# FP6–004171 HEARCOM

# Hearing in the Communication Society

**INTEGRATED PROJECT**

**Information Society Technologies**

# D-7-1: Speech recognition tests for different languages

| | |
|---|---|
| Contractual Date of Delivery: | August 31, 2005 (+ 45 days) |
| Actual Date of Delivery: | September 30, 2005 |
| Editor: | Rob Drullman |
| Sub-Project/Work-Package: | SP3/WP7 |
| Version: | 2.0 |
| Total number of pages: | 26 |

| Dissemination Level | | |
|---|---|---|
| PU | Public | X |
| PP | Restricted to other programme participants (including the Commission Services) | |
| RE | Restricted to a group specified by the consortium (including the Commission Services) | |
| CO | Confidential, only for members of the consortium (including the Commission Services) | |
| Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) | | |

# Deliverable D-7-1

| VERSION DETAILS | |
|---|---|
| Version: | 2.0 |
| Date: | 29 September 2005 |
| Status: | Final |

| CONTRIBUTOR(S) to DELIVERABLE | |
|---|---|
| **Partner** | **Name** |
| NL-TNO | Rob Drullman |
| DE-HZO | Kirsten Wagener |
| BE-LEU | Jan Wouters, Koen Eneman |
| | |
| | |
| | |

| DOCUMENT HISTORY | | | |
|---|---|---|---|
| **Version** | **Date** | **Responsible** | **Description** |
| 0.1 | 14.02.2005 | Rob Drullman | First draft outline |
| 0.2 | 21.04.2005 | Rob Drullman | Revised outline + content |
| 1.0 | 30.06.2005 | Rob Drullman | complete document |
| 2.0 | 30.09.2005 | Rob Drullman | revised final version |
| | | | |
| | | | |

| DELIVERABLE REVIEW | | | |
|---|---|---|---|
| **Version** | **Date** | **Reviewed by** | **Conclusion*** |
| 1.0 | 16.09.2005 | Stig Arlinger | Accept with minor modifications |
| 1.0 | 29.09.2005 | Tammo Houtgast | Accept |
| | | | |

* e.g. Accept, Develop, Modify, Rework, Update

# Table of Contents

# List of Tables

# 1 Executive Summary

This report describes the currently available material for testing speech intelligibility in different languages. We selected those tests that are most promising in view of the required harmonisation across European countries. Both speech material and measurement procedures were studied, with the focus on tests to determine the speech reception threshold (SRT) in noise. Standardised measurement procedures will be used within HEARCOM for the purpose of rehabilitation of the hearing impaired and evaluation of signal enhancement techniques.

The speech tests considered in this report are based on sentences, monosyllabic words (consonant-vowel-consonant – CVC), or number triplets. For sentences and CVCs we selected those tests that are frequently used in research and/or clinical practice. Number triplet tests have gained new interest (see below) as a method for quick and reliable screening of hearing acuity.

The survey of speech tests in this report can be used to give researchers and clinicians some guidance in finding an appropriate test, given the information they need. This may be a thorough insight in the communication abilities under adverse conditions, or only a quick screening. For each test a number features are given, such as specific properties of the material (number of lists, items per list), type of speaker, SRT measurement procedure, reference data for normal-hearing listeners, and suitability of the test for severely hearing impaired or cochlear-implant patients.

The speech tests mentioned in this report are given for the following seven languages: Danish, Dutch, (British) English, French, German, Polish, and Swedish. While most languages have one or more standardised tests for sentences and CVCs, some do not (yet) cover the complete range. This applies to French and Polish in particular. The number triplet test has recently become quite popular, because of its powerful and reliable performance, relatively easy construction, and ease of (unsupervised) use over telephone or the internet. Following the Dutch/German examples, number triplet tests are also being developed for the other languages within the HEARCOM project.

# 2 Introduction

There are a large number of different tests in Europe to measure speech perception in various conditions. The nature and availability of these tests, however, varies from country to country and language to language, as does the frequency of use in research and in clinical practice. In view of the rehabilitation of the hearing impaired and the evaluation of speech enhancement techniques, HEARCOM WP7 studies the available test material and measurement procedures. The purpose is to select the ones most suitable for the required harmonisation across different languages.

This deliverable is restricted to the evaluation of speech intelligibility tests. The material and procedures treated here have a close relation to the work in WP1, viz. the definition of tests to estimate an individual's communication performance. A review of relevant sentence tests was consulted for the present report (Lyzenga, 2005). Other relevant material was found in the open literature and in the documentation of the former EU-project NATASHA (Network and Tools for the Assessment of Speech/Language and Hearing Ability), which made a first attempt to document and compare audiological measuring methods across Europe.

The speech tests described in this deliverable are for general-purpose evaluation of intelligibility, but will particularly be used for the evaluation of signal enhancement techniques developed in HEARCOM WP5 and for the validation of rehabilitation procedures developed in WP6. The enhancement techniques aim at improving the reception of speech signals by the hearing impaired (usually by increasing the signal-to-noise ratio). Signal enhancement can be used in hearing assistive devices or in telephone or public address (PA) systems. For rehabilitation purposes, diagnostic speech tests can be used in the fitting of hearing aids and cochlear implant devices.

Speech recognition tests can be considered at different levels. The material involved can consist of sentences, words, or numbers, which will be outlined in the sections below. The choice of materials and procedures will depend on several factors, among which the nature of the hearing impairment. Our intention is that the survey in this report will give future users of speech reception tests a good starting point to make a selection for an appropriate test, given the information they need. This may be a thorough insight in the communication abilities under adverse conditions, or, on the other hand, only a quick screening.

In section 3 we will present the different speech tests that should enable us to obtain comparable data on speech recognition across languages. Section 4 describes ways in which tests are (or will become) available to the community and section 5 will give the main conclusions.

# 3 Speech material and measuring procedures

## 3.1 Basic considerations

The basic goal of the speech tests considered here is to measure the speech reception threshold (SRT) against a background of noise. That is, we want to find the speech level that yields 50% intelligibility. For the noise we will in principle take the common stationary speech-shaped noise, rather than fluctuating noise or multi-talker babble. In most measurement procedures the speech signals are presented at a fixed or varying level relative to the noise, and the items (sentences, keywords, syllables, …) which are reproduced correctly are scored for each individual subject. The SRT is usually expressed as a signal-to-noise ratio (SNR) in dB. The value of the SRT depends on the speech material and on the capacities of the listener (normal hearing or hearing impaired). For the purpose of this report, we assume the use of native listeners for each language. Although speech tests may very well be employed to measure a subject's capability to understand speech in a different language, this is not dealt with here.

Given the many different materials and different talkers, it is an illusion to strive for identical SRT-values across languages. Each test will have its specific reference score for (young) normal-hearing listeners. We should aim at a maximum degree of similarity across the tests for the difference-scores between the various conditions (degradations, simulations). That is, the difference between scores for clean speech and scores for degraded speech should be similar for all languages involved (for identical types of degrading). On the other hand, given clean speech, the difference between scores for normal-hearing and hearing impaired subjects with a specific auditory profile should also be similar across languages. So, the objective is to harmonise the delta-scores on well-defined conditions, not the absolute scores.

### 3.1.1 Test types

When considering a speech test, the type of speech material is an important aspect. In this report we will consider three major types of tests: sentences tests, word (CVC – consonant-vowel-consonant) tests, and number triplet tests. For some languages there are more tests of a specific type, intended for a wide range of normal-hearing and hearing-impaired subjects. We have selected those tests that are easily available and frequently used in research and/or clinical practice.

#### 3.1.1.1 Sentence tests

The use of sentence materials has many advantages. Sentences are representative for everyday communication (more than single words, syllables, or phonemes), involving a context and incorporating the whole language system. Moreover, from a measurement point of view, the psychometric function is steeper, making sentences particularly suitable for accurate estimation of the SRT. A disadvantage of the use of (a limited set of) sentences is that the material can often only be used once for each subject and that the procedure usually requires the presence of an instructor. However, modified methods for sentence tests overcome these problems to a large extent (see below).

Basically, two types of sentence tests will be discussed in this report. Firstly, tests which use lists of everyday sentences (5-8 words) with varying semantic and syntactic structures, and similar intelligibility for individual sentences, as developed for Dutch by Plomp (Plomp & Mimpen, 1979). This type of sentence tests will therefore be referred to as Plomp-tests. Secondly, tests which use lists of sentences with a fixed grammatical structure, each consisting of 5 words (name-verb-number-adjective-object). There is a base list consisting of ten sentences with five words each. The test lists were generated by choosing one of the ten alternatives for each word group in a pseudo-random way that used each word of the base list exactly once. Although first described for Swedish by Hagerman (1984), this type of test is commonly known within HEARCOM as OLSA-test, after the German version (**Ol**denburger **Sa**tztest; Wagener et al., 1999; for a short English description see Wagener and Brand, 2005). The OLSA-test sentences have low redundancy, and can be used more than once with the same subjects.

As mentioned by Lyzenga (2005), sentence tests differ in the way responses are collected from the listeners. Plomp-tests use sentences that have to be reproduced as closely as possible. Scoring may be based on whole sentence correct, separate words, or keywords. These tests normally use an open response set. OLSA-tests are mostly used as open tests as well, although it is possible to use them as closed tests. Such tests are mostly performed by elderly hearing-impaired listeners. They may be confused by a closed response set, leading to incorrectly poorer intelligibility results. In general, open-set tests are better for quick and accurate assessment of speech recognition, because of relatively small learning effects. Closed-set tests are better for prolonged and/or repeated testing, as redundancy is lower (although the learning effect is stronger).

#### 3.1.1.2 CVC tests

For a more diagnostic speech test, words rather than sentences may be used. We will restrict ourselves to tests employing CVC syllables, either meaningful or nonsense, that are presented against a noise background

(or in quiet). CVC tests are quite common in the clinical practice. Apart from determining the SRT or a mere percentage correct score, the subjects' responses can be analysed on a phonemic level. The confusions listeners make can give a more detailed indication of the (loss of) hearing acuity. CVC syllables can be presented in isolation or in a carrier phrase, and, depending of the measurement procedure, data can be obtained with or without an instructor.

As with the sentence tests, either open-response sets or closed-response sets can be used for CVC tests. In the latter case, the paradigm is that of a rhyme test, with about 6 alternatives for either the initial or the final consonant (cf. House et al., 1965). In some tests, the middle vowel is also tested (cf. Wallenberg and Kollmeier, 1989).

#### 3.1.1.3 Number triplet tests

Finally, we will consider a speech-in-noise test with digit triplets as speech material. This is a relatively quick and easy test for screening purposes (20-30 triplets in a list) with high sensitivity and specificity. The test can be repeated with the same subject (low risk of remembering triplets), is suitable for a large number of hearing impaired subjects, and can be done by telephone or over the internet (Smits et al., 2004; Smits and Houtgast, 2005). No instructor is needed to administer the test. The listener has to indicate which 3 digits were spoken in noise (in the correct order), in a closed response-set. The psychometric curve shows a steep slope of the psychometric curve, hence enabling accurate determination of the SRT.

### 3.1.2 Languages

In view of the composition of the HEARCOM consortium and the availability of speech material, the following languages will be treated in this report: Danish, Dutch, (British) English, French, German, Polish, and Swedish. As will be clear from the remainder, some languages have many (accepted and validated) speech tests available, whereas other languages do not. Some tests are under development, and will be available during the HEARCOM project (WP1 or WP7) by the end of 2005 or early in 2006.

### 3.1.3 Subjects

The speech tests are generally suitable for normal hearing subjects. The tests described here also assume that the subjects are adults (or adolescents) and not children. Particular tests have been developed for (very) young children, but this report will not focus on those tests. The use of normal hearing subjects is an important point here: not to test their normal performance, but rather to test the effects of simulated hearing impairment and to validate the performance of signal processing algorithms.

Not all speech tests are suitable for the hearing impaired. Experience has shown that Plomp-tests with an open-response set and a demand to reproduce the entire sentence, cannot be used for all severely hearing impaired subjects or subjects with cochlear implants. Sentences are sometimes spoken at a fast rate, leading to considerable coarticulation and assimilation. Plomp-tests in which only (key)words are scored, or OLSA-tests with easy word material and a medium speech rate are more suitable for these kinds of subjects. In the sections to follow, it will be noted whether tests are unsuitable for specific groups of hearing impaired.

### 3.1.4 Speech and noise levels

As mentioned in paragraph 3.1, the speech tests considered in this report focus on measuring speech intelligibility in noise. Speech and noise material for SRT tests can be defined in different ways, but the common aspect is that the results (xx% intelligibility score, were xx = 50 in most adaptive procedures) are expressed in terms of a SNR. It should be noted that SNR = 0 dB does not necessarily mean that the individual speech item (sentence, CVC, number triplet) has the same level as the noise. For an entire list of speech items this may be true (long-term), but not always for each single item. For example, in the original Plomp-test for Dutch (Plomp and Mimpen, 1979) sentence lists were constructed on the basis of iso-intelligibility, with different levels per sentence. The SRT-test developed later at NL-VUMC (Versfeld et al., 2000) uses sentences with the same (linear) rms as the noise.

When considering speech and noise levels, one needs to know how these levels were defined for the various test materials. This concerns frequency weighting (linear, A- or C-weighting) and time weighting (e.g., rms, equivalent SPL, discarding of low-level parts, …). It is beyond the scope of the current report to discuss the speech and noise level issues for each test separately – details can be found in the papers referenced. However, the user of a speech tests must be aware of the fact that the particular speech and noise materials belong together. The (normal-hearing) reference data for the SRTs mentioned in the next sections are based on these default data .

## 3.2 Sentence tests

### 3.2.1 Danish

The basic Danish speech material DANTALE, designed by Elberling et al. (1989), does not contain a sentence test in noise. Recently, such an SRT test was developed by Wagener et al. (2003). This sentence test is called DANTALE II and is an OLSA-type test. Main features of this test are:

- 16 tests lists of 10 sentences with a fixed structure (5 words), which can be combined to lists of 20 sentences
- untrained female speaker
- SRT in speech-shaped noise measured either
  - at fixed SNRs or
  - in an adaptive procedure
- mostly open response set, closed-response set is possible, scoring of words correct
- normal hearing references (logistic model function, unmodulated speech-shaped noise): SRT = –8.4 dB, slope at SRT = 13.2%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

### 3.2.2 Dutch

The basic Dutch SRT test for sentences in noise is the Plomp-test (Plomp & Mimpen, 1979). It is a standard in audiological research and in clinical practice. The main features are:

- 10 lists of 13 everyday sentences (8-9 syllables)
- trained female speaker
- SRT in speech-shaped noise measured in an adaptive procedure (monaural)
- open-response set, scoring of whole sentence correct
- normal hearing references: SRT = –5.6 dB, slope at SRT = 15-20%/dB
- not suitable for severely hearing-impaired and cochlear-implant subjects

A similar test, but with an untrained male speaker, was later developed by Smoorenburg (1992). It follows the same procedure. Reference data for this test are: SRT = –5.1 dB, slope at SRT = 17.7%/dB.

In order to create a large set of sentence material, Versfeld et al. (2000) developed another test (VU98), with the following features:

- 78 lists of 13 everyday sentences (39 lists per speaker)
- trained male and a trained female speaker
- SRT in speech-shaped noise measured in an adaptive procedure (monaural)
- open-response set, scoring of whole sentence correct
- normal hearing references:
  - female speaker: SRT = –4.1 dB, slope at SRT = 16.6%/dB
  - male speaker: SRT = –4.0 dB, slope at SRT = 15.2%/dB

- not suitable for severely hearing-impaired and cochlear-implant subjects

Because of the encountered difficulties in using the existing Dutch sentence tests with cochlear-implant subjects, a new test was developed in Belgium – the Leuven Intelligibility Sentence Test (LIST; Van Wieringen & Wouters, 2005). It is basically a Plomp-type test, but with simple speech material and scoring of keywords instead of entire sentences. The LIST speech material is highly articulated at a relatively low rate. The main features are:

- 35 lists of 10 sentences (90 syllables per list)
- trained female speaker
- SRT in speech-shaped noise measured in an adaptive procedure or at fixed SNR (monaural)
- open-response set, scoring of keywords (32-33 per list)
- normal hearing references: SRT = –7.8 dB, slope at SRT = 17.5%/dB
- particularly suitable for severely hearing-impaired and cochlear-implant subjects

A Dutch version of the OLSA-test has been developed (HEARCOM WP1). It is expected to be available by the end of 2005.

### 3.2.3 English

The Adaptive Sentence Lists (ASL) were developed for British English by Macleod & Summerfield (1990). The speech material was recorded for both auditory and audio-visual testing, but will consider the auditory test only. It is a Plomp-type sentence test, with simple speech material and scoring of keywords instead of entire sentences. The main features are:

- 10 lists of 15 sentences (4-6 words)
- untrained male speaker
- SRT in lowpass filtered (10 kHz cutoff) white noise measured in an adaptive procedure (monaural)
- open-response set, scoring of keywords (3-4 per sentence), 'loose' scoring, i.e., ignoring errors in case or declension
- normal hearing references: SRT = –16.8 dB, slope at SRT = 9.9%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects (audio-visual version also available)

An English version of the OLSA-test has been developed (HEARCOM WP1). It is expected to be available early in 2006.

### 3.2.4 French

It appears that different speech tests are used in France, but little is known about standardisation. Audiological centres have custom-developed sets, and many evaluations are made with live voice, sometimes recorded. A recent paper in French by Wable (2001) and a CD with accompanying speech material was made available through HEARCOM-partner BE-LEU. It is a Plomp-type test, with the following main features:

- 17 lists of 8 sentences (5-8 words)
- untrained male speaker
- SRT in speech-shaped noise measured in an adaptive procedure (monaural)
- open-response set, scoring of whole sentence correct
- normal hearing references: not known
- suitable for hearing-impaired subjects (see below).

To our knowledge, this test was administered with a large number of hearing-impaired subjects (with $PTA_{0.5,1,2}$ of about 45 dB). A proper validation yielding reference data for normal-hearing subjects is lacking. Within the HEARCOM-project it was agreed that BE-LEU will perform the validation on the available sentence material. Data are expected to be published by the end of 2005.

Within the HEARCOM-project (WP7) a French OLSA-test will also be developed. Recordings were made and evaluation will start at the beginning of 2006.

### 3.2.5 German

Kollmeier & Wesselkamp (1997) developed the Göttingen sentence test with objective and subjective assessment procedures, as an extension of older existing German tests. We will only consider the objective intelligibility assessment. The Göttingen test is a Plomp-type test which can both be used to measure speech performance at fixed SNRs or adaptively determine the SRT. The adaptive procedure was developed by Brand & Kollmeier (2002). The test has the following features:

- 20 lists of 10 everyday sentences (5-8 words)
- untrained male speaker
- Sentences presented in speech-shaped noise (monaural), either
  - at fixed SNRs
  - in an adaptive procedure

- open-response set, scoring of words correct (weighted to obtain a sentence score)
- normal hearing references (logistic model function, speech-shaped unmodulated noise):
  - with fixed SNRs: SRT = –6.2 dB, slope at SRT = 19.2%/dB
  - adaptive method: SRT = –5.6 dB, slope at SRT = 16-20%/dB
- suitable for moderately hearing impaired, too difficult for severely hearing impaired and for cochlear-implant subjects due to a rather fast speech rate.

The second German test for sentences in noise is the Oldenburg sentence test (Wagener et al., 1999; Wagener and Brand, 2005). This OLSA-test can both be used for testing speech recognition against fixed-level noise or in an adaptive procedure. The same adaptive procedure as for the Göttingen sentence test is used in this test (Brand and Kollmeier, 2002). The main features are:

- 10 test lists of 10 sentences with a fixed structure (5 words), combined to lists of 20 or 30 sentences
- untrained male speaker
- Sentences presented in speech-shaped noise (monaural), either
  - at fixed SNRs
  - in an adaptive procedure
- mostly open-response set, closed-response set is possible, scoring of words correct
- normal hearing references (logistic model function, speech-shaped unmodulated noise):
  - with fixed SNRs: SRT = –7.1 dB, slope at SRT = 17%/dB
  - adaptive method: SRT = –6.3 dB, slope at SRT = 14-16%/dB
- suitable also for severely hearing impaired and cochlear-implant subjects.

### 3.2.6 Polish

There are no validated Polish sentence tests of the Plomp- or OLSA-type available at this moment. Preparations for developing the speech material under the current HEARCOM-project have started, though (Ozimek et al., 2005).

The speech material consists of 20 lists of 12 Plomp-sentences and 20 lists of 12 OLSA-sentences, recorded with a trained male and female speaker. Preliminary SRT-measurements were done in speech-shaped noise at fixed SNR levels. Optimisation of speech material and procedure is

currently carried out, and normal-hearing reference data are expected to be available by the end of 2005 or early 2006.

### 3.2.7 Swedish

The standard Swedish sentence test in noise was developed by Hagerman (1982). This test was the prototype for the OLSA-tests. The original test was done by presenting sentences at fixed SNRs. Later on an adaptive paradigm (converging to 40% intelligibility) was evaluated (Hagerman & Kinnefors, 1995). The test has the following features:

- 11 tests lists of 10 sentences with a fixed structure (5 words)
- untrained female speaker
- SRT in speech-shaped noise (slightly amplitude modulated at 2.1 Hz) measured monaurally either
  - at fixed SNRs
  - in an adaptive procedure
- mostly open-response set, closed-response set is possible, scoring of words correct
- normal hearing references:
  - with fixed SNRs: SRT (50% int.) = –8.1 dB, slope at SRT = 25%/dB
  - adaptive method: SRT (40% int.) = –7.8 dB, slope at SRT = ?
- suitable for severely hearing-impaired and cochlear-implant subjects

Recently, new Swedish sentence material was developed for a Plomp-like test, to be used in fixed and adaptive procedures (Hällgren et al., 2005). The main features are:

- 25 lists of 10 everyday sentences (5-9 syllables each)
- Trained female speaker (male speaker in preparation)
- Noise spectrum filtered to be equal to spectrum of the actual speech, unmodulated noise.
- Speech level measured as C-weighted equivalent sound pressure level with silent intervals between sentences eliminated. Noise level measured as C-weighted equivalent sound pressure level.
- Open response set with adaptive procedure
- Scoring of whole sentence correct and single words correct
- Normal hearing references:

  SRT = –3.0 dB, slope 17.9%/dB for sentence scoring

  SRT = –4.9 dB, slope 15.4%/dB for word scoring
- Limited clinical experience with the material as yet.

## 3.3 CVC tests

### 3.3.1 Danish

The DANTALE speech material (Elberling et al., 1989) contains 'adult' word lists consisting of common nouns, adjectives and verbs. Normative data for the speech material was later presented by Keidser (1993). The main features of the test are:

- 8 lists of 25 meaningful monosyllables (80 phonemes per list); no carrier phrase.
- trained female speaker
- SRT in amplitude-modulated speech-shaped noise, measured at fixed SNRs (monaural)
- open-response set, scoring of whole words, phoneme scoring, or triple score (CVC)
- normal hearing references (word scoring in noise):
  - SRT = -8.7 dB SNR, slope at SRT = 6 %/dB
- suitable for severely hearing-impaired and cochlear-implant subjects (audio-visual version also available)

### 3.3.2 Dutch

The NVA-test (NVA: Dutch Audiological Society) is the common word intelligibility test for Dutch, based on work by (Bosman, 1989). It uses lists of meaningful CVC words and has the following features:

- 45 lists of 33 meaningful CVC syllables; no carrier phrase.
- trained female speaker
- SRT in speech-shaped noise measured at fixed SNRs (monaural)
- open-response set, scoring of whole words
- normal hearing references: SRT = –10.1 dB; slope at SRT = 6.0%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

A southern Dutch (Flemish) variant of the NVA-lists has also been recorded with a Flemish male voice (Wouters et al., 1994).

### 3.3.3 English

A British English CVC test was developed in different stages, as described by Markides (1978). Based on the isophonemic word lists constructed by Arthur Boothroyd (Boothroyd, 1968), this test is known as AB word test. The main features are:

- 15 lists of 10 meaningful CVC syllables; no carrier phrase.
- untrained male speaker
- SRT in quiet (monaural)
- open-response set, scoring of whole words
- normal hearing references (adults, audio-cassette recordings): SRT = 32 dB; slope at SRT = 5.1%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

The Four Alternative Auditory Feature test (FAAF) was developed and evaluated by Foster and Haggard (1986). It uses meaningful CVCs in multiple pairs with minimal phonemic differences. The main features of the FAAF-test are:

- 80 meaningful CVCs in 20 sets of 4 binarily and minimally paired words; embedded in a carrier phrase.
- untrained male speaker
- SRT in speech-shaped noise, measured at fixed SNRs (monaural)
- closed-response set, four-alternative forced-choice (4AFC) scoring of whole words
- normal hearing references: SRT = –8.6 dB; slope at SRT = 6%/dB
- maybe suitable for severely hearing-impaired and cochlear-implant subjects (but with low scores)

### 3.3.4 French

No standardised word test is known to us. The former EU-project NATASHA mentions some acoustic sets, e.g., CVCVs with 20 consonants and semi-consonants, originally constructed by Fournier ("Test d'autiométrie vocale," Dahlberg Audible International CD & booklet), but details and reference data are not available.

### 3.3.5 German

The Freiburger word test (Hahlbrock, 1953) is a common test in the German clinical practice. The test is only used to determine the benefit in hearing aid provision. It uses lists of meaningful words and has the following features (CVC part only):

- 20 lists of 20 meaningful CVCs, no carrier phrase
- trained male speaker
- SRT in quiet
- open-response set, scoring of whole words

- normal hearing references (DIN 45621-1, 1995): SRT = 29.3 dB SPL, slope at SRT = 4%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

The monosyllabic rhyme test (Sotscheck, 1982; von Wallenberg and Kollmeier 1989) uses meaningful CVCs and a closed-response set. Five response alternatives per test word vary either in the initial consonant, the middle vowel, or the final consonant. It has the following features:

- 10 lists of 72 meaningful CVCs (possibly abbreviated to 47 or 25 CVCs), carrier phrase or 3 repetitions of the test word
- untrained male speaker
- SRT in quiet or in noise
- closed response set (5 response alternatives), scoring of whole words (equivalent to scoring of target phoneme)
- normal hearing references[1]:
  - 72 CVCs in quiet: SRT = 14.1 dB SPL, slope at SRT = 4.8%/dB
  - 47 CVCs in quiet: SRT = 16.1 dB SPL, slope at SRT = 5.4%/dB
  - 25 CVCs in quiet: SRT = 13.8 dB SPL, slope at SRT = 5.8%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

### 3.3.6 Polish

There are a few Polish word tests used in the clinical practice, but it is hard to find standardised procedures. A test developed by Pruszewicz et al. (1994a, b) consists of 10 lists of 24 words from among the most frequent monosyllabic Polish nouns. The lists are (poorly) balanced phonemically, semantically, acoustically and structurally. The test was not evaluated in noise and gives rather high standard deviations for the SRT in quiet.

The development of a Polish CVC test is not foreseen within the HEARCOM project. Priority is given to the development of Polish sentence tests.

### 3.3.7 Swedish

A test with monosyllabic words was developed by Magnusson (1995), based on old standardised speech material from the 1950s. Not all words

[1] In fact, the values given refer to $L_{mid}$ which slightly differs from the SRT. Based on a logistic model function and considering random hit rate, the midpoint level $L_{mid}$ of the intelligibility function corresponds to 0.5 × (maximum intelligibility + random hit rate).

in this list are truly CVC (some have double initial or final consonants), but we have nevertheless decided to mention the test here. Its main features are:

- 12 lists of 50 meaningful syllables (not all CVC), preceded by a carrier phrase.
- trained male speaker
- SRT in speech-weighted noise measured at fixed SNRs (monaural)
- open-response set, scoring of whole words
- normal hearing references: SRT = –2.8 dB; slope at SRT = 6.5%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

## 3.4 Number triplet tests

### 3.4.1 Danish

The DANTALE speech material (Elberling et al., 1989) contains digit triplets to determine the SRT in noise. Evaluation of the lists was done by Olsen (1996). The test has the following features:

- 8 monosyllabic digits, combined to 3 lists of 20 triplets
- trained female speaker
- SRT in quiet or in amplitude modulated speech-shaped noise, measured at fixed SNRs (monaural)
- closed-response set, scoring of triplets
- normal hearing reference in quiet: SRT = 15 dB SPL, slope at SRT = 8.3%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

### 3.4.2 Dutch

A Dutch version of the number triplet test was recently developed for screening by telephone (Smits et al., 2004; Smits and Houtgast, 2005). The reference data given below, however, apply to headphone listening in the laboratory. The test has the following features:

- 8 monosyllabic digits, combined to a list of 23 triplets (with coarticulation)
- trained female speaker
- SRT in speech-shaped noise measured in an adaptive procedure (monaural)
- closed-response set, scoring of triplets

- normal hearing references: SRT = –11.2 dB; slope at SRT = 16%/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

An alternative Leuven Intelligibility Number Test (LINT) was developed by BE-LEU (Van Wieringen & Wouters, 2005). It is not a real triplet test, but a closed-response set test of numbers between 1 and 100. The test was especially constructed for use with severely hearing impaired and cochlear-implant patients. Main features of LINT are:

- 40 lists of 10 numbers between 1 and 100 (with coarticulation)
- 4 trained speakers (2 male and 2 female)
- SRT in speech-shaped noise measured in an adaptive procedure (monaural)
- closed-response set, scoring of whole number
- normal hearing references: SRT = –10 dB; slope at SRT = 15%/dB
- particularly suitable for severely hearing-impaired and cochlear-implant subjects

### 3.4.3 English

An English version of the number triplet test, following the German format, was developed and is currently under evaluation (HEARCOM WP1). It is expected to be available at the beginning of 2006.

### 3.4.4 French

A French version of the number triplet test, following the German format, is currently under development (HEARCOM WP1). It will be available in 2006.

### 3.4.5 German

A German version of the number triplet test was recently developed in Oldenburg (Wagener et al. 2005), with the following features:

- 9 monosyllabic digits, combined to a list of 27 triplets (with coarticulation)
- untrained female speaker
- triplets in speech-shaped noise to measure either
  - intelligibility at fixed SNRs or
  - adaptively SRT
- closed-response set, scoring of triplets

- normal hearing references (headphone listening): SRT = -9.3 dB SNR, slope at SRT = 19.6 %/dB
- suitable for severely hearing-impaired and cochlear-implant subjects

### 3.4.6 Polish

There is no Polish version of the number triplet test. At this moment, the development of such a test is not foreseen within the HEARCOM project. Priority is given to the development of Polish sentence tests.

### 3.4.7 Swedish

A Swedish version of the number triplet test, following the German format, is currently under development (HEARCOM WP1). It is expected to be available by the end of 2005 or at the beginning of 2006.

## 3.5 Compendium of the speech tests

For the sake of easy comparison, Table 1 below gives a survey of the main features of the different speech tests described in the previous sections.

**Table 1.** Main properties of the different speech tests per language.

| Speech material | Language | Test ID | SRT in noise (normal hearing) | Suitable for sev. HI / CI | Available |
|---|---|---|---|---|---|
| Plomp-type sentences | Danish | - | - | - | - |
| | Dutch | Plomp<br>Smoorenburg<br>VU98<br>LIST | –5.6 dB<br>–5.1 dB<br>–4.0 dB / –4.1 dB<br>–7.8 dB | no<br>no<br>no<br>yes | CD, software<br>CD<br>CD, software<br>CD |
| | English | ASL | –16.8 dB | yes | CD |
| | French | Wable | to do (BE-LEU) | ? | 2005/2006 |
| | German | Göttingen | –6.2 dB / –5.6 dB | no | CD, software |
| | Polish | under construction | - | - | End of 2005 |
| | Swedish | Hällgren | –3.0 dB / –4.9 dB | ? | 2005/2006 |
| OLSA-type sentences | Danish | DANTALE II | –8.4 dB | yes | CD |
| | Dutch | under construction | - | yes | End of 2005 |
| | English | under construction | - | yes | End of 2005 |
| | French | under construction | - | yes | 2006 |
| | German | Oldenburg (OLSA) | –7.1 dB / –6.3 dB | yes | CD, software |
| | Polish | under construction | - | yes | 2005/2006 |
| | Swedish | Hagerman | –7.8 dB / –8.1 dB | yes | CD |
| CVC | Danish | DANTALE | –8.7 dB | yes | CD |
| | Dutch | NVA | –10.1 dB | yes | CD |
| | English | AB<br>FAAF | only in quiet<br>–8.6 dB | yes<br>yes (?) | tape, CD<br>tape, CD |
| | French | Fournier | - | - | CD |
| | German | Freiburg<br>WaKo rhyme | only in quiet<br>only in quiet | yes<br>yes | CD, software<br>CD, software |
| | Polish | Pruszewicz | - | - | - |
| | Swedish | Magnusson | –2.8 dB | yes | CD |
| Number triplets | Danish | DANTALE | only in quiet | yes | CD |
| | Dutch | VU<br>LINT | –11.2 dB (headphone)<br>–10.0 dB (headphone) | yes<br>yes | CD, software<br>CD |
| | English | under construction | - | yes | 2006 |
| | French | under construction | - | yes | 2006 |
| | German | Oldenburg | –9.3 dB (headphone) | yes | CD, software |
| | Polish | - | - | - | - |
| | Swedish | under construction | - | yes | 2005/2006 |

# 4 Dissemination and Exploitation

The material and procedures described in this report are mostly of interest to audiological professionals (researchers and clinicians). This applies to sentence and VCV tests in particular. Professionals can find many of the test characteristics in the open literature (see references). Speech material and measurement procedures (software packages) can be made available by universities or research institutes, or sometimes by commercial parties in the field of audiological equipment. Depending on the specific test, a purchase or license fee may apply.

For end users (among whom hearing-impaired people) the number triplet test can be a simple way for automatic screening. This kind of self-test can be made accessible over the telephone or over the internet. Evaluation of the test by telephone in the Netherlands showed very good results and indicates that it fulfils the need for a functional hearing screening test, enhancing public awareness about hearing loss (cf. Smits and Houtgast, 2005).

# 5 Conclusions

This report gives a survey of speech tests in seven European languages. It focuses on tests to determine the SRT in noise for different kinds of speech material: sentences, CVCs, and number triplets. The tests described in this report can be used within HEARCOM for the purpose of rehabilitation of the hearing impaired and evaluation of signal enhancement techniques. Tests for seven languages are reported: Danish, Dutch, (British) English, French, German, Polish, and Swedish.

A summary of the main features is given for each SRT-test in a language: properties of the speech material (number of lists, items per list), type of speaker, SRT measurement procedure, reference data for normal-hearing listeners, and suitability of the test for severely hearing impaired or cochlear-implant patients.

Although we found one or more standardised tests for sentences and CVCs in most languages, we could not get full descriptions for some (i.c. French and Polish). However, development of new material and/or evaluation of existing material is or will be done within the HEARCOM project. This also applies to the number triplet tests, following the Dutch/German examples for a reliable automatic self-test.

# 6 References

Boothroyd, A. (**1968**). "Developments in speech audiometry," Br. J. Audiol. **2**, 3-10.

Bosman, A.J. (1989). *Speech perception by the hearing impaired*. Unpublished PhD. thesis Utrecht University.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801-2810.

DIN 45621-1 (**1995**). *Sprache für Gehörprüfung - Teil 1: Ein- und mehrsilbige Wörter.* Deutsches Institut für Normung.

Elberling, C., Ludvigsen, C., and Lyregaard, P.E. (**1989**). "Dantale: a new Danish speech material," Scand. Audiol. **18**, 169-175.

Foster, J.R. and Haggard, M.P. (**1986**). "The Four Alternative Auditory Feature test (FAAF)–linguistic and psychometric properties of the material with normative data in noise," Br. J. Audiol. **21**, 165-174.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**, 79-87.

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

Hahlbrock, K. H. (**1953**). "Über Sprachaudiometrie und neue Wörterteste," Arch. Obienheilk. 162, 394.

Hällgren M., Larsby B., Arlinger, S. (**2005**). "A Swedish version of the Hearing in Noise Test (HINT) for measurement of speech recognition," submitted for publication in Int. J. Audiol.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (**1965**). "Articulation-testing methods: Consonantal differentiation with a closed-response set," J. Acoust. Soc. Am. **37**, 158-166.

Keidser, G. (**1993**). "Normative data in quiet and in noise for "DANTALE"--a Danish speech material," Scand. Audiol. **22**, 231-236.

Kollmeier, B., and 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.

Lyzenga, J. (**2005**). *Literature review for HearCom, SP1, WP1.* BSCW server, WP1 Taks1.

Macleod, A., and Summerfield, A. Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise," Br. J. Audiol. **24**, 29-43.

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

Markides, A. (**1978**). "Speech discrimination functions for normal hearing subjects with AB isophonemic word lists," Scand. Audiol. **7**, 239-245.

Olsen, S.Ø. (**1996**). "Evaluation of the List of Numerals in the Danish Speech Audiometry Material," Scand Audiol **25**, 103-107.

Ozimek, E., Sek, A., Wicher, A, Kutzner, D., and Szczepaniak, O. (**2005**). "Intelligibility of the Polish sentence tests in interfering noise," Personal communication.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiol. **18**, 43-52.

Pruszewicz A, Demenko G, Richter L, and Wika T. (**1994a**). "New articulation lists for speech audiometry. Part I," Otolaryngol. Pol. **48**, 50-55. (in Polish)

Pruszewicz A, Demenko G, Richter L, and Wika T. (**1994b**). "New articulation lists for speech audiometry. Part II," Otolaryngol. Pol. **48**, 56-62. (in Polish)

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

Smits, C. and Houtgast, T. (**2005**). "Results from the Dutch speech-in-noise screening test by telephone," Ear Hear. **26**, 89-95.

Smoorenburg, G.F. (**1992**). "Speech reception in quiet and in noisey conditions by individuals with noise-induced hearing loss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421-437.

Sotscheck, J. (**1982**). "Ein Reimtest für Verständlichkeitsmessungen mit deutcher Sprache als ein verbessertes Verfahren zur Bestimmung der Sprachübertragungsgüte," Der Fernmeldeingenieur, **36**, 1-83.

Versfeld, N. J., Daalder, L., Festen, J. M., and 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.

Wable, J. (**2001**). "Normalisation d'un test de reconnaissance de la parole dans le bruit chez le sujet deficient auditif," Cahier de l'Audition **14**, 29-38.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a German sentence test part III: Evaluation of the Oldenburg sentence test)," Z. Audiol. **38**, 86-95.

Wagener, K., Josvassen, J.L., and Ardenkjær, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10-17.

Wagener, K.C., Brand, T. (**2005**). "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters," Int. J. Audiol. **44**, 144-157.

Wagener, K.C., Eenboom, F., Brand, T., and Kollmeier, B. (**2005**). "Ziffern-Tripel-Test: Sprachverständlichkeitstest über das Telefon," 8. Jahrestagung der DGA, Tagungsband

Wallenberg, E. L. von, and Kollmeier, B. (**1989**). "Sprachverständlich-keitsmessungen für die Audiologie mit einem Reimtest in deutscher Sprache: Erstellung und Evaluation von Testlisten," Audiol. Akustik. **28**, 50-65.

Wieringen, A. van, and Wouters, J. (2005). "LIST and LINT: Dutch speech audiometry lists with sentences and numbers," CD and booklet by Lab.Exp.ORL - NKO, University of Leuven. (in Dutch)

Wouters, J., Damman, W., and Bosman, A.J. (**1994**). "Vlaamse Opname van Woordenlijsten voor Spraakaudiometrie," CD and booklet.