FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-1-6b: Development of a digit-triplet test in modern Greek

Contractual Date of Delivery:	28 February 2009 (+45 days)
Actual Date of Submission:	07 April 2009
Editor:	Johannes Lyzenga
Sub-Project/Work-Package:	SP1/ WP1
Version:	2.0
Total number of pages:	23

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)

This information is confidential and may be used only for information purposes by Community Institutions to whom the Commission has supplied it

VERSION DETAILS	
Version:	2.0
Date:	7 April 2009 /22-2-2010 (public)
Status:	Final version

CONTRIBUTOR(S) to DELIVERABLE	
Partner	**Name**
NL-VUMC	Johannes Lyzenga
GR-ILSP	Theologos Athanaselis
GR-ILSP	Georgios Giannopoulos

DOCUMENT HISTORY			
Version	**Date**	**Responsible**	**Description**
1.0	04-04-2009	Johannes Lyzenga	First version
2.0	06-04-2009	Johannes Lyzenga	Final version
2.0	22-2-2010	Marcel Vlaming	Final version made public

DELIVERABLE REVIEW			
Version	**Date**	**Reviewed by**	**Conclusion***
1.0	06-04-2009	Michael Buschermöhle	Accept with modifications
1.0	06-04-2009	Marcel Vlaming	Accept with modifications

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

Table of Contents

List of Figures

List of Tables

Acknowledgement

Supported by grants from the European Union FP6, Project 004171 HEARCOM. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Pre-Amble

The objective of this deliverable is to describe the Greek digit-triplet test for hearing self screening using speech intelligibility. The first part of the deliverable deals with the construction of the Greek digit-triplet test, recording, selecting, homogenising and construction of the triplets. The second part presents validation results of the digit-triplets scoring, a test-retest evaluation, and an investigation of learning effects.

1 Executive Summary

A digit triplet is a sequence of three digits, e.g. 3-4-1, which are read out as separate digits, e.g. three-four-one. The present study describes the construction and validation of a digit-triplet test in modern Greek. The triplets were constructed from recorded triplets by cutting the original triplets up into separate digits for the first, second, and final position in the triplet. From these three lists of digits, the best tokens were selected to form three lists with one token of each included digit for each position. To acquire statistically homogenous triplet material (i.e. characterized by similar SRTs and steep psychometric functions), psychometric functions were measured for all these digits in twelve listeners (5 females and 7 males), and a set of triplets was constructed with optimised steepness of the psychometric function of each triplet. Next, the psychometric functions of the digit triplets were recorded using an adaptive procedure. During all measurements, the sound pressure level of the interfering noise was kept constant at a self chosen level. Twenty otologically normal subjects took part in the validation measurements (5 females and 15 males). The results of this validation experiment showed that for the Greek digit-triplet test, the average speech reception threshold (SRT) was −7.7 dB, and the average slope of the psychometric function (S_{50}) was 16.8 %/dB. The test-retest reliability was found to be 0.7 dB, and learning effects were found to be smaller than 1 dB in the first 5 measurements. In conclusion, the Greek digit-triplet test was found to be a reliable and accurate test for fast intelligibility screening. For the Internet version of the test, criteria for 'good,' 'insufficient,' and 'poor' test results are derived from the data.

2 Introduction

Digits have been used as the speech material for hearing screening for some time. They have been used for basic speech intelligibility measurements, for clinical purposes, and in studies aimed at the influence of speech context on intelligibility (Miller et al., 1951, Kalikow et al., 1977, Rudmin, 1987, Ramkissoon et al., 2002, Smits, 2005). An often used format of digit tests are digit triplets, such as 1-4-3 and 9-2-1 which are pronounced as: *one-four-three* and *nine-two-one*, respectively. Like standard CVC elements, words, or sentences, the triplets are usually presented to the listeners in a background of interfering noise at varying signal-to-noise ratio (SNR). On the basis of the listeners' responses, the speech-reception-threshold (SRT) can be obtained, i.e. the SNR yielding 50% speech intelligibility.

Although the limited set of material of a digit-triplet test may make it less suitable for clinical testing, using digit triplets has several advantages compared to using words or sentences. For example, such a test can be applied automatically via the Internet or telephone, which makes it a good candidate for a self screening test. In addition, digit triplets produce relatively steep psychometric functions compared to single digits or digit pairs. Since a given digit triplet is composed of only three elements, it is still relatively easy to remember all the successive digits, so memory effects do not play a large role in the test. So, SRT estimations using triplet material are characterised by relatively small standard deviations, without making any considerable demands on a listener's cognitive abilities (memory). Since complexes of different digits mostly have no context, it is extremely hard to learn particular triplets by heart, especially if they are presented in a random order. This enables prolonged testing. The SRTs obtained for digit-triplet tests are usually lower than those for sentence tests, but the results of the both tests are highly correlated (Smits, 2005). Thus, it is often possible to predict the SRT for sentence intelligibility on the basis of the SRT obtained for digit-triplet material. Furthermore, since the measurements of the SRTs can be done via telephone or Internet, digit-triplet tests are widely used for self screening of speech-reception-in-noise abilities. So far, digit-triplet tests were developed in Dutch (Smits, 2005), German (Wagener et al., 2005), English (Lutman, 2006), Polish (D1-6a), Swiss German, and French, while tests in Swedish, Australian and American English are under construction.

3 Development and generation of the triplet test

A single unit of the digit-triplet test is a complex composed of 3 digits with values from 0 to 9. In Greek the corresponding words are: "μηδέν," "ένα," "δύο," "τρία," "τέσσερα," "πέντε," "έξι," "επτά," or "εφτά," "οκτώ," or "οχτώ" and "εννέα," or "εννιά," or in Arabic characters: "miden," "ena," "dio," "tria," "tessera," "penta," "hexi," "hepta," "octo," and "ennea." Though many versions of the digit-triplet test concentrate on monosyllabic digits, it was decided that the Greek triplet test should be composed of disyllabic digits. Since in Greek, nine out of ten digits are disyllabic, only the '4' then needed to be eliminated. The used pronunciations are: "midèn," "èna," "dìeo," "trìea," "pènde," "èxie," "ephtà," "ochtò," "ennià."

A set of triplets containing all digits in all three positions were read out in a radio studio by a female speaker. The speaker was asked to keep a natural intonation, and approximately the same loudness level and vocal effort over time. Each separate digit was read out and recorded at least three times. Special care was taken to keep an approximately constant root-mean-square (rms) level during the recording sessions. The microphone signal was pre-amplified and converted to digital at a sampling rate of 44.1 kHz and with a resolution of 16 bits. The recorded triplet lists were then edited and stored on a computer hard disc as wave files (*.wav). Finally, the best digits for each position were selected and stored in separate wave files, creating three times nine signal files. In these files, silent periods were left at the start and end of each file to represent the original natural pauses as closely as possible. Finally, a speech noise was generated with the same long-term average spectrum as the concatenation of the 27 individual digits. Fig. 3.1 shows the power spectrum of this speech noise.

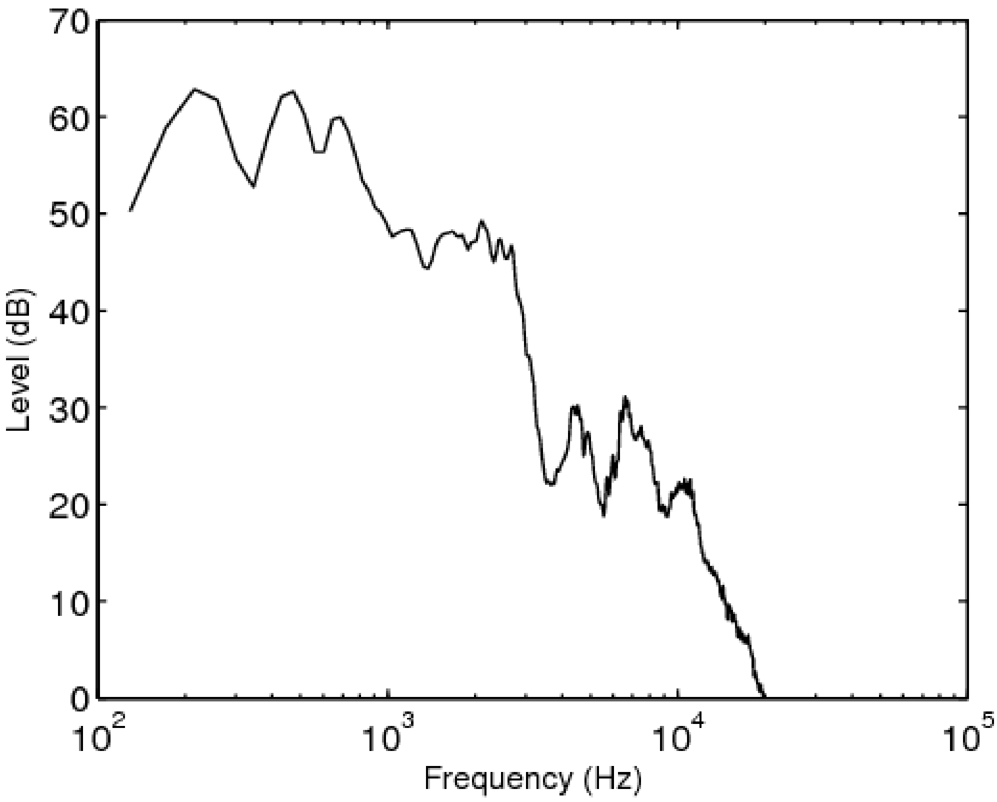

Fig. 3.1. *Power spectrum of Greek speech noise for digits.*

4 Measurements of triplet intelligibility versus SNR

4.1 Apparatus, procedure and subjects

A dedicated program was used to measure 50-percent correct thresholds for individual digits in noise. The signals were presented from a Windows XP(SP3) PC, with a Creative Audigy 2 soundcard, through Technics RP

F200 headphones. The subjects were asked to adjust the volume to a comfortable level, just as is the case in the Internet versions of the digit-triplet test. The digits were mixed digitally with the interfering noise and presented monaurally to the subjects. An adaptive paradigm was used. The noise sound-pressure level was kept constant, so the SNR value was determined entirely by the speech level. A run consisted of 32 presentations of randomly selected digits, for a specific digit position in the triplet, at varying SNRs. The start SNR was 0 dB, and it was increased by 2 dB after every incorrect response and decreased by 2 dB after every correct response. An average threshold was calculated over the last 27 presentation levels plus the presentation level of the imaginary 33^{rd} presentation. Each listener performed 15 runs per digit position per ear (3 * 15 * 32 = 1140 presentations), where 10 listeners performed a set for each ear, one listener for just the left ear, and one listener for just the right ear. So, on average, each digit was presented to a given listener at least 50 times (usually over a range of about 9 different SNRs).

The recorded SRTs were between −10.8 and −14.0 dB SNR, with an average value of −12.3 dB (SD = 1.1 dB). Over all listeners, the number of data points per digit per SNR was over 100 for all percent correct scores between 0.3 and 0.8 (the guessing chance was 0.11). The total number of data points per digit per SNR was about 250 near the threshold of 50-percent correct.

An experimental session lasted about 2 hours for one subject, including breaks which the subjects were allowed to take whenever they wished to. Twelve, self reported, otologically normal subjects took part in the experiments (5 females and 7 males). Their average age was 25.9 years with a standard deviation of 2.5.

4.2 Determination of the intelligibility function parameters: SRT and steepness (S_{50})

In the next step, the intelligibility data were determined for each digit over all subjects, i.e. the proportions of correct responses obtained for the respective SNRs. As expected, the intelligibility score depended strongly on SNR and increased monotonically with growing SNR value. To ultimately obtain psychometric functions for the triplets with optimal steepness, the average psychometric functions for the individual digits were determined after correcting the data set of each listener by their average single-digit SRT. On the basis of these intelligibility data, the psychometric functions were fitted using the least-mean-square (LMS) method. The intelligibility data were fitted using cumulative density functions (CDF functions, Versfeld et al., 2000, Smits, 2005, Ozimek et al., 2006) to obtain SRT values and slope values (S_{50}) for each digit at each position in the triplet. Such intelligibility functions are characterized by four main parameters: (1) the SRT, i.e. the SNR value yielding 50%-probability of a correct response, (2) the S_{50}, i.e. the steepness of the function at the SRT point, (3) the guess rate, which in this case was 1/9 = 0.11, and (4) the lapse rate, which we were able to chose at unity as the average scores for high SNR were all very close to 1.0. An example of a data set for the digit '2' in the second digit position is shown in Fig. 4.1, where the top panel shows the average correct scores as function of the SNR, and the bottom panel shows the number of data points over which each entry in the top panel was averaged.

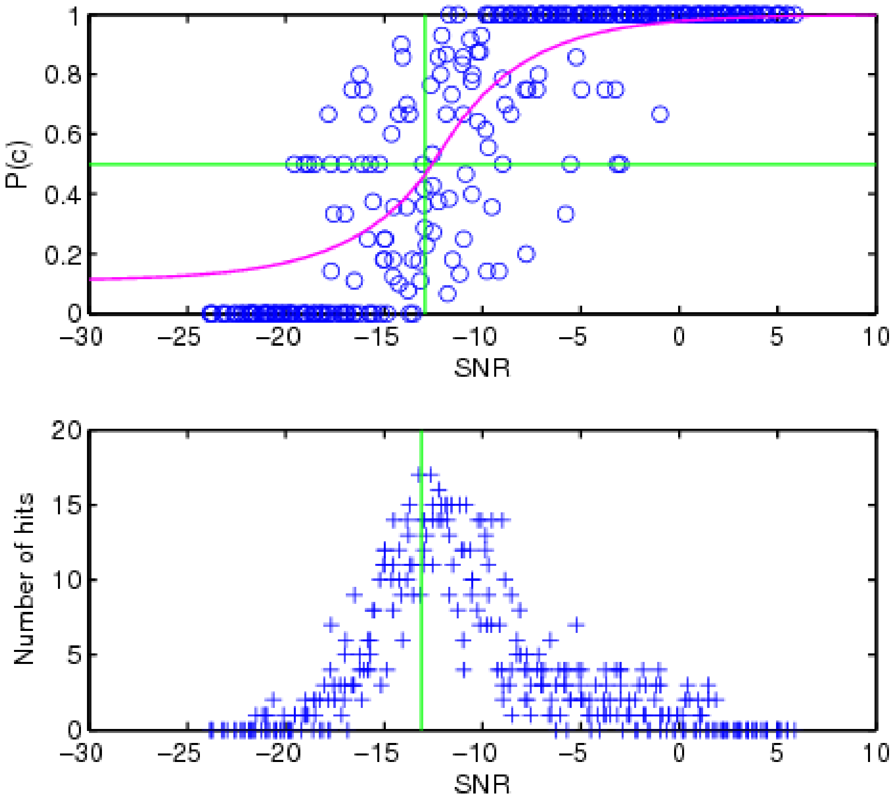

Fig. 4.1. *Psychometric functions for the first digit position.*

Figure 4.1 shows that most data was collected near the threshold, which was about −13 dB for this particular digit. There are many data points at zero and unity at the lower and higher SNR values, respectively, but the number of hits in the bottom panel shows that these entries scored only a small number of hits, and correspondingly, received a low weight in the fitting procedure.

All fitted functions for the individual digits are shown for the first, second, and third digit positions in the Figs. 4.2, 4.3, and 4.4, respectively. The psychometric functions in Figs 4.2 and 4.3 all show a very good steepness. In Fig 4.4, the psychometric function for the digit '3' shown in cyan is clearly less steep than the other functions.

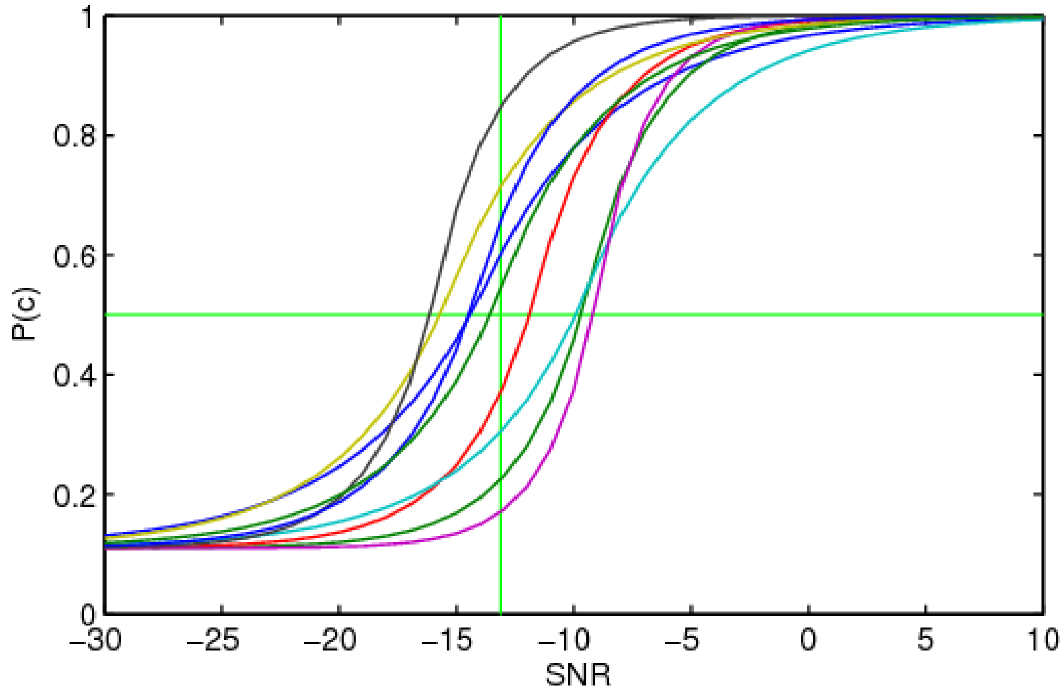

Fig. 4.2. *Psychometric functions for the first digit position. The curves for the digits '0,' '1,' '2,' '3,' '5,' '6,' '7,' '8, and '9' are displayed in blue, green, red, cyan, magenta, yellow, and black, respectively.*

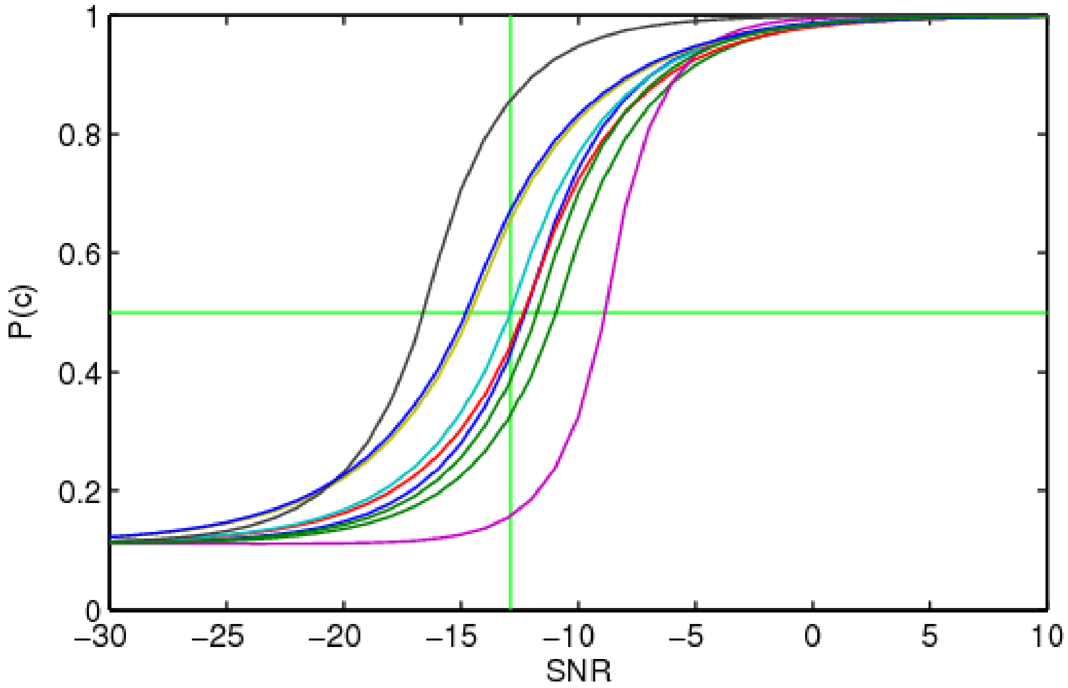

Fig. 4.3. *Psychometric functions for the second digit position.*

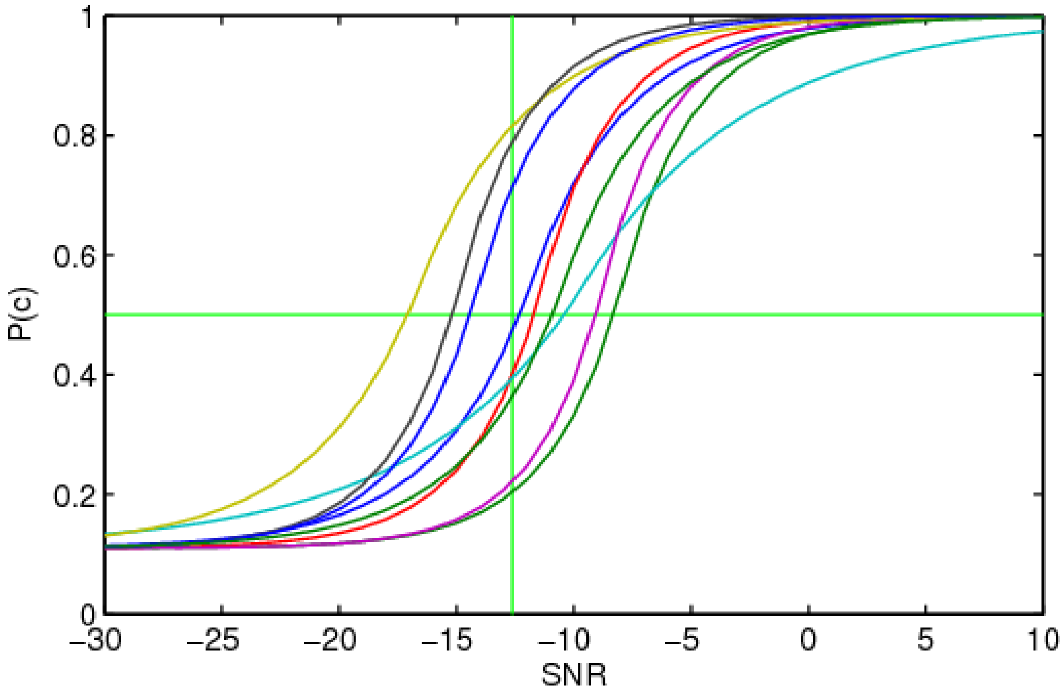

Fig. 4.4. *Psychometric functions for the third digit position. The function of the digit '3' (cyan) is shallower than the others.*

5 Composition of the final triplet test

5.1 Selection of optimised triplets

Reliable and accurate tests for speech intelligibility measurements need to be composed of statistically homogenous speech material, i.e. material that is characterized by similar and steep psychometric functions (Plomp and Mimpen, 1979, Kollmeier and Wesselkamp, 1997, Versfeld et al., 2000, Wagener, 2003, Smits, 2005, Ozimek et al., 2006). A set of 92 triplets was chosen that had a fairly even distribution of the digits over the three positions in the triplet (see Appendix A). Using the psychometric functions of the individual digits, psychometric functions of these triplets could be predicted. In order to maximise the slope, rms adjustments of −2 to +2 dB were allowed for each digit, with respect to the average level of all the digits. However, to obtain a good naturalness of the produced triplets, the levels of the relatively soft digits '6' were kept at their original levels at approximately 3 dB below the average digit level. In addition, for the digits '1,' '2,' '3,' '6,' '7,' and '9' the digit at the third triplet position needed to reduced in level by 1 dB for good naturalness. The constructed triplets met the following conditions:

- SRT value individually corrected to the predicted average SRT, as obtained from the measurements of the individual digits. This average SRT was taken at the cube root of the 50-percent correct point: an average 79.4-percent correct point ($100 * 0.5^{1/3}$) for the digits gives an average correct point of 50-percent for the triplets.

- the S_{50} was at least 11 percent/dB, or 9 percent/dB when taken as a straight line through the points at p(c) = 0.25 and p(c) = 0.75.

The average SRT for all the triplets was adjusted to –7.7 dB and the average S_{50} was 15.3 (SD = 2.8), or 11.9 (SD = 2.1) percent/dB when taken as a straight line through the points at p(c) = 0.25 and p(c) = 0.75.

Fig. 5.1 depicts a plot of the average psychometric functions for the digit triple tests for a number of European languages and the newly developed Greek digit triple test (for measurements via headphones).

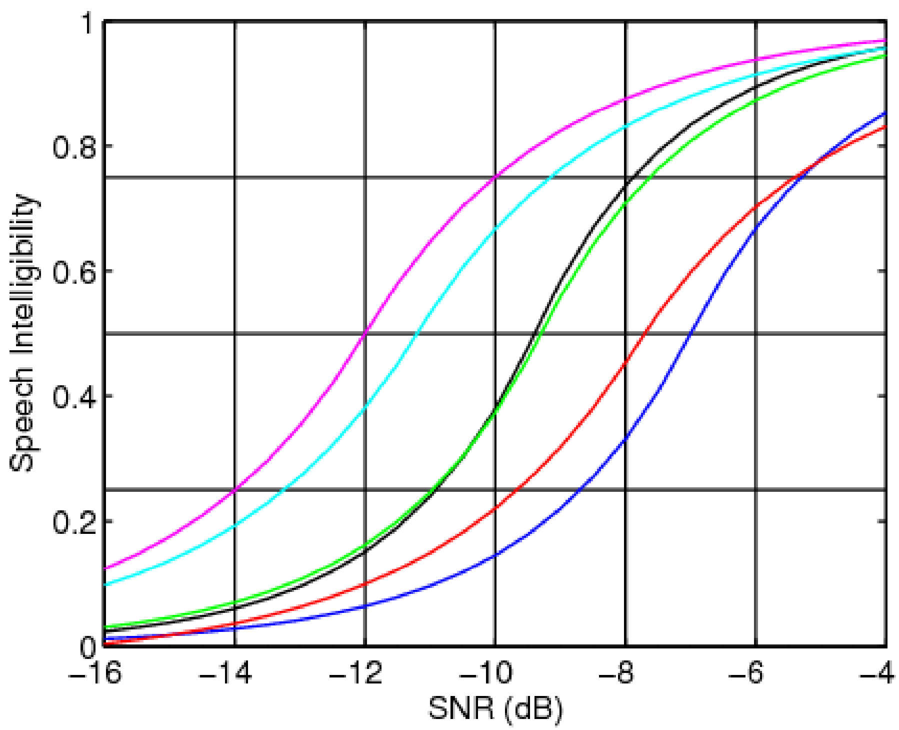

Fig. 5.1. *Comparison of average psychometric functions for digit-triplet tests across languages: English (magenta); Dutch (cyan), German (green), Polish (black), Swedish (blue), and expected Greek (red).*

As can be seen, there are some clearly visible differences in SRTs and slopes between the psychometric functions obtained for the respective languages. The functions obtained for Polish and German are very similar. The slope of the Greek expected psychometric function is close to those of the English, Dutch, and Swedish functions, albeit slightly shallower. The thresholds for the Swedish and Greek functions are somewhat higher (closer to zero) than those of the other countries.

6 Validation measurements

6.1 Apparatus, procedure and subjects

A dedicated program was used to measure 50-percent correct thresholds for triplets lists in noise. The stimuli were presented using the same setup as was used in the homogenisation experiment. As in that experiment, the subjects were asked to adjust the volume to a comfortable level, just as is the case in real Internet versions of the digit-triplet test. The triplets were mixed digitally with the interfering noise and presented monaurally to the subjects. An adaptive 1-up/1-down paradigm was used, almost identical to the procedure used by Smits et al. (2004). They used 23 presentation, where we used 24 for the determination of the psychometric functions, but only 23 for the calculation of the SRT. The noise sound-pressure level was kept constant, so the SNR value was determined by the speech level. A run consisted of 24 presentations of randomly selected triplets (from the set of 92) at varying SNRs. The start SNR was 0 dB, and it was decreased by 2 dB after every correct response and increased by 2 dB after every incorrect response. The average threshold was calculated over the last 18 presentation levels, excluding the presentation level of the imaginary 25th presentation, to make this calculation equal to that of Smits et al (2004). Each listener performed 5 runs for the right ear (making a total of 20 * 5 * 24 = 2400 presentations). So, on average, each triplet was presented about 25 times (usually over a range of about 8 different SNRs).

An experimental session lasted about half an hour for one subject. Twenty otologically normal subjects took part in the experiments (5 female and 15 male). Their average age was 30.4 years with a standard deviation of 6.1.

6.2 Determination of SRT, test-retest reliability, and steepness of the psychometric function

The average threshold of the first run (the test run) was found to be –7.72 dB with an SD of 1.12 dB. For the second run (the retest run) the average was -8.08 dB with an SD of 0.96 dB. These two runs led to a test-retest reliability (SD of the individual differences divided by 1.41) of 0.72 dB. The learning effect of the test was investigated using all five runs that the listeners performed. The average thresholds for those runs are presented in Fig. 6.1. The average threshold goes down by 0.3 dB from the first to the second run, from -7.7 to -8.1 dB. After that it keeps decreasing with about 0.1 dB per additional run to -8.5 dB for the fifth run. This means that the learning effect for five runs is smaller than 1 dB, and that for the first two or three runs it is smaller than the test-retest reliability. Nevertheless, if the test is to be used repeatedly for examining the SRT under different conditions, it is probably wise to add one practice list at the beginning of the series to minimise any learning effects.

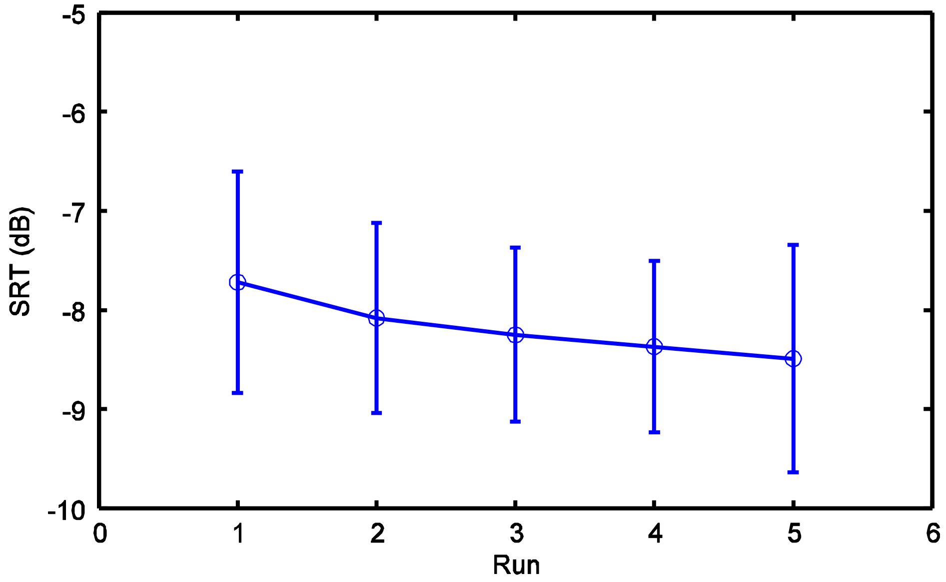

Fig. 6.1. *The average thresholds and their SDs over all five runs.*

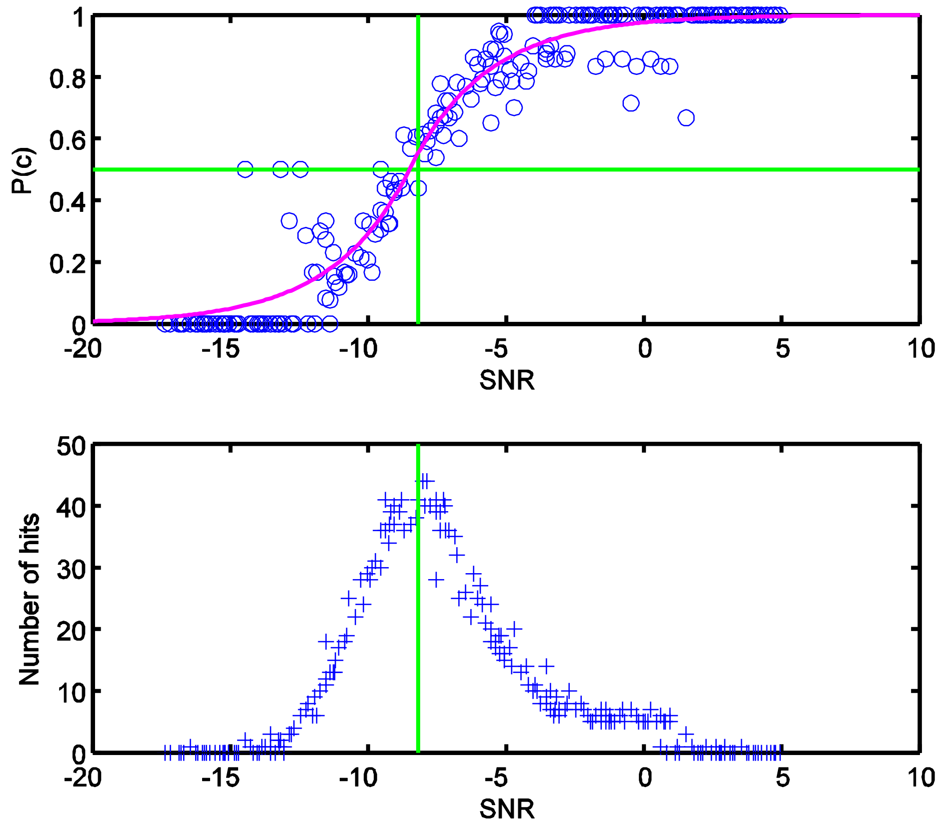

Fig. 6.2. *Average psychometric function for the triplets.*

Next, the average psychometric function of all the triplets was determined over all adaptive, five 24-stimulus, runs of the 20 subjects. The average SRT over all five runs and 20 listeners was -8.2 dB. To obtain a maximal steepness, the individual results were corrected to the average value of -8.2 dB. On the basis of these intelligibility data, the average psychometric function was fitted using a least-mean-square (LMS) method. As earlier, the intelligibility data were fitted using cumulative density functions (CDF functions) to produce an SRT and a slope value (S_{50}). In this estimation, the guess rate was set to zero, and the lapse rate was set to unity as the average scores for high SNR were all very close to 1.0. The SRT and the slope were used as parameters in the fitting procedure. The full data set for this fit is shown in Fig. 6.2, where the top panel shows the average correct scores as function of the SNR, and the bottom panel shows the number of data points over which each entry in the top panel was averaged (the SRT corrections produced data for a large set of SNRs).

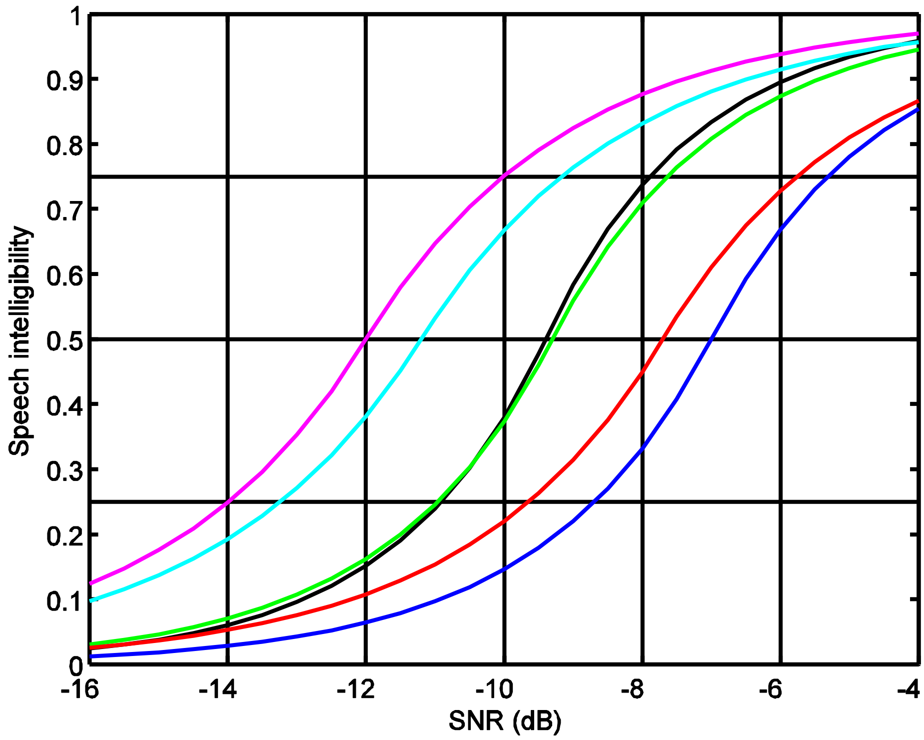

Fig. 6.3. *Comparison of average measured psychometric functions for digit-triplet tests across languages: English (magenta); Dutch (cyan), German (green), Polish (black), Swedish (blue), and Greek (red).*

The bottom panel of Fig. 6.2 shows that data was collected for a fair spread of SNR values. The top panel shows a spread of the p(c) data around the fitted curve (magenta), but most of the data points lie in the near vicinity of this fitted curve. The average SRT found for the optimally fitted curve is -8.5 dB, which lies near the arithmetic average of -8.2 dB.

The slope found for the current optimal least-squares fit is 16.8 percent/dB. Assuming that the shape of the psychometric function is the same for all runs, we can estimate the psychometric function of the first run of the tests by shifting the function from Fig 6.2 to the average SRT of the first run. In Fig. 6.3, the resulting psychometric function for the average SRT of -7.7 dB is compared to the results from the other countries, as was done earlier for the expected psychometric function in Fig 5.1. The measured curve has the same average SRT than the expected curve, but is even a bit steeper. The mean SRT of -7.7 dB for the novel Greek digit-triplet test lies at the higher end of the range of the SRTs obtained for the other languages ($SRT_{English}$ =-12 dB; SRT_{German} =-9.3 dB; SRT_{Dutch} =-11.2 dB; SRT_{Polish} =-9.4 dB; $SRT_{Swedish}$ =-7.0 dB). This means that the test is relatively robust to environmental noises present when clients perform the test, just as the Swedish test. The mean steepness of the psychometric function (S_{50}= 16.8 %/dB) lies in the middle of the range found for the other languages ($S_{50Dutch}$ = 16 %/dB; $S_{50English}$ = 16.4 %/dB; $S_{50Swedish}$ = 19.3 %/dB; $S_{50German}$ = 19.6 %/dB; and $S_{50Polish}$ = 21.6 %/dB). This means that the test has a good sensitivity.

6.3 Preparation of the Internet version

In order to be able to implement the Greek digit-triplet test on the Internet, 12 test lists were created (HörTech, Oldenburg), which are shown in Appendix B. Each list contains 23 triplets, and as a total of 92 triplets were available, each triplet occurs exactly three times in the lists (12 * 23 = 3 * 92), but never more than once per list. In order to check the performance of these lists with respect to the validation experiment, in which the triplets were chosen at random, their psychometric functions have been examined. This exercise has been performed for the psychometric functions without the correction for the individual SRTs, so the slopes will be slightly less steep than reported in the previous section, but the SRTs can be expected to be accurate. The results of this exercise are shown in Table 6.1, which shows the average SRTs and slopes, and in Fig. 6.4, where the uncorrected psychometric function of the validation experiment is shown in the top panel, and the psychometric functions of the 12 test lists are shown in the bottom panel.

Table 6.1 shows that the steepness of the uncorrected psychometric function is about 1 percent/dB shallower than that of the corrected function, while the SRT is unchanged. The same can be expected for the psychometric functions of the 12 test lists, they should have steepness values near the average of 15.7 percent/dB, and SRTs of about -8.5 dB. Table 6.1 shows that this is true for all test lists, except for lists 7 and 11, which have rather low steepness values. These two shallow slopes can be recognised as the two overlapping black lines in the bottom panel of Fig 6.4. Deleting these two lists from the Internet test will enhance its sensitivity and reliability. The remaining 10 lists all have steepness values and SRTs that are very close to target.

Psychometric functions:	Slope (%/dB)	SRT (dB SNR)
Corrected	16.8	-8.5
Uncorrected	15.7	-8.5
List 1	15.2	-8.9
List 2	19.6	-8.4
List 3	16.2	-8.5
List 4	14.2	-8.2
List 5	18.8	-8.6
List 6	18.1	-8.4
List 7	13.1	-8.3
List 8	14.7	-8.8
List 9	16.8	-8.5
List 10	17.4	-8.7
List 11	12.4	-8.4
List 12	18.1	-8.3

Tab. 6.1. *SRT and slope values for the corrected and uncorrected psychometric functions of the validation experiment, and for the uncorrected psychometric functions of the 12 test lists.*

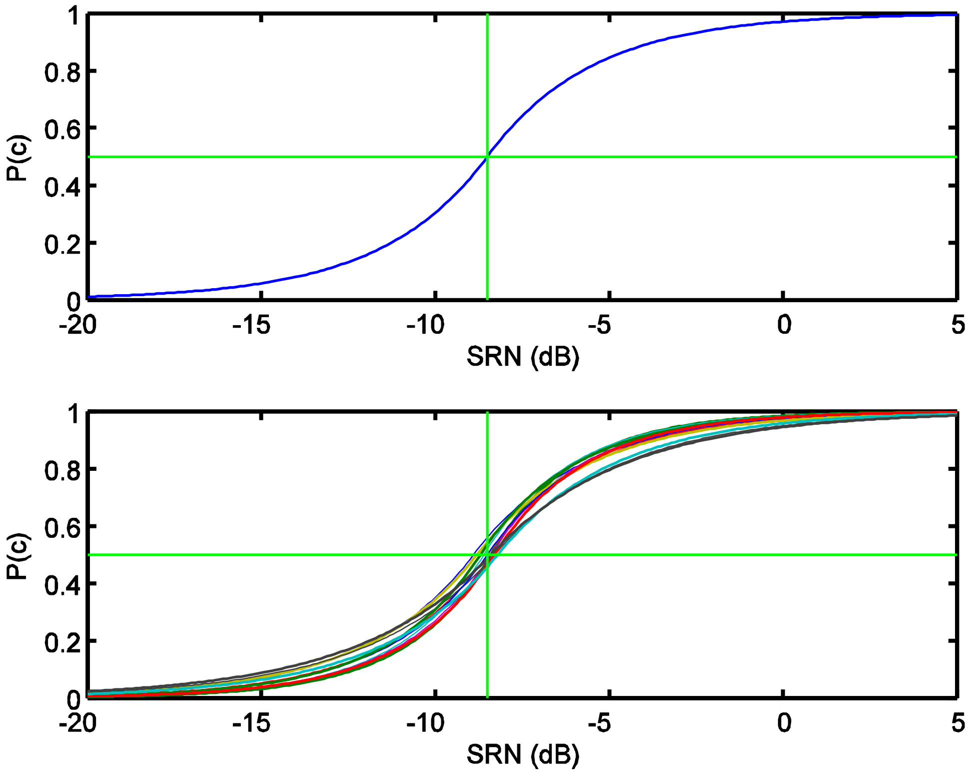

Fig. 6.4. *Comparison of average uncorrected psychometric functions for the validation experiment (top panel) and the 12 test lists (bottom panel).*

The average SRT is not affected by eliminating test lists 7 and 11. The average slope of all 12 test lists is 16.2 percent/dB, corresponding to average uncorrected and corrected slopes from the validation experiment of 15.7 and 16.8 percent/dB, respectively. After eliminating lists 7 and 11, the average slope of the 10 remaining test lists is 16.9 percent/dB. In comparison with the relation between the average slope of all test lists and those form the validation experiments, we may expect slopes of about 1 percent/dB more for the psychometric functions of the selected lists when they are corrected for the individual SRTs. This would give an estimation of the overall slope of the Internet version of the Greek digit-triplet test of approximately 18 percent/dB. This may, in future, be confirmed from the data of the Internet version of the test.

For self-screening purposes via Internet and telephone, the results of the test need to be presented as 'good,' 'insufficient,' or 'poor.' For this, the measured monaural SRT can be categorised using the criteria of -5.5 dB (SRT + 2 * SD) for 'insufficient' and -3.2 dB (SRT + 4 * SD) for 'poor' results. For binaural presentations, a 1-dB correction may be needed (-6.5 dB for 'insufficient' and -4.2 dB for 'poor' results).

7 Dissemination and Exploitation

The Greek digit-triplet test has been implemented for Internet use. A telephone implementation can be considered. The test will be offered to the Greek public as a hearing self-screening test, to be performed either at home of in the clinic of a professional.

8 Conclusions

The following conclusions can be drawn:

- The mean SRT of -7.7 dB for the Greek digit-triplet test lies at the higher end of the range of the SRTs obtained for the other languages, implying that the test is relatively robust to environmental noises.

- The mean steepness of the psychometric functions for the Greek digit triplets (S_{50}= 16.8 %/dB) lies in the middle of the range found for the other languages, implying that the test has a good sensitivity.

- The Greek digit-triplet test was shown to be a reliable and accurate test for a fast Internet hearing self screening (especially when deleting the test lists 7 and 11). For self-screening purposes a measured monaural SRT can be categorised using the criteria of -5.5 dB for 'insufficient' and -3.2 dB for 'poor' results. For binaural presentations, a 1-dB correction may be needed (-6.5 and -4.2 dB).

9 Literature

Hall, J. L. (1983). "A procedure for detecting variability in psychophysical thresholds." Journal of Acoustical Society of America 73: 663-667.

Kalikow, D. N., Stevens, K. N., and Elliot, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability." J Journal of Acoustical Society of America 61(1337): 1337-1351.

Kollmeier, B. (1990). Messmetodik, Modellierung und Verbeserung der Verstandlichkeir von Sprache. Gottingen, Georg-August-Universtat.

Kollmeier, B., and Gilkey, R. (1983). "Adaptive staircase methods-A comparison using Markov theory." Journal of Acoustical Society of America 1(73 suppl.): 104.

Kollmeier, B., Gilkey, R., and Sieben, U. (1988). "Adaptive staircase techniques in psychoacoustic: A comparison of human data and a mathematical model." Journal of Acoustical Society of America 83(5): 1852-1862.

Kollmeier, B., and Wesselkamp, M. (1997). "Development and evaluation of a sentence test for objective and subjective speech intelligibility assessment." Journal of Acoustical Society of America 102(4): 1085-1099.

Lutman, M. E. (2006). "Development of English digit triplet test for earphones and telephone." HearCom Meeting - Alphen (presentation).

Miller, G. A., Heise, G. A., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials." Journal of Experimental Psychology 41: 329-335.

Ozimek, E., Kutzner, D., Sek, A., Wicher, A., and Szczepaniak, O. (2006). "The Polish sentence test for speech intelligibility evaluations." Archives of Acoustics 31(4): 431-438.

Plomp, R., and Mimpen, A. M. (1979). "Improving the reliability of testing the speech reception threshold for sentences." Audiology 18: 43-53.

Ramkissoon, I., Proctor, A., Lansing C. R., and Bilger, R.C (2002). "Digit speech recognition thresholds (SRT) for non-native speakers of English." American Journal of Audiology 11(23-28).

Rudmin, F. (1987). "Speech perception thresholds for digits." Journal of Audiological Research 27: 15-21.

Smits, C., Kapteyn, T. S., and Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. International Journal of Audiology, 43(1), 15-28.

Smits, C.. and Houtgast, T. (2005). Results from the Dutch speech-in-noise screening test by telephone. Ear and Hearing. 26(1), 89-95.

Smits, C., and Houtgast, T. (2006). "Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests." Journal of Acoustical Society of America 120(3): 1608-1621.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (2000). "Method for the selection of sentence material for efficient measurement of the speech reception threshold." Journal of Acoustical Society of America 107: 1671-1684.

Versfeld, N. J., and Dreschler, W. A. (2002). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners." Journal of Acoustical Society of America 111(1): 401-408.

Wagener, K. (2003). Factors influencing sentence intelligibility in noise. PhD Dissertation, University of Oldenburg.

Wagener, K., Eeenboom, F., Brand, T., and Kollmeier, B. (2005). Ziffer-Tripel-Test: Spracherverstandlichkeitstest uber das Telefon. Tagungs-CD der DGA Jahrestagung.

Appendix A: the Greek triplets

Triplet no.	Triplet	SRT (dB) before correction	Slope (%/dB) 0.25 – 0.75	S_{50} (%/dB)
1	025	-6.49	11.06	14.40
2	026	-8.17	8.57	11.16
3	031	-5.58	10.60	13.81
4	038	-9.34	9.78	12.74
5	058	-5.42	15.26	19.87
6	065	-7.05	10.77	14.02
7	067	-10.09	9.59	12.50
8	071	-8.86	10.61	13.82
9	079	-9.99	9.19	11.98
10	087	-9.99	9.62	12.53
11	102	-6.94	12.54	16.33
12	109	-5.47	11.71	15.25
13	152	-5.48	16.37	21.32
14	156	-5.35	16.25	21.17
15	157	-5.39	16.55	21.56
16	160	-6.01	11.26	14.66
17	162	-7.35	12.31	16.04
18	167	-9.43	11.92	15.53
19	180	-5.98	11.27	14.67
20	186	-6.96	11.02	14.35
21	198	-7.36	12.26	15.96
22	205	-7.61	12.53	16.32
23	207	-8.98	11.89	15.49
24	219	-7.59	11.06	14.41
25	230	-8.00	10.85	14.14
26	275	-8.75	12.87	16.76
27	283	-7.33	8.48	11.05
28	293	-6.31	8.91	11.61
29	296	-7.89	11.57	15.07
30	315	-5.96	11.26	14.66
31	321	-5.04	10.51	13.69
32	350	-5.21	14.45	18.82
33	356	-5.37	14.67	19.10
34	361	-5.46	10.34	13.46
35	376	-9.81	8.72	11.36
36	379	-8.27	9.07	11.82
37	389	-6.34	8.75	11.40
38	501	-6.18	14.89	19.39
39	529	-6.69	14.08	18.33
40	532	-6.35	14.79	19.27
41	560	-6.72	14.05	18.30
42	572	-6.59	15.14	19.72
43	578	-10.57	14.99	19.52
44	581	-6.40	14.71	19.16
45	582	-6.49	14.71	19.16
46	583	-5.99	11.67	15.20
47	593	-5.96	11.84	15.42
48	596	-6.63	14.49	18.88
49	608	-8.31	10.94	14.25
50	609	-6.31	10.11	13.17
51	615	-7.14	11.98	15.60

52	632	-7.22	10.89	14.18
53	639	-7.38	9.55	12.44
54	652	-5.45	15.84	20.63
55	657	-6.57	15.73	20.49
56	672	-11.30	11.32	14.75
57	685	-7.31	10.98	14.31
58	698	-9.90	10.92	14.23
59	713	-6.80	8.63	11.24
60	718	-11.46	12.04	15.68
61	720	-10.40	10.51	13.69
62	721	-9.37	11.49	14.97
63	730	-9.99	10.71	13.95
64	732	-10.41	11.89	15.49
65	735	-10.20	12.23	15.94
66	750	-6.50	15.75	20.51
67	782	-11.73	11.84	15.42
68	792	-10.81	12.14	15.81
69	820	-9.15	9.90	12.90
70	821	-7.96	11.14	14.51
71	831	-8.34	11.23	14.62
72	837	-10.21	11.21	14.60
73	856	-5.56	15.62	20.35
74	860	-9.47	9.75	12.70
75	862	-8.40	11.06	14.41
76	867	-10.86	10.90	14.19
77	890	-8.26	10.51	13.69
78	892	-7.77	11.74	15.29
79	905	-5.85	11.75	15.30
80	912	-7.36	11.28	14.70
81	915	-6.28	11.86	15.45
82	916	-6.51	9.87	12.86
83	917	-9.11	10.78	14.04
84	920	-6.36	9.40	12.24
85	931	-6.19	10.87	14.16
86	950	-5.38	15.27	19.89
87	957	-5.48	15.64	20.37
88	960	-6.01	8.93	11.63
89	967	-10.01	10.17	13.24
90	978	-9.96	10.89	14.18
91	985	-6.21	11.02	14.35
92	986	-8.12	8.55	11.14

Appendix B: the 12 test lists

List	1	2	3	4	5	6	7	8	9	10	11	12
	293	890	581	031	916	582	685	732	720	198	867	205
	867	025	986	985	198	315	275	931	837	283	361	532
	207	632	379	067	186	915	296	156	356	572	609	109
	931	856	275	283	985	067	560	632	230	860	275	912
	960	821	950	296	735	321	960	862	186	529	931	792
	376	102	315	596	529	698	652	160	379	376	152	071
	529	560	639	079	609	230	087	730	892	632	058	581
	532	698	198	157	856	578	672	820	657	916	750	296
	058	180	219	831	596	837	957	079	102	583	672	721
	912	917	071	732	379	219	713	109	219	639	65	905
	038	109	685	792	180	205	792	831	593	160	157	718
	967	087	205	978	718	920	038	162	856	180	031	615
	609	735	615	657	905	967	283	860	862	782	293	890
	920	389	583	608	581	782	361	912	920	652	915	735
	156	750	892	572	608	986	293	031	582	167	079	713
	321	862	820	230	890	350	501	389	162	960	389	978
	501	652	720	782	750	065	356	867	732	315	207	156
	915	160	152	837	657	157	892	978	985	831	038	560
	356	718	730	916	102	917	721	532	067	321	820	957
	721	713	065	578	615	583	593	376	608	986	596	730
	593	905	162	582	025	821	720	026	967	698	026	025
	672	350	186	167	058	207	572	167	821	950	578	917
	026	361	957	860	639	950	071	152	685	350	501	087