



FP6-004171 HEARCOM Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-4-5: Conclusions on the impact of hearing impairments on the quality measures w.r.t. telecom degradations

Contractual Date of Delivery: M 42 (+45 days)

Actual Date of Submission: 26.05.2008

Editor: Ute Jekosch

Sub-Project/Work-Package: SP2/WP4

Version: 2.0

Total number of pages: 22

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

Deliverable D-4-5

VERSION DETAILS

Version: 2.0

Date: 22.05.2008 (small delay for review agreed)

Status: Final

CONTRIBUTOR(S) to DELIVERABLE					
Partner	Name				
DE-IAS	Ute Jekosch				
DE-IAS	Margitta Lachmann				
DE-IAS	Sebastian Merchel				

DOCUMENT HISTORY							
Version	Date	Responsible	Description				
1.0	25.04.2008	Ute Jekosch, DE-IAS	Draft				
2.0	19.05.2008	Ute Jekosch, DE-IAS	Draft				
2.0	22.05.2008	Ute Jekosch, DE-IAS	Final				

DELIVERABLE REVIEW							
Version	Date	Reviewed by	Conclusion*				
1.0		Tammo Houtgast	Accept after modifications				
1.0	09.05.2008	Rainer Huber	Modify				
1.0		Torben Poulsen	Accept after modifications				
2.0	21.05.2008	Rainer Huber	Accept				
2.0	22.05.2008	Torben Poulsen	Accept				
2.0	25.05.2008	Tammo Houtgast	Accept				

^{*} e.g. Accept, Develop, Modify, Rework, Update

Table of Contents

Pre	-Amble	5
Exe	ecutive Summary	7
1	Introduction	8
2	Scope of Validity of Extended Models	9
3	Future Prospects	.13
4	Summary	.18
5	References (Selected literature)	. 20
Apr	pendix 1	.22

List of Figures

Fig. 1: Quality projection model for voice transmission quality perceived by hearing-impaired listeners

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

This deliverable is related to the HEARCOM sub project 2 (SP2) *Adverse Condition in Communication Acoustics*, work package 4 (WP4) *Telecommunication Systems*.

The main objective of this work package is the investigation of speech quality in telecommunication networks as perceived by hearing-impaired listeners and its prediction by quality models. Two series of listening-only tests have been performed with normal and hearing-impaired listeners within this work package. They are reported in the deliverables D-4-2: Report on Experiments on the Performance of Normal and Non-normal Hearing Listeners for a Range of (Simulated) Transmission Conditions and D-4-3: Report on Experiments on the Performance of Normal and Nonnormal-hearing Listeners for a Range of (Simulated) Transmission Conditions with Combined Technical Disturbances. The purpose of these listening tests is two-fold: Firstly, they serve to investigate possible differences in the quality perception between normal-hearing and hearingimpaired listeners and to conclude possible consequences from these results for the demands on telecommunication networks. Secondly, they serve to build a data base that is needed to test, possibly extend and/or "train", and finally validate quality models that aim at predicting auditory quality ratings of hearing-impaired telephone users. The completion of this database is represented by milestone M-4-2: Database containing auditory test results of hearing-impaired and normal-hearing listeners.

Promising model candidates are described in deliverable **D-4-1**: Report on the Selection of Quality Models for Telecommunications. Three models have been pre-selected and applied to the data gathered so far (i.e. data from the first set of listening test). This represents the milestone M-4-3: Instrumental measuring results. As a consequence of these first model results, two of the pre-selected models have been considered further. Using the data having been collected in the first listening tests and a part of those ones having been obtained from the second series of listening tests, these models have been extended and enhanced in order to improve the correlation between model ratings and auditory quality judgments. The results of the extended models on these "training" data are represented in milestone M-4-4: Preliminary results on model extensions. In D-4-4: Report on Model Performance for Different User Groups the model extensions are described. They are based on the results from the listening tests, taking into account different languages, hearing impairment levels and test conditions.

This deliverable **D-4-5**: Conclusions on the Impact of Hearing Impairments on the Quality Measures w.r.t. Telecom Degradations summarises our achievements, with specific emphasis on the question of validity, reliability and objectivity of both types of measurements we performed, the auditory and instrumental ones. After a brief review of the

approach taken, the question is discussed whether an initiative for a new ITU-T recommendation on an extended E-model and/or PESQ algorithm for hearing-impaired users can already be launched, specifically taking into consideration issues of model extensions performed in SP2/WP4.

Executive Summary

In SP2/WP4 a number of auditory and instrumental data were collected, all related to hearing-impaired listeners. The aim of this work package was to first collect data that are typical for hearing-impaired listeners using traditional, narrow-band and handset terminated networks, and then to check whether recommended network planning and speech quality models can be used to predict the specific auditory behaviour of this group of listeners. The intention was to qualify and quantify individual hearing losses with regard to telecom degradations and to enhance tools accordingly which planners, operators, and developers regularly use for the design of new networks, the (possibly online) evaluation of existing networks, and the development of new codecs, respectively. The aim was to provide auditory and instrumental data in order to also account for the needs of hearing-impaired listeners in network planning activities. Another aim of extending existing network-planning/speech- quality models towards hearing impairment listeners was to recommend them to the ITU-T for standardisation. Model extensions lead to promising results, nevertheless, the data do not yet suffice as a basis for standardisation. D-4-5 gives some of the reasons why. Main aspects are discussed that have to be investigated further much more analytically than could be done in the framework of this HEARCOM work package.

1 Introduction

One of the main tasks in HearCom WP4 is the extension of existing models for estimating the quality in telecommunication networks towards hearing-impaired users. As described in D-4-4, the models were successfully extended with regard to the collected auditory reference dataset.

The procedure taken to develop and/or improve an instrumental measure for modelling perceived speech quality is a fairly typical one: subjects are asked to listen to speech samples and then judge and describe the perceived speech quality. The data related to the listeners are then used to test the direct performance of the instrumental method without any modification made. Then modifications of the instrumental method(s) are done whenever necessary. New auditory quality test data are used for verification, and if, after a range of system performance tests and system optimisations, the listeners' judgments agree with the instrumentally measured data with regard to perceived quality, the instrumental method in question is considered to be promising.

At the present point in time, results of this loop of performance testing and model modifications can be summarised as follows:

PESQ for hearing-impaired listeners (PESQ-HI):

- Provided that listeners are allowed to set the loudness of the speech to a comfortable level individually, the basic PESQ algorithm is able to predict quite well how hearing-impaired listeners suffering from a mild to moderate hearing loss perceive the quality of speech degraded by transmission networks. This holds for both the unaided and aided situations (ITU-T COM 12-C79, D-4-4)
- When subjects are not allowed to control loudness individually, and when they suffer from a moderate to severe loss, the basic PESQ algorithm showed low correlations in a first run (below 0.5). This shortcoming was overcome by introducing the overall level of the stimuli as a parameter, i.e. a separate level-related quality indicator for the enhanced PESQ-HI (PESQ for the Hearing Impaired).
- PESQ-HI leads to reasonable good correlations between auditory and instrumental measurements. However, the database was too restricted to verify PESQ-HI extensively with regard to other impairment factors.

Conclusion 1: It is recommended to start a further verification of PESQ-HI possibly leading to an extension of the ITU-T recommendation P.862, specifically for the use in situations where hearing-impaired subjects are the main users of a telecommunication service.

E-Model for hearing-impaired listeners (E-Model-HI):

• Level (SLR, RLR, etc.) is a critical parameter for the basic E-model. A first verification of the E-model revealed that a model extension was necessary with regard to level. The extension was done with reference to the auditory data base of the first listening test. The model improvement becomes obvious when the model prediction values are compared to data of the second listening test. This listening test is so far not used for model verification purposes.

- Taking level into account, the correlation value for German listeners with severe hearing loss and hearing aid increases from 0.3 to 0.84.
- For Flemish listeners with moderate hearing loss and hearing aid, the correlation value increases from 0.52 to 0.71. The Flemish data base was used again for model enhancements.

Conclusion 2: The E-Model-HI is a valuable tool for predicting the influence of the presentation level on the overall quality perceived by listeners with hearing impairments. The extended model achieves reasonable to good prediction accuracy. (ITU-T COM 12-C101, D-4-4)

Conclusion 3: The development of the extension is based on mean values for different hearing impairment groups. The individual data for hearing-impaired listeners shows some variation. Therefore, the extension should be used very carefully.

Conclusion 4: A verification of the E-model is a complex task. The E-Model includes many different parameters. They interact with each other in a complex way. Therefore, a thorough verification requires a listening test database, which includes further combined variations of main parameters.

Conclusion 5: The proposed model extension must be handled with care. A thorough model verification with further experimental data is necessary.

2 Scope of Validity of Extended Models

The scope of validity is dependent on the data set that the model performance is tested against. Single as well as multiple technical disturbances have been investigated. They are briefly summarised below:

Single technical disturbances:

In the first series of experiments, the perceived quality of single technical disturbances on a telephone connection was evaluated for normal-hearing and hearing-impaired listeners as follows:

- Type of test: listening-only test
- 34 processing conditions were applied to the speech signals with the following parameters at each test site
 - o level
 - o frequency shaping
 - o additive noise
 - o different codecs
 - time variant transmission impairments (packet losses and different bit error rates)
- most comfortable level
- sentence stimuli
- two male and two female speakers
- basic transmission highest quality for a narrow-band connection (Nc = -70dBm0p, Nfor =-64 dBmp, G.711 a-law, Ppl = 0, SLR = 8, RLR = 2, Dr = 3, Ds = 3, Pr = 35 dBA, Ps = 35 dBA).
- subject groups:
 - o normal hearing, Belgian (Flemish language, 9 females (f), 6 males (m), mean age 22.2 years (y))
 - o mild hearing loss, Belgian (Flemish language, 6 f, 10 m, mean age 58.1 y)
 - moderate hearing loss, Belgian, unaided (Flemish language, 4 f, 13 m, mean age 70.4 y)
 - o moderate hearing loss, Belgian, aided (same as unaided)
 - o normal hearing, German (German language, 7 f, 14 m, mean age 21.9 y)
 - o mild hearing loss, German (German language, 9 f, 6 m, mean age 59.4 y)
 - moderate hearing loss, German, unaided (German language, 8 f, 7 m, mean age 58.3 y)
 - o moderate hearing loss, German, aided (same as unaided)
 - o normal hearing, Dutch (Dutch language, 14 f, 1 m, mean age 24 y)

Results:

- Hearing impairment has an effect on auditory quality ratings
- Hearing-impaired, aided participants never rated the ideal ISDN connection as good as normal-hearing participants did
- Too low signal presentation levels represent the worst single technical disturbance when comparing normal and hearing-impaired participants
- Noise, whatever kind, is perceived at higher levels only by the hearing-impaired compared to normal-hearing persons

Multiple technical disturbances

In the second series of experiments, the effect of combined technical disturbances on the perceived overall quality at the receiver side was examined as follows:

- type of test: listening-only test
- 34 processing conditions applied to the speech signals with the following parameters at each test site (German and Belgian):
 - o level
 - o frequency shaping
 - o additive noise
 - o different codecs
 - o time variant transmission impairments (packet losses and different bit error rates)
 - o German test site:
 - combination of different types of noise (Ps, Nc and noise cancellation algorithms)
 - non-standard presentation levels (21 conditions comprised relative deviations of -15 to +10 dB from the standard presentation level)
 - o Belgian site:
 - speech signal enhancement approaches
 - different codecs and cascades of codecs
 - non-standard presentation levels. (16 conditions with relative level deviations from -15 to +10 dB)
- subject groups:
 - o normal hearing, Belgian (Flemish language, 12 f, 3 m, mean age 22.3 y)
 - moderate hearing loss, Belgian (Flemish language, 4 f, 11 m, mean age 75.3 y)
 - severe hearing loss, Belgian, (Flemish language, 7 f, 7 m, mean age 68.5 y, all aided)
 - o normal hearing, German (German language, 8 f, 9 m, mean age 24.8 y)
 - moderate hearing loss, German (German language, 11 f, 6 m, mean age 64.3 y)
 - severe hearing loss, German, (German language, 6 f, 9 m, mean age 58.3 y, 10 participants were aided, 5 participants unaided)

Results:

- Normal-hearing listeners give highest quality ratings around the standard presentation level
- Hearing-impaired participants prefer higher levels (even higher than the most comfortable level)
- Moderately hearing-impaired listeners without hearing aids give higher quality ratings than normal-hearing listeners in conditions with noise distortions
- Severely hearing-impaired listeners with hearing aids give lowest quality ratings

The results show that the SP2/WP4 work has been quite successful. Nevertheless, there are some open questions for future work that will be discussed in the following chapter.

3 Future Prospects

An ideal outcome of SP2/WP4 would have been an enhanced PESQ-HI and an enhanced E-model-HI, both ready for standardisation within the ITU-T. As discussed above, all objectives of WP4 were successfully met, however, the aim of a standardisation was too ambitious. The reasons are as follows:

Both auditory and instrumental measurements are performed in order to collect information on the quantitative relation between a measurand and a unit. The measurand is a feature of the measured object to be quantitatively described in the course of the measurement. Although the measured object is the same for auditory and instrumental measurements (e.g., a speech event), their measurands differ: Auditory measurements are directed towards perceptually motivated measurands whereas instrumental measurements are directed towards physical ones. Even when both types of measurements comply with objectivity, reliability and validity principles, sometimes significant discrepancies occur in the measured values (e.g. values for voice quality). One of the reasons is that not all constitutive measurands are defined, and that the relation of their contributions to an integrated construct (e.g., voice quality) is not known. This is specifically true for hearing-impaired listeners using telecom equipment.

After standardisation, instrumental voice and/or speech quality measurement procedures such as PESQ-HI or the E-Model-HI will be used to calculate or predict voice quality of networks on the basis of physically measured values alone. Hearing-impaired listeners are no longer involved in this context. PESQ-HI and E-Model-HI deliver a calculation or a prediction index for voice quality, and amongst others, it is this index network planners base their decision on. The index is taken as an indicator for the quality of the speech transmission system used by hearing-impaired listeners. This shows that the measurement procedure is a central tool for decision making, and the decisions are far-reaching.

However, by using the instrumental measurement tools it is ultimately neither completely clear which perceptually relevant aspect this value illustrates (it is an overall voice quality value), nor under which circumstances this index has to function. Provided that the database is big enough, the first point might be disregarded. The second, however, is of major importance and cannot be neglected here. Its relevance becomes clear when discussing the scope, the components, the structures and the correlations of measurements for voice transmission quality in view of metrological theory.

Each measurement that is carried out is influenced by certain factors, and, as already said, that holds for auditory as well as for instrumental measurements. Widely known aspects that generally have an influence on voice and/or speech quality measurements are, e.g.:

- · stimulus speech material
 - o perceptibility of speech
 - o articulation
 - o speaking style (e.g., speed)
 - o pronunciation
 - linguistic aspects
 - o communicative intention of the speaker
 - intelligibility
 - 0 ...
- network and transmission characteristics
 - signal distortions
 - o circuit noise
 - o non-optimum loudness
 - o ambient noise
 - o delay, echo, sidetone
 - o time-variant impairments
 - send loudness rating/received loudness rating
 - o handset direct-diffuse sensitivity factor, send side/receive side
 - room noise at sender/receiver
 - o electric circuit noise
 - o noise floor, receive side
 - sidetone loss
 - o echo loudness loss
 - o ..
- subject-related characteristics
 - o individual differences in the perception of degraded speech
 - o the internal reference system when assessing speech stimuli
 - training effects
 - contextual effects
 - o cognition
 - o attention and its control
 - о ..

And for hearing-impaired listeners

- o hearing impairment, audiological profile
- o hearing aid

Usually, when a quality assessment of a telecom network is carried out (e.g., within ITU-T activities), people with normal hearing are asked to give their assessment. Thus, the measurand is a specific characteristic of the transmitted speech. Apart from this measurand, all other influencing characteristics are mostly kept constant.

In our case hearing-impaired listeners are asked to assess the network voice quality. However, when looking at Fig. 1 and Appendix 1, it can be seen that there are a number of additional accompanying situational features that may have an influence on the true value of the measurand. Within SP2/WP4 we did not have to examine all the situational features and their relationships per se, but we had to detect variables and constants when running the tests for the entire scope. We succeeded in identifying a number of these, but not all of them could be investigated. Following this HEARCOM approach further would increase the reliability of our data base, motivated by an underlying metrological system.

For sake of time and money, a selective approach had to be taken. Consequently, some questions had to remain unanswered. Experiments reported in D-4-2 and D-4-3 showed that there are some unforeseen results that could not easily be explained. Before standardising PESQ-HI and the E-model-HI, an auditory data base is necessary that includes – dependent on auditory profiles – answers to the following questions:

- How do hearing-impaired listeners process and assess features of voice and speech quality?
- Which aspects do they concentrate on?
- Are they able to compensate for them?
- Do hearing-impaired listeners perform differently than normal listeners do?
- Are there age-dependent quality reactions to perception?
- Which reaction types can be identified?
- How do these correlate with auditory profiles?
- How far-reaching is their ability to judge (generalisation)?
- Are their judgments prototypical (representativity)?

Or, formulated in terms of metrology:

- How certain and valid is their assessment (reliability and validity)?
- What is their certainty based on when they pass judgment (reasoning)?
- How representative is their assessment (general applicability)?
- How good is their ability to judge (range and limitations)?

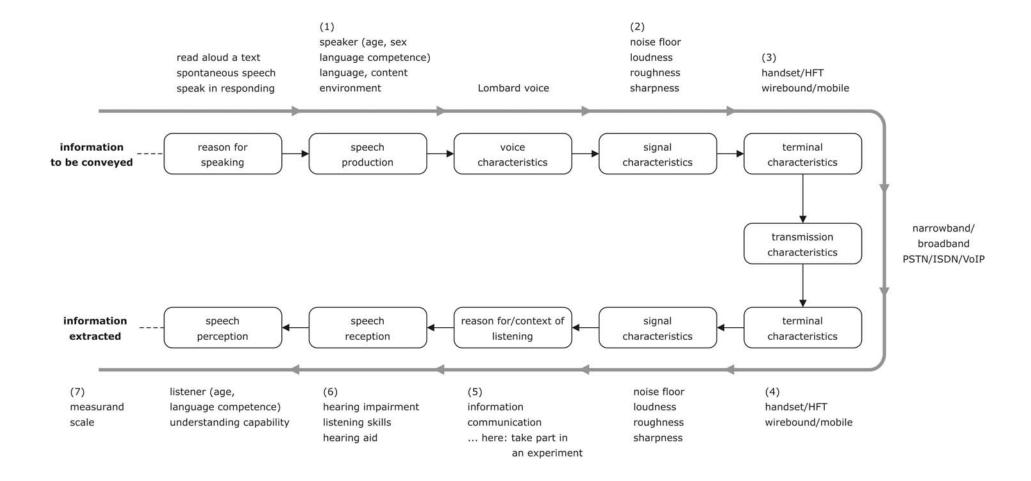


Fig. 1: Quality projection model for voice transmission quality perceived by hearing-impaired listeners

These questions have to be answered more deeply before the data set can be used for a thorough verification of the enhanced instrumental models for hearing-impaired listners. Such a thorough verification is the basic requirement for standardisation. Only then it can be guaranteed that the instrumental quality models reliably calculate or predict the auditory event of a hearing-impaired listener. So far we assume that they do, but we have no extensive experimental data which prove our assumption.

Nevertheless, our experimental data clearly showed the feasibility of adapting existing speech-quality models to hearing-impaired persons. On the basis of collected data, a number of algorithmic modifications were done for PESQ and the E-Model.

Because the Auditory Profile, for classifying different types and degrees of hearing impairment, was not yet fully defined when the experiments in WP4 were carried out, the potential of these measures in refining the speech quality prediction for a given person could not be addressed.

With two white papers (one for PESQ and one for the E-model) the issue has now been brought to the attention of the ITU-T. It is obvious that much more experimental verification is required before reaching a standard for including the characteristics of the hearing-impaired in a general quality index for perceived speech quality. This should include a broader range of transmission characteristics, and a broader range of hearing impaired persons, preferably characterised by the Auditory Profile.

The processes involved in normalisation are expensive and time consuming. As an example, the verification of the E-model has been a work item within the ITU-T for more than 5 years now. Many international partners use to perform different auditory tests, with different languages and cultures involved. This is done in a concerted way. The aim is to check E-model performance for all physical transmission parameters and parameter values against this auditory data base (but for normal-hearing users). After some periods of enhancement and verification, the E-model was given kind of a standard as described in Rec. G.107.

Within HEARCOM SP2/WP4 a new factor of influence is introduced, namely hearing-impaired listeners. This is another new dimension the influence of which has to be measured thoroughly in the advent of the standardisation process.

4 Summary

Within WP4 we are facing the fact that neither the hearing-impaired listener nor the hearing impairment does exist (in contrast to the normalhearing listener which is a construct already). However, some entries of their distinctive auditory profiles are important influential factors of the assessment event. A standardisation of PESQ-HI and E-model-HI within the ITU-T at this point in time is not advisable, because it would imply that the models are able to entirely register or predict the perceptive and assessment behaviour of the hearing-impaired person (or a sub-type of it). This may be true for a simulation of a listener's assessment behaviour in situations compatible with those ones in WP4 experiments 1 and 2, but cannot be generalised as it does not necessarily apply to a calculation or prognosis on how a hearing-impaired listener will assess transmitted speech sound events in different contexts. So far, a hearing-impaired listener's behaviour in telecom listening situations is not completely comprehensible, and consequently entirely transparent and not predictable.

Within SP1/WP2 an Auditory Profile is available now. This profile allows to characterise and classify hearing impairments in terms of perceptual categories. Hearing-impaired subjects who take part in experiments can be characterised with regard to perceptual parameters, and test data can be analysed and interpreted accordingly.

Undoubtedly, HEARCOM has set the first steps with regard to ITU-T standardisation efforts: the feasibility of adapting PESQ and the E-model to the hearing-impaired has been shown, and the Auditory Profile has been defined as a potentially relevant tool in characterising hearing-impaired listeners.

Conclusion:

Modelling auditory behaviour of hearing-impaired subjects in connection to the technological processes in transmission system planning requires an extension of auditory methodologies and methods as well as of instrumental models to capture dominant aspects of voice and/or speech quality. Investigating further the perceptual behaviour of hearing-impaired individuals when using telecom devices is absolutely necessary, because more analytic information is urgently needed for hearing impairment compensation methods. One essential tool, the Auditory Profile, has been developed within HEARCOM SP1/WP2 and is ready for use.

Work in SP2/WP4 has impressively shown that the perceived quality by hearing-impaired subjects can be modelled. However, more effort and experimental data is needed for an ITU-T standardisation of PESQ-HI and E-Model-HI.

5 References (Selected literature)

Beerends, J. G., Hekstra, A. P., Rix, A. W., and Hollier, M. P. (2002). *PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II - Perceptual model.* J. Audio Eng. Soc., vol. 50, pp. 765-778.

Beerends, J. G., Krebber J. and Huber, R. (2007), *Speech quality measurement for the hearing impaired on the basis of PESQ*, ITU-T Study Group 12, White contribution COM 12-C79.

Beerends, J. G., Krebber J., Huber, R., Eneman, K., and Luts H. (2008), "Speech quality measurement for the hearing impaired on the basis of PESQ," to be presented at the 124th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, Pre-print.

Chan, D., Fourcin, A, Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., Zeiliger, J. (1995). *EUROM- A Spoken Language Resource for the EU*, in Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology. Madrid, Spain, 18-21 September, 1995. Vol 1, pp. 867-870.

D-4-1 (2005): Report on the selection of quality prediction models for telecommunications. FP6-004171 HEARCOM SP2/WP4, deliverable

D-4-2 (2005): Report on experiments on the performance of normal and non-normal listeners for a range of (simulated) transmission conditions. FP6-004171 HEARCOM SP2/WP4, deliverable

D-4-3 (2007): Report on Experiments on the Performance of Normal and Non-normal-hearing Listeners for a Range of (Simulated) Transmission Conditions with Combined Technical Disturbances. FP6-004171 HEARCOM SP2/WP4, deliverable

D-4-4 (2007): Report on model performance for different user groups. FP6-004171 HEARCOM SP2/WP4, deliverable

Hansen, M. and Kollmeier, B. (2000). *Objective modelling of speech quality with a psychacoustically validated auditory model.* J. Audio Eng. Soc., vol. 48, p. 395-409.

Huber, R. and Kollmeier, B. (2006). *PEMO-Q - A new Method for Objective Audio Quality Assessment using a Model of Auditory Perception*. IEEE Transactions on Audio, Speech and Language processing 14(6): 1902-1911.

- ITU-T COM 12-C79: Speech quality measurement for the hearing impaired on the basis of PESQ. (TNO Information and Communication Technology, The Netherlands)
- ITU-T COM 12-C101: Tests on overall quality as perceived by high frequency hearing impaired subscribers. (Federal Republic of Germany)
- ITU-T Rec. P.48 (1989). *Specification for an Intermediate Reference System*. International Telecommunication Union, CH-Geneva.
- ITU-T Rec. O.41.(1993). Psophometer for Use on Telephone-Type Circuits. International Telecommunication Union, CH-Geneva.
- ITU-T Rec. P.880 (1996). *Continuous evaluation of time-varying speech quality.* International Telecommunication Union, CH-Geneva.
- ITU-T Rec. P.861 (1996). *Objective Quality Measurement of Telephoneband (300-3400 Hz) Speech Codecs*. International Telecommunication Union, CH-Geneva.
- ITU-T Study Group 12 (2000). *Performance of the Integrated KPN/BT Objective Speech Quality Assessment Model.* Delayed Contribution D.136.
- ITU-T Rec. P.862 (2001). Perceptual Evaluation Of Speech Quality (PESQ), An Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs. International Telecommunication Union, CH-Geneva.
- ITU-T Rec. G.107 (2002). *The E-Model, a Computational Model for Use in Transmission Planning.* International Telecommunication Union, CH-Geneva.
- ITU-T Rec. P.862.1 (2003). *Mapping function for transforming P.862 raw result scores to MOS-LQO.* International Telecommunication Union, CH-Geneva.
- ITU-T Rec. P.862.2 (2005). Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs. International Telecommunication Union, CH-Geneva.
- Jekosch, U. (2005). *Voice and speech quality perception. Assessment and evaluation.* Springer, Berlin.
- Rix, A. W., Hollier, M. P., Hekstra, A. P., and Beerends, J. G. (2002). *PESQ, the new ITU standard for objective measurement of perceived speech quality, Part 1 Time alignment*. J. Audio Eng. Soc., vol. 50, pp. 755-764.
- Raake, A. (2006) *Speech Quality of VOIP Assessment and Prediction.* John Wiley, West Sussex, England

Appendix 1

Items of the quality projection model for voice transmission quality perceived by hearing-impaired listeners

(1) Speech production:

- Which languages should be involved?
- Which stimulus material should be taken? Words, sentences, semantic-bearing or non-semantic bearing?
- How many stimuli?
- Is the complexity of the material comparable across languages (as to speech perception requirements)?
- How many speakers?
- Where do they speak? In a 'natural' environment? In a sound-proof chamber?

(2) Signal characteristics:

- Should the speech material be recorded in acoustically adverse conditions? If so: How are the conditions determined and specified?
- Is a talking head available or how should the recording be specified?

(3) Terminal characteristics (send side):

- When using handset: Should it be held in a defined way or in the way the speaker usually helds it?
- When using a hands-free-terminal: Where should the speaker's position be?

(4) Terminal characteristics (receive side):

- When using the handset: Should the receiver be held in a defined way or in the way the listener usually holds it?
- When using a hands-free-terminal: Where should the listener's position be?

(5) Reason for/context of listening:

- How many subjects?
- Which experiments are performed (e.g., listening-only-test (LOT), conversational)?
- Which metrics is used?
- How long does the experiment take place?
- Are specific experimental setups necessary for children and/or elderly people?
- How often does a subject have to come to the experimental site?
- How many stimuli are they able to scale?

- Are they allowed to listen to the offered stimuli again if they would like to?
- Is there a fixed or a flexible inter-stimulus interval?

(6) Speech perception:

- How is 'hearing impairment' classified?
- What are the auditory criteria to classify and select subjects?
- How are hearing aids classified?
- Are hearing-impaired listeners used to wear specific hearing aids?
- Should test subjects wear their hearing aid at all?
- Are they allowed to change the setting during the experiment?
- Should they wear other devices (fitted to their hearing loss) that they are not used to wear?
- Would they be able to adapt to new devices?
- If so: how long would it take?

(7) Measurand/scale:

- Which kind of information (e.g., degree of comprehensibility, intelligibility, clarity, overall quality)?
- How is extracted information scaled:
 - by freely noting down? (guided or unguided identification)
 - by responding verbally or non-verbally? (behavioural communicative reaction)