorton julkort
4. Vännerna var på bio igår
5. Sonen kör traktor på åkern
6. Han svettades under matchen
7. En grön cykel stod olåst
8. Tjejerna spelar handboll ikväll
9. Köket ska städas nästa lördag
10. Det är picknick på stranden vid sju

List 12

1. Bilen kör på skogsvägen
2. Skjortorna ska hängas i skåpet
3. De små tomaterna är gröna
4. Refrängen kom tre gånger
5. De tvättade i kallt vatten
6. Flickan fick sjutton julklappar
7. Paret gick genom parken
8. Matchen spelades i gassande sol
9. Ödlan smet ner under bron
10. En häst betade bland fåren

List 13

1. Han brände sig på strykjärnet
2. Växten i fönstret har en blomma
3. Han tappade kniven i sjön
4. Tolv muffins gräddades i ugnen
5. Farmor åker till golfbanan
6. Såsen kokade över
7. Två muggar gick sönder
8. De sprang runt det gula huset
9. I aulan hänger tolv tavlor
10. Flickan lyssnade på talet

List 14

1. Morfar fiskar hela dagarna
2. Pappa tvättar den blå bilen
3. Målaren beställde två öl
4. Det finns en spegel i sovrummet

5. Äggen ska kokas sju minuter
6. Flickan känner min kusin
7. Spaden i sandlådan var gul
8. Min morbror ska cykla till Norge
9. Två svarta skjortor hängde på tork
10. Flickan hittade inte rätt väg

List 15

1. Läraren tränar efter skolan
2. En gul ros blev gruppens symbol
3. Tyskarna bodde hos morfar
4. Ljuset brinner långsamt
5. Mamma ska baka en paj
6. Flickan sökte en vit sjal
7. Pojken hann tvätta arton bilar
8. Pappa slängde den brända kakan
9. Tjejen kom in i rummet
10. Den vita stolen står på gruset

List 16

1. Hon sköljde kläderna i bäcken
2. Det står en blomma på pallen
3. Köksklockan går fel
4. Det stod en ren vid vägkanten
5. En röd lampa tänds
6. Flickan säljer glass på stranden
7. Stadsborna åker till havet
8. Grodan ska fånga en fluga
9. Mamma strök byxorna
10. Mormor ska till Grekland i morgon

List 17

1. Ledarna ordnade en lekkväll
2. Han åker skridskor med barnen
3. Lillasyster öppnade paket
4. Pojken åt allt utom kärnhuset
5. Femåringen samlade burkar
6. Pappa har bytt frisyr
7. Den svarta hunden var hungrig
8. Tanten plockade snäckor
9. Lillebror vattnade gräsmattan
10. Gästen bad om en filt

List 18

1. Anden simmade i dammen
2. Bollen studsade ut på vägen
3. Engelsmannen spelar tuba
4. Hon köpte sex liter mjölk
5. Farfar lagar mat åt barnen
6. Mannen vaknade på soffan
7. Pojken skrattade åt sin kompis
8. Ungdomarna köper varsin glass
9. Flickan har kort rött hår
10. Ett grönt kuvert låg kvar i köket

List 19

1. De röda bladen föll från trädet
2. Det finns två gungor i parken
3. Bilen står fem gator bort
4. Dansken reser till Sverige
5. Flickan spiller sås på tröjan
6. Han skänkte vinsten till ett sjukhus
7. Killarna spelar squash på lördag
8. Tjejen hoppade i snödrivan
9. Alla äpplen låg på marken
10. Katten ska få ungar

List 20

1. Tanten säljer en gammal soffa
2. Vi ska sy åtta par byxor
3. Det låg fem kritor i badkaret
4. Hon dök ner i floden
5. Temperaturen sjönk under noll
6. Optikern ser en fågel
7. Fem svarta valpar låg på filten
8. Studenten vann en biobiljett
9. Barnen smög i buskarna
10. Kvinnan dansade hela natten

List 21

1. Staketet målades vitt
2. Längs muren växte röda liljor
3. Barnen paddlade runt sjön
4. Fyra män sjöng i kyrkan
5. Mannen låg på stranden
6. Läkaren hälsar på sin dotter
7. Blommorna i vasen är blåa
8. Kvinnan i kiosken har vitt hår
9. Flickan letar efter strumporna
10. Pojken såg det gröna tyget

List 22

1. Kossan betar grönt gräs i hagen
2. Marken blir hård när det är tjäle
3. Farmor köpte femton fiskar
4. Valpen svalkade sig i skuggan
5. Kvinnan hjälper sina barn
6. Den blå flaskan står på hyllan
7. Stegen är två meter lång
8. Ungarna satt framför datorn
9. Pojken hade sönder staketet
10. Rektorn rättade till slipsen

List 23

1. Pojken har tretton kusiner
2. Fyra gula bilar stod i kön
3. Tavlan var målad i rött
4. Mattan vädrades på balkongen

5. Fem tjejer sprang på skolgården
6. Huset hade nio sovrum
7. Tjusningen med vintern är snön
8. Farmor rensade i klädskåpet
9. Flickan handlade ost och korv
10. Barnet lärde sig alfabetet

List 24

1. Soldaten springer över bron
2. Chefen pratade med honom
3. I glaset fanns en röd servett
4. Hundarna rullade runt i snön
5. Flickan lånade ut sitt hårband
6. Bilen fick två repor
7. Vaktmästaren städar matsalen
8. Killen målade om väggarna
9. Byxorna glömdes i torkskåpet
10. Jackan hängde i garderoben

List 25

1. Elefanten åt ur en blå hink
2. Servitrisen bar ut grädden
3. Under fönstret stod en säng
4. Det fanns mycket svamp i gläntan
5. Klockan i köket slog tolv
6. Pojkarna sprang nedför backen
7. Flickan har tio väskor
8. Dockan låg gömd i buskarna
9. Kompisarna delar på pizzan
10. Grannen lånade en mejsel

Practice list 1

1. Kvinnan plockar fram sin medicin
2. Killen köpte en glass
3. Han sprang fort genom skogen
4. Svenskarna dricker hemgjord saft
5. Morfar odlar potatis
6. Grannen hade fest igår
7. Pojken kräver högre veckopeng
8. Barnet lekte med hunden
9. Båda tröjorna var svarta
10. Lyktan gav ett grönt sken

Practice list 2

1. Kvinnorna satt på bänken
2. Pojkarna lekte på stranden
3. Syltburken står på köksbordet
4. Fåtöljen står en trappa upp
5. Det låg en vit snäcka på botten
6. Mamma dricker ur sin nya kopp
7. Boken är i trädgården
8. Skivan innehöll femton låtar
9. Hennes skor är svarta
10. Handskarna låg på hatthyllan

Practice list 3

1. Kusinerna satt i fåtöljen
2. Mormor äter gröt varje dag
3. Damen handlade sju skjortor
4. Sönerna pussade sin mamma
5. Pappa väntade utanför
6. Killarna stod framför kiosken
7. Mamma stökade till i köket
8. Det sitter tio män i bastun
9. Brandbilen välte ner i diket
10. Trädet hade bara sex löv kvar

Practice list 4

1. Kvinnan hade blå naglar
2. Pojken i grönt åt tre portioner
3. Barnen fick måla nere vid ån
4. På lovet ska jag till simhallen
5. Det låg tretton knivar i lådan
6. Påsen innehöll ett gult paket
7. Författaren såg en pjäs
8. Tjejen satt i trappan
9. Barnen lekte framför huset
10. Löparen rengjorde sina skor

Practice list 5

1. Klänningen i skyltfönstret är vit
2. Affärsmannen körde för fort
3. Servitrisen tappar besticken
4. Månen lyste på de sex männen
5. I diskhon låg tjugo gafflar
6. Flickan stod länge i duschen
7. Han tappade en sko
8. Åtta tallrikar föll från hyllan
9. Fem filmer recenserades
10. En spegelram målades grön