INFLUENCE OF SPEAKER, EXPERIMENTAL CONDITION AND RATING OF SPEECH QUALITY

C. A. Sust[1], H. Lazarus[2], L. Volberg[1] & M. Kulka[2]
[1]ABoVe GmbH, Gießen, [2]Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund

Speech intelligibility and speech quality The fact that high noise levels can lead to hearing damage and other forms of health impairment is general knowledge. Day-to-day experience also includes the interference of noise with spoken communication, be it at the workplace or in leisure activities – due to environmental noises, specifically traffic and, in the immediate surroundings, machines and conversations of others. Interference-free speech communication is a major prerequisite of troublefree working and humanely designed work activity. In dwellings, in the leisure domain and at school a high quality of speech communication is taken for granted, otherwise it would not be possible to learn language, to exchange (personal) information or to conduct a relaxed conversation.

With the standards to be set in the new version of ISO 9921 it is intended to ensure the minimum degree of necessary speech intelligibility in different communication situations. Speech intelligibility – in other words the percentage of correctly recognised speech items with defined speech material – is a necessary, but not a sufficient criterion for describing the quality of speech. This criterion is inadequate, in particular in the case of high speech intelligibility, because it is no longer sufficiently differentiated. Because of the so-called ceiling effects, it would appear meaningful to describe the quality of spoken communication in addition in terms of further features. A number of perspectives should be considered here:

- hearer-side: hearer satisfaction and effort
- transmission-side: interference form background noises, communication of other persons
- speaker-side: volume and natural character of the language, dialect, accent (Cox et al. 1987).

To rate communication situations quickly and simply with respect to their speech quality, it is desirable to have a scale to rate the quality of the speech communication, but one which imposes rating criteria going beyond that of speech intelligibility (Purdy & Pavlovic 1992).

To date different classification procedures have been used:

The basis is a defined vocal effort (sound pressure level at a distance of 1 metre from the speaker's mouth) with defined speech material (meaningless syllables, monosyllables, counting words, sentences etc.).

- Subjective procedures for speech quality: the quality of the communication situation is determined by human recipients classifying on rating scales.
- Subjective procedures for speech intelligibility: This is determined correctly as a percentage of directly understood and reproduced speech items of defined speech material.
- Objective procedures: by determining the physical parameters of the communication situation, for example by predicting speech intelligibility on the basis of the signal-to-noise ratio (AI, SIL, STI, RASTI etc.), and using a known relationship between this and the speech intelligibility of defined speech material (see ISO 4870, SII etc.).

In order to agree a rating scale for the quality of speech communication in relation to the signal-to-noise ratios, experiences and tests were evaluated and summarized. During the formulation of the standard governing the "Ergonomic Assessment of Speech Communication" such a quality scale was fixed in the 80s from the expert judgements known at that time and from experiments conducted (literature). In addition a relationship with the signal-to-noise ratio was established (SNR_A, SIL, AI) (ISO 9921-1, 1996).
An extensive, international study has since been conducted with respect to 16 different communication situations. It was implemented and evaluated in individual experiments in 11 different countries using appropriate speech tests (Houtgast and Steeneken 1984). Unfortunately the classification of speech intelligibility on quality scales was only included retrospectively and it is thus not adequately documented (for a critical assessment see Lazarus et al. 2002). This study served initially mainly to evaluate the RASTI index. This means that speech intelligibility tests were conducted independently in 11 countries as reference values for the RASTI measurements. In addition to the speech intelligibility test 9 out of 11 laboratories each attached an assessment on a five-stage scale from 'excellent' to 'bad'. There is no exact description of how these assessments of the speech intelligibility values arose. In addition 4 of the 11 laboratories conducted subjective evaluations on rating scales for speech quality and/or hearing effort. The study results were used, among other things, to undertake a reallocation of RASTI values on a five-stage scale from 'excellent' to 'bad'. This allocation included, alongside the basic values taken from an earlier study, the results of the subjective evaluations conducted by the four laboratories. Furthermore the evaluations of the speech intelligibility values attached by 9 laboratories were included in the calculation.
Owing to the inconsistencies in this study, such as the restricted range of the RASTI used (0.31...0.79), different speech materials, the number of speakers and listeners, evaluation procedures, the results are only suitable to a limited extent as standard values.

Experimental studies of the evaluation of speech quality One of the declared goals of the project was to expand the database with respect to speaker-relevant aspects and a qualitative evaluation of the communication situations. The communication situations were broken down. In other words, first the speech material was spoken on tape by different, non-professional speakers. These recordings were played to the test subjects, who had to repeat and assess the speech material according to various criteria. Specifically the test plan contained the following:

- Participation of 4 speakers (all untrained, two native speakers and two non-native speakers, in each as male/female)
- Use of two speech materials (monosyllables in carrier sentences as well as grammatically correct, but semantically nearly unpredictable sentences); The sentences are constructed according to the principle of “article-adjective-noun-verb-article-object” or “article-noun-verb-article-adjective-object”
- 6 signal-to-noise ratios (pink noise LNA = 60 dB, SNRA = -10, - 5 ... + 15 dB)
- Experimental situations: control condition, hearing impairment, parallel activity (which was included to simulate normally attention absorbing activity)
- Rating questionnaire with the features of subjectively felt speech intelligibility, coping, concentration, annoyance (evaluation on a five-stage scale after presenting four speech items)

32 subjects took part in the test. First they took part in a hearing test and then in a training session. The two test sessions took place at an interval of two to three days. The test subjects were on average 27.6 years old (between 20 and 46 years), had for the most part 'Abitur' (A-levels,

high-school graduation) or 'Fachhochschulreife' (entrance qualification for higher vocational training), one third had completed a vocational or higher qualification. The subjects were paid for taking part.

The proportion of correctly recognised speech items (monosyllables, sentences) were registered. In the case of the sentences both the percentage (proportion of correctly recognised words in the sentence) and the absolute proportion of completely understood sentences were determined, as were the scale ratings of the evaluation questionnaire. The percentage of speech items as a function of the signal-to-noise ratio exhibits as expected a rise with the increase in the signal-to-noise ratio. If one differentiates the results according to speakers there is clear evidence of differences between the individual speakers (Fig. 1). For the sentence intelligibility there is a difference between the speakers of almost 20 % and the masked threshold differs by 5 dB.

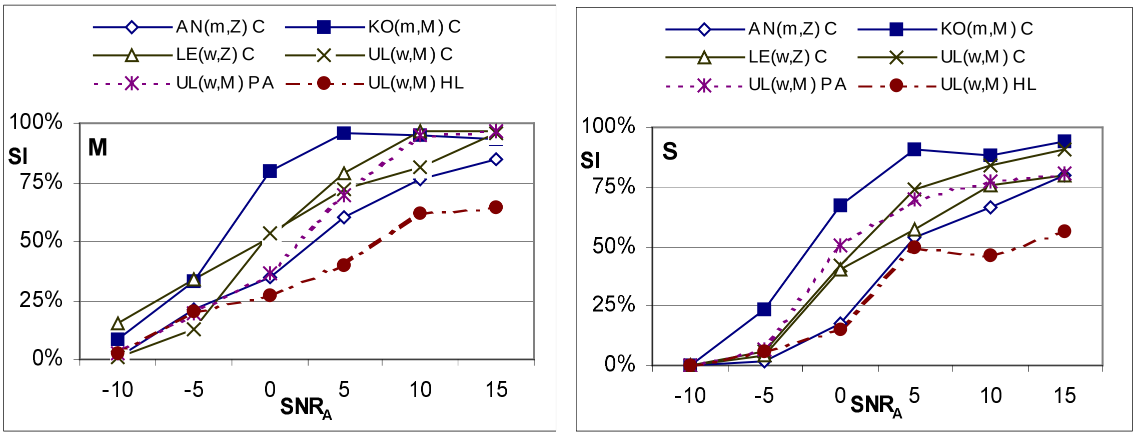

Fig.1: Speech intelligibility (SI) for four speakers (m: male; w: female; M: native speaker; Z: non-native Speaker) with monosyllabic words (M) and sentences (S) for the experimental conditions “Control” (C), “Parallel Activity” (PA); and “Hearing Impairment” (HL) at S-N-Level (SNR_A in dB)

After the presentation of 4 speech items in each case – under constant conditions with respect to signal-to-noise ratio, speaker, speech material and hearing situation – the subjects answered four questions on their rating of the preceding situation. The scales for subjective intelligibility (sSI), coping (CP), concentration (CC) and annyoance (NS) included the following steps:

sSI, CP:	5-excellent	4-good	3-fair	2-poor	1-bad
CC, NS:	1-not	2-not very	3-moderately	4-fairly	5-very

When answering the questions the test subjects had to imagine a communication situation in which it was a matter of correctly understanding what was being communicated so as to pass it on. There was also clear evidence of speaker effects in the ratings. According to the questions concerned, these effects diverge by 0.5 to 0.8 units on the scale (1 to 5).

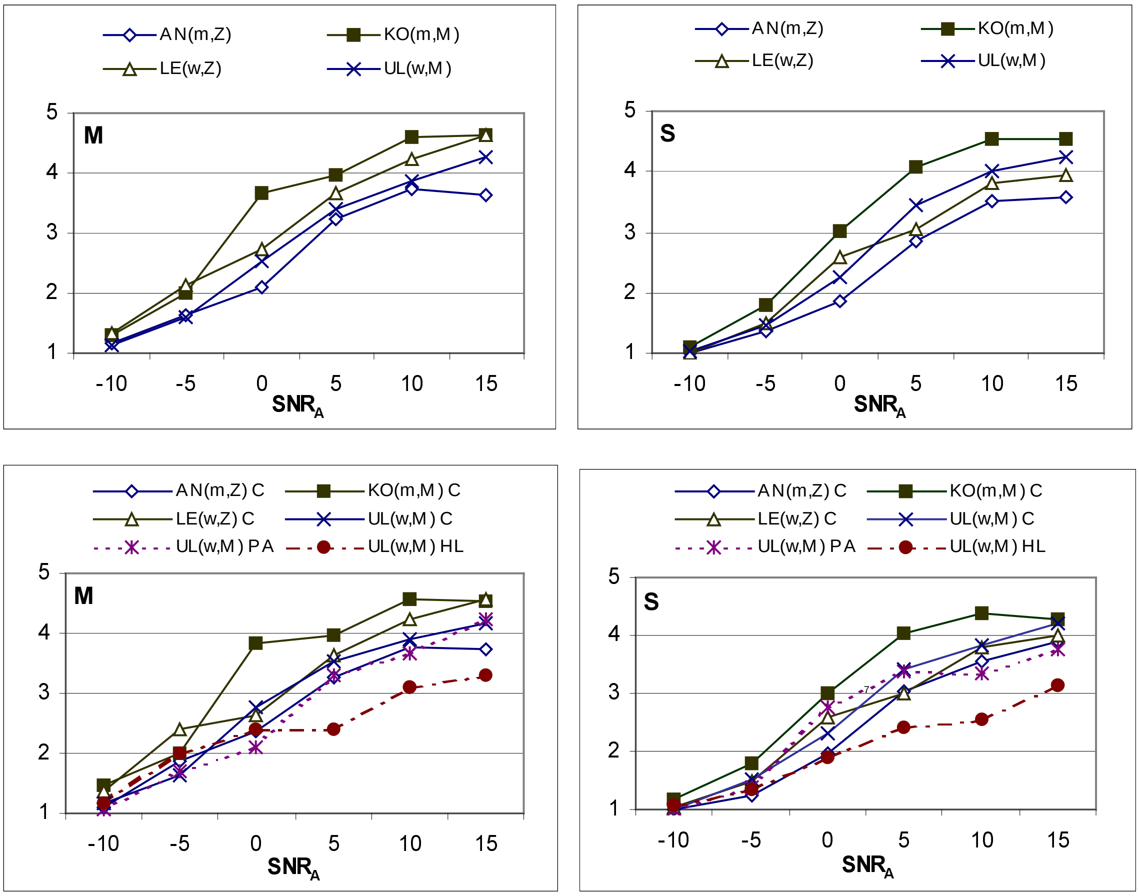

Fig 2: Quality rating of communication: "subjective speech intelligibility" (sSI) for control condition (above) and "Coping" (bottom) for four speakers

The differing speech intelligibility is reflected in the judgement of the test subjects. The more positive subjective evaluation corresponds to a higher speech intelligibility – measured in terms of the percentage of correctly understood items. The correlation of the four evaluation statements and the speech intelligibility is 0.57 to 0.82. In the judgement of the test subjects better speech intelligibility makes it easier for them to cope with their task (hearing, understanding, reproducing; Figure 2b). On the other hand, a poor speech intelligibility forces them to concentrate more and they feel exposed to greater interference.

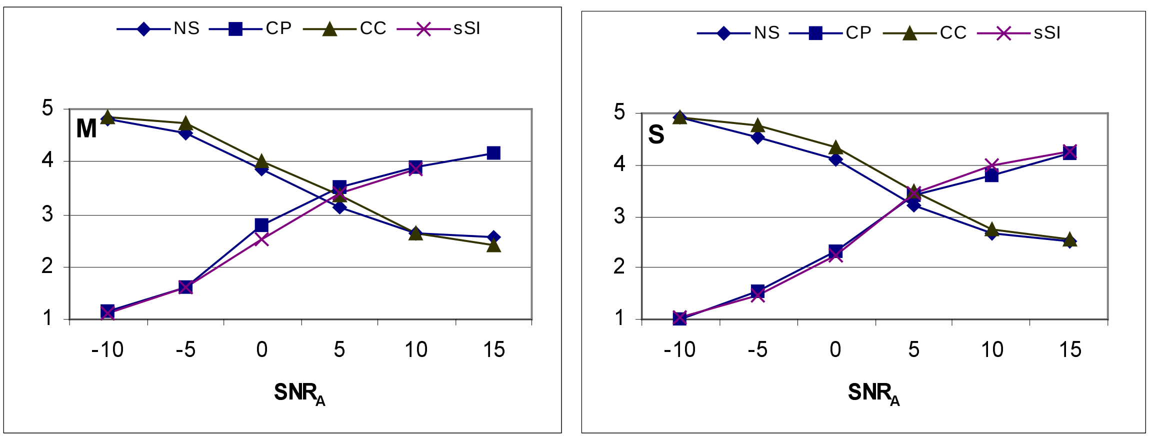

Fig 3: Quality ratings of communication for "Annoyance" (NS); "Coping" (CP); "Concentration" (CC); " subjective Speech Intelligibility" (sSI...) for the control condition (speaker UL)

Conclusion and discussion As described above, one goal of this study is to determine the individual evaluation steps (1 to 5) or their boundaries and to describe them in terms of signal-to-noise ratios.

The results are entered in Table 1 (columns 7 to 4) for the control situation and parallel activity; initially only one speaker (moderate quality, native speaker) was involved. To determine the results a psychometric function was determined for the average ones. With the psychometric function, the boundaries between the evaluation steps (scale 1...5, boundaries: 1.8, 2.6, 3.4, 4.2) give the corresponding signal-to-noise ratios in Table 1 (columns 7 to 4). As a psychometric function a linear and an ArcTan regression has been calculated. Tab 1 presents the results given by the ArcTan; the linear regression gives familiar results (column 8: Coping (CP; sentences) for linear regression: 13.3; 7.5; 1.6; -4.2). The results clearly show that a higher S-N-ratio is necessary than provided for to date in ISO 9921 (2001) (column 6), especially for the higher quality of speech intelligibility. The boundary between good and excellent has to be at SNR_{eff} = 10 to 16 dB.

1	2		3	4	5	6	7	8	9	10	11	12	13	14
Classification of communication			**Exp. 84**	**ISO 9921-1 96**	**SII 97**	**ISO 9921 01**	**Exp. 02/control**				**Exp. 02/parallel activity**			
							sentences		**monosyl.**		**sentences**		**monosyl.**	
							sSI	**CP**	**sSI**	**CP**	**sSI**	**CP**	**sSI**	**CP**
excell.	**5**													
		4.2	7.5	12.0		7.5	10.3	11.5	11.5	11.2	14.6	16.3	13.7	14.2
good	**4**													
		3.4	3.0	6.0	7.5	3.0	5.3	5.8	5.4	4.7	8.2	9.4	10.0	8.1
fair	**3**													
		2.6	-1.5	0.0	-1.5	-1.5	2.2	2.3	1.6	0.6	1.8	2.5	2.2	1.9
poor	**2**													
bad	**1**	1.8	-5.4	-3.0		-6.0	-2.8	-3.5	-4.5	-5.9	-4.5	-4.4	-3.5	-4.2

Table 1: Boundaries between the classifications (excellent to bad) given for the various experiments (coloumn 3, 7 – 14) and standards (column 4 – 6) as an S-N-Level (effective signal-to-noise ratio SNR_{eff} in dB); sSI: Question as to the quality of subjective speech intelligibility, CP: Question as to the quality of coping). (The following was assumed STI = (SNR_{eff} + 15) / 30)

In all the classifications proposed in ISO 9921 must be regarded as decidedly critical. In a standard governing ergonomics a speech intelligibility would only be described as "good" or "excellent" if the situation permits. A speech intelligibility with a signal-to-noise ratio of 7.5 dB cannot be described as "excellent" (Tab 1, Col 6) because very substantial improvements in speech intelligibility are certainly possible with a higher signal-to-noise ratio (10 to 15 dB).

Even considering the many different speakers one is confronted with in everyday situations alone it would appear absolutely essential to create the highest possible signal-to-noise ratio to ensure interference-free communication, a ratio that indeed earns the rating of "excellent". A ratio of 7.5 dB, rated as "excellent" in ISO9921, suggests under certain circumstances a quality of speech communication which we were not able to reproduce in our studies. It must also be considered that with such low signal-to-noise ratios the persons whose hearing ability is not longer optimum owing to lengthy noise exposure or age-related wear are at a considerable disadvantage.

In the study the speech material, and the noise, was also modified through a hearing impairment simulator (Guski et al. 1996) with moderate hearing impairment (frequency response and recruitment. The evaluation of this communication situation clearly shows (Fig. 3) an increase in the signal-to-noise ratio leads to an improvement in the situation, that means, that the rating “Coping” (sentences) is only realised at the boundaries of SNR_{eff} = 25.9; 16.2; 10.2; 0.5 (given by the ArcTan regression).

References

Cox, R. M., Alexander, G. C. & Gilmore, C. (1987) Intelligibility of average talkers in typical listening environments. JASA 81 (5) 1598-1608.

Guski, R., Genuit, K., Hottenbacher, A. & Wühler, K. (1996). Mangel an präventiven Gestaltungsmaßnahmen zum Lärmschutz. Forschungsbericht Fb 745. Bremerhaven: Wirtschaftsverlag NW.

Houtgast, T., & Steeneken, H. J. M. (1984). A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria. Acustica, 54 (4), 185-199.

ISO 9921 (DIS 2001). Ergonomics - Assessment of speech communication.

ISO 9921-1. (1996). Ergonomic assessment of speech communication; part 1: speech interference level and communication distances for persons with normal hearing capacity in direct communication (SIL method). DIS,.

Lazarus, H.; Sust, Ch.A.; Volberg, L.; Kulka, M. (2002) Bewertung der Sprachkommunikation. Bemerkungen zu einer internationalen Norm (ISO 9921). Zeitschrift für Lärmbekämpfung 49 (3), 103-105.

Purdy, S. C. & Pavlovic, C. V.(1992) Reliability, sensitivity and validity of magnitude estimation, category scaling, and paired-comparison judgements of speech intelligibility by older listeners, Audiology 31, 254-271.