Does the Content of Speech Influence its Perceived Sound Quality?

Alexander Raake

Institute of Communication Acoustics (IKA)
Ruhr-University Bochum
D-44780 Bochum, Germany
raake@ika.ruhr-uni-bochum.de

Abstract

From a user's perspective, the speech quality of modern telecommunication systems often differs from that of traditional wireline telephone systems. One aspect is a changed sound of the interlocutor's voice – introduced by an expansion of the transmission-bandwidth to wide-band, by low-bitrate coding and/or by the acoustic properties of specific user-interfaces. In order to quantify the effect of transmission on speech quality, subjective data to be correlated to transmission characteristics have to be collected in auditory tests. In this paper, a study is presented investigating in how far the content of specific speech material used in a listening-only test impacts its perceived sound quality. A set of French speech data was presented to two different groups of listeners: French native speakers and listeners without knowledge of French. The speech material consists of different text types, such as everyday speech or semantically unpredictable sentences (SUS). The listeners were asked to rate the sound quality of the transmitted voice on a one-dimensional category rating scale. The French listeners' ratings were found to be lower for SUS, while those of the non-French listeners did not show any major dependency on text material. Hence, it can be stated that if a given speech sign is understood by the listeners, they are unable to separate form from function and reflect content in their ratings of sound.

1. Introduction

Modern transmission techniques such as Voice over Internet Protocol (VoIP) allow the narrow-band telephone channel (300 – 3400 Hz) to be expanded to wide-band (50 – 7000 Hz), yielding higher speech sound quality and thus higher overall quality. New application scenarios such as mobile telephony have given rise to the introduction of low bit-rate coding as well as of innovative terminal equipment. These, especially from the user's perspective, differ significantly from classical systems with telephone handsets – both in their acoustic properties and in their usage. Consequently, the quality dimension of speech sound quality, which is multidimensional itself, is becoming increasingly important.

Speech sound quality can be regarded as the subjective quality related to the timbre of a transmitted speech signal as perceived by a user. For the development of network components, as well as for network planning and monitoring, it is necessary to have a quantitative estimate of the impact of specific transmission parameters on speech quality at hand. Hence, for network designers or providers and for terminal equipment manufacturers it is desirable to dispose of a corresponding one-dimensional "speech sound quality" rating, in order to quantify the effect of a specific transmission condition – leading to a specific timbre – on speech quality.

For this aim, corresponding quality ratings have to be collected. Such quality ratings can reliably only be obtained in auditory tests from human subjects. The acoustic properties of the terminal and/or the channel specifically affect (one-way) voice transmission quality (for a quality scheme cf. Möller, 2000), which is commonly assessed in listening-only tests (LOTs). Such tests are carried out using databases of speech processed via the transmission system under investigation.

In the process of developing test methods for speech sound quality and its dimensions at our laboratory, appropriate text material to be read by different speakers had to be chosen. A central question that was raised during test design was, to what extent the listeners' judgements of speech sound or vocal quality – judgements aimed at the form of the perceived signal – depend on the content (or 'sense' from the listener's perspective) of the presented text material.

2. Speech Data in Listening Tests

In the past, different types of text material were used in various contexts of speech quality assessment for listening tests, e.g.:

- Phoneme-specific material to study the influence of specific signal-processing on phonetic properties of speech (Huggins and Nickerson, 1985).
- A limited number of short, meaningful sentences to study the multidimensionality of speech quality in a telecommunications context; the attention of listeners is drawn to the measurement object and away from specific speech material (McGee, 1964; Pascal and Boyer, 1990).
- A greater number of meaningful sentences to investigate the overall quality of speech, reflecting the variety of utterances typical for real-life telephone conversations (ITU-T P.800, 1996; Krebber, 1995; Raake, 2000).
- Syntactically correct, but semantically unpredictable sentences (SUS), initially developed to study the intelligibility of speech synthesis systems (for a test method cf. e.g. Benoît et al, 1996). SUS-material has also been used to study the quality of high-quality speech processing devices such as low-bitrate codecs (Bappert and Blauert, 1994).
- Different other text-material on syllable, word and sentence level, primarily developed to investigate the intelligibility of speech synthesis systems; for an overview cf. (Jekosch, 2000).

However, with respect to quality of natural speech, the content or predictability of the text material has only been looked at in terms of its influence on intelligibility, cf. e.g. (Stickney and Assmann, 2001). In these studies, the impact of semantic context on the perception of speech filtered with narrow bandpass filters has been investigated. Low predictability sentences were found to be less intelligible than high predictability sentences. The influence of content in tests related to speech sound

can of course be avoided, if only a small number of meaningful sentences is used (i.e. the same sentence for each condition, repeated with different sentences), as described above in case of the analytical tests aiming at the multidimensionality of speech quality. This stands in contrast to traditional (listening-only) quality tests in the telecommunications context, where a more general evaluation of a system is sought and a higher number of more or less independent sentences are used, in order to reduce training effects and to represent the real-life telephone situation. The last aspect is of major importance, as many tests on sound related features of transmitted speech described in the literature put the subject in a situation that is untypical for a real-life telephone conversation. This possibly reduces the validity of the obtained results. Therefore, the study described here was specifically designed to account for the application context of telephone speech.

In the following, the speech material used in a listening test will be dealt with on the basis of a semiotic view. This provides an illustrative way of test interpretation. For the purpose of this paper, only a very general description of semiotics will be given, which allows the illustration of the described speech sound quality studies.

Language (and speech as a part of it) can be regarded as a sign system, whose signs can be available in written or acoustic form (or in other more specific forms such as Braille). Triadic sign models of semiotic theory differentiate three constituents (correlates) of a sign, which form the so-called semiotic triangle (Peirce, 1986; Ogden and Richards, 1923), which is here shown in general representation (Figure 1: Nöth, 1990).

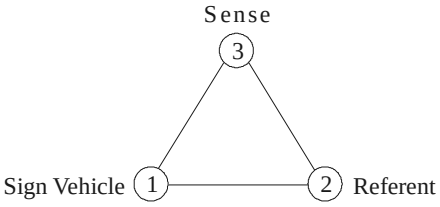

Figure 1: Triadic sign model in general representation: Semiotic triangle (Nöth, 1990).

The three constituents of a sign are closely related to each other. In case of language, the "sign vehicle" or sign carrier are the written word or the acoustic speech signal, thus the form, in which the sign is presented. The "referent" is the object (possibly abstract) the sign stands for. The "sense" is the sense made of the sign by its interpreter. Hence, used as the correlate of a sign, "sense" can be regarded as the role the sign plays or the function it has for a sender or receiver. Here, the situation in which the sign "happens" – e.g. a telephone conversation – is an important factor both for reception and for the sending of the sign. The discussion of the different correlates will not be taken further here, for an overview cf. (Nöth, 1990).

In the phase of test preparation at our lab, the question was raised in how far a more general evaluation of the sound quality of transmitted speech (aiming at the form of the speech sign) can be achieved independently of the speech material used in the test. Speech has sign-character both for a speaker or sender and a listener or receiver. However, the sign appears differently regarded from the sender's or the receiver's perspective. For listening tests – here to measure the sound quality of a specific speech transmission link – speech samples have to be recorded. Already the underlying text material is presented as a sign in written form and is perceived as such by the sender. The sender reads the text and consequently converts it into a sign in acoustic form. The receiver, too, interprets the acoustic speech signal as a sign and is now requested to judge its sound quality, concerning the sign carrier or form. If sentences of low predictability such as SUS are used as text material, which do not have a clear "sense", different possibilities of processing exist both for sender and receiver, e.g.:

(a) The *sender* reads the text aloud like a meaningful statement, in spite of its unpredictability and meaninglessness. The *receiver* is supposed to rate the form or sign carrier, but may not be able to detach himself from the meaninglessness or incompleteness of the sign. In this case, the sound quality is rated lower than in case of meaningful sentences. If the sense of the sentence material does not play a role for his judgement, no difference will be found for SUS and other text material.

(b) The *sender* reflects the meaninglessness or unpredictability of the text material in a corresponding, unusual form (e.g. over-articulated speech). The *receiver* can now process this speech sign in different ways: Either, he is able to abstract in his sound-judgements from the sense and the corresponding form given to the speech sign by the sender, so that the senselessness does not have any influence on the judgement. Or, his/her attention is specifically drawn to the senselessness of the speech material, and speech sound quality is rated lower than for other text material. A third possibility is that the listener rates the unusual form given to the sign by the speaker lower, independently on the sense or predictability of the speech signal.

(c) The unpredictability and thus lower intelligibility of the SUS material leads to a higher listening effort. As only a one-dimensional rating is asked from the listener, SUS is rated more critically, as the quality dimension of intelligibility cannot be separated clearly from the target dimension of "speech sound".

On the basis of these considerations, the results of two auditory listening tests will be discussed, trying to approach answering the question: Are subjects able to differentiate the form from the sense of the speech sign, the latter affected by predictability/intelligibility as well as by the function of the speech sign with respect to the test-situation and -task?

3. Speech Test Material

For this aim, a small database of four sets of different types of French text material read by six native speakers of French (4 female, 2 male) was established. The text material consists of:

- Semantically unpredictable sentences (SUS), cf. (Gibbon, Moore and Winski, 1997).
- Everyday sentences likely to occur e.g. in a telephone conversation. The text material is taken

from the EUROM-database, cf. (Gibbon, Moore and Winski, 1997). The French version of the sentences has been extracted from (Institut de la Communication Parlée, 1994). The sentences were shortened in order to correspond more to the SUS-material in length.[1]

- Short passages taken from "Le petit prince", Saint-Exupéry.
- More complex sentences of philosophical content, namely short excerpts from the collection of thoughts – "Pensées" – by Blaise Pascal.

The untrained speakers were instructed to read the material aloud in a natural way. The speech material was digitally recorded in an anechoic environment at 32 kHz sampling-rate, with the microphone placed 30 cm in front of the speaker's mouth. The active speech level of the recorded samples was adjusted to –26 dB (rel. ovl. of the digital system) to yield telephone typical conditions (cf. ITU-T Rec. P.56, 1993).

The pre-recorded samples were bandpass-filtered using three different transmission bandwidths (Table 1; the 150-3550 Hz condition corresponds to the telephone channel filter defined in ITU-T Rec. G.712). For each text type, a different sentence was used so that every sentence was unknown to the listeners, yielding 6 x 4 x 3 = 72 short speech samples (6 speakers, 4 text types, 3 bandpass-filters). With the bandpass-filters, the spectral characteristics of the speech samples were modified in order to represent different transmission conditions. The bandpass-filters had flat spectral characteristics in their transmission band and were adjusted to yield equal loudness ratings, determined according to (ITU-T Rec. P.79, 1999).

Bandpass [Hz]	mid-frequency [Hz]	Band-width [Hz]
350-2550	944	2200
150-3550	730	3300
50-7000	591	6950

Table 1: Bandpass filters as used for sample processing.

4. Listening Tests: Set-Up

For sample presentation, a specifically built wide-band handset-telephone was used. This way, it was tried to put the subjects as far as possible in a "normal" telephone situation and to establish a corresponding expectation (as shown in Raake, 2000). The handset-telephone consists of the left receiver of a STAX LAMBDA PRO headphone in combination with the lower part of a classical 'Type-7-handset', the latter a model which was used in Germany for many years. Traditional handset-telephones could not be used in the test, as they do not satisfactorily couple the lower frequency range to the listener's ear, which is necessary in case of wide-band transmission.

In two series of auditory tests, the set of speech files was presented in randomised order to two different groups of listeners: (1) French native speakers and (2) German listeners without knowledge of French. In this way, one group of listeners was able to understand and – using semiotic considerations – to assess the speech sign as a whole, while the other group mainly assessed formal aspects of the samples. The subjects were asked to judge the sound quality of the transmitted voice on a one-dimensional category rating scale, as depicted in Figure 1.

The instructions for the listeners were a crucial point in the experiment, as only a one-dimensional rating aiming at the sound quality due to a specific timbre was asked. In order to reflect the application to telecommunications, the subjects were instructed to ignore the characteristics of the voice itself as far as possible and to concentrate mainly on the effect of transmission. The instructions handed to the test subjects in written form

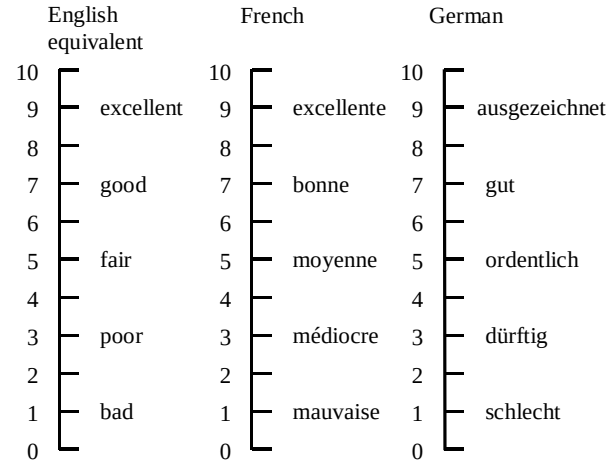

Figure 2: Category rating-scale used in the listening-only test, according to (IEC Publ. 268-13, date unknown). The version on the left shows the English equivalents to the French and German scale-labels, as obtained from the ACR (MOS)-scale traditionally used in telecommunications (ITU-T Rec. P.800, 1996).

read as follows (English equivalent):

[…]The aim is to judge the sound quality of speech transmitted across telephone lines of different transmission quality.[…]

What impression do you have of the sound quality of the transmitted voice?[2]

[1] The text material was not explicitly tested on phonemic balance. However, the SUS-material is composed of words taken from phonemically balanced word-lists, and the setting up of the Eurom-material used here was carried out taking phonemic balance into consideration.

[2] French original:
[…]Le but est de juger de la qualité sonore de la parole transmise à travers des lignes téléphoniques de différentes qualités de transmission.[…]
Quelle impression aviez-vous de la qualité sonore de la voix transmise?
German original:
[...]Ziel ist es, die Klangqualität der über Telefonleitungen unterschiedlicher Übertragungsqualität übertragenen Sprache zu beurteilen.[...]
Welchen Eindruck haben Sie von der Klangqualität der übertragenen Stimme?

Twenty native French listeners, naive with respect to the test, participated in the first test series (10 female, 10 male). As the French listeners were able to understand the presented speech material, they were able to judge the speech sign as a whole. This first test series was carried out at the LMA, CNRS, Marseille, France, and the test subjects were recruited from different labs of the CNRS, Marseille. The age range of the subjects in this test was from 22 to 58 years. The attention of the subjects was not specifically drawn to the content or sense of the speech material. However, it was pointed out to them that also meaningless sentences would be presented. This way, it should be avoided that the semantically unpredictable sentences would be rated worse only because of their unexpected occurrence and lower intelligibility.

Fifteen listeners, by their own account without knowledge of the French language, took part in the second listening experiment carried out at IKA in Bochum, Germany (3 female, 12 male). As they were not able to understand the presented speech samples, their judgments are referring mainly to the form of the speech sign, while they were still able to identify the signal as speech[3]. Hence, it can be assumed that they were able to make analogies to the German language (or others) with respect to the other two sign correlates "sense" and "referent", or to ignore these completely. Here, the listener's age range was from 23 to 45 years. The test subjects, scientific and non-scientific members and students of the lab, were naive listeners with respect to the test.

5. Test Results

The results show that the sense of the text material obviously played a role in the listeners' judgments: The French listeners rated the SUS-sentence material significantly lower than the Eurom-sentences for all three bandwidths, as is depicted in the upper graph in Figure 3. The two literary text types were rated to have speech sound quality between that of the SUS and that of the Eurom materials. Analysis of variance revealed that text type and bandwidth are statistically significant factors for the variance of the sound ratings (0.05 significance-level). The interaction between text-type and speaker was found to be significant, too (0.05 significance-level). In the judgments of the Non-French listeners, on the other hand, no significant difference between the four test sets could be found (lower graph of Figure 3), except for the wideband-case, which will be addressed in more detail in the next paragraph. From the semiotic viewpoint, this implies that the form or carrier of the sign ("sign vehicle") is rated as being similar for the different text types, if due to the incomprehensibility of the language no connection exists to the transported sense.

In case of wideband transmission, the SUS-material was rated slightly lower than the Eurom-material by the non-French listeners. However, the excerpts from "Le Petit Prince" and Pascal were rated significantly better than the SUS material. An analysis of the dependency

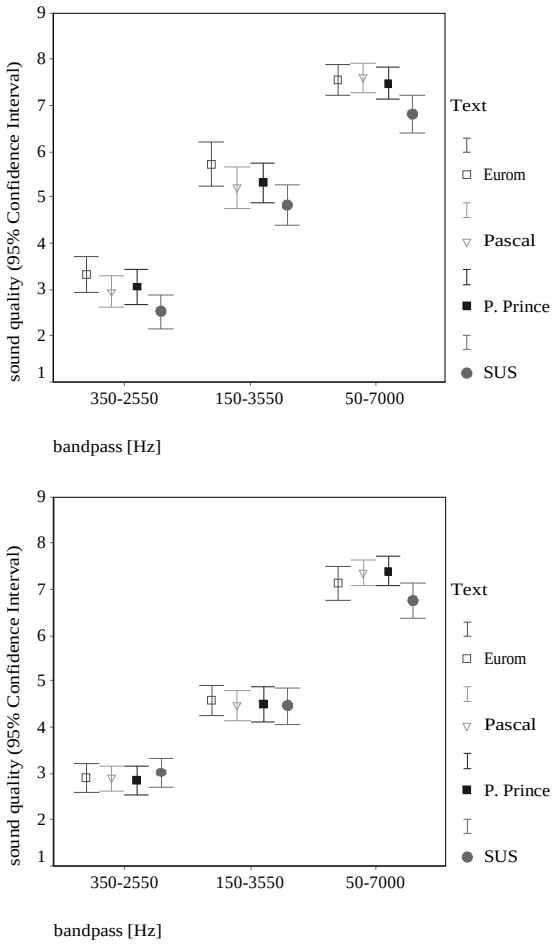

Figure 3. Results averaged over all speakers. French listeners (upper graph) and German listeners (lower graph).

of the judgments on the speaker showed that the SUS-material uttered by one specific (female) speaker ('D') was rated worse by the non-French listeners than the samples of all other speakers, and to some extent also by the French listeners, cf. Figures 4 and 5. If this speaker is discarded in the data analysis, the four text types can no longer be differentiated from the average ratings of the non-French listeners. The average ratings of the French listeners, on the other hand, are only very little effected by the discarding of this particular speaker. The difference between SUS- and Eurom-sentence material is still statistically significant. This implies that the speaker has read the SUS material in a way that especially for non-French listeners was unusual or conspicuous (if the speech sign is not understood, the particular form cannot be anticipated). This could be confirmed by informal listening to the corresponding speech material, which is read in a monotonous way at high pitch. Hence, the speaker as sender of the speech sign has obviously expressed the senselessness with a corresponding carrier, which is rated more negatively by the listener (cf. section 2, case (b)).

[3] This is also the reason why not otherwise processed listening material was used such as modified speech (specifically made "senseless").

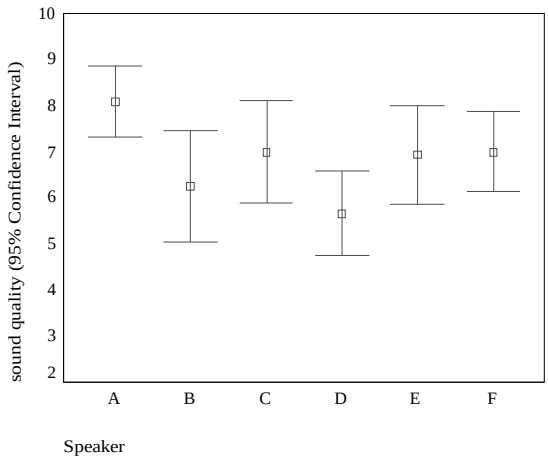

Figure 4: French listeners: Ratings for SUS-text and wideband transmission (50-7000 Hz) as a function of the speaker.

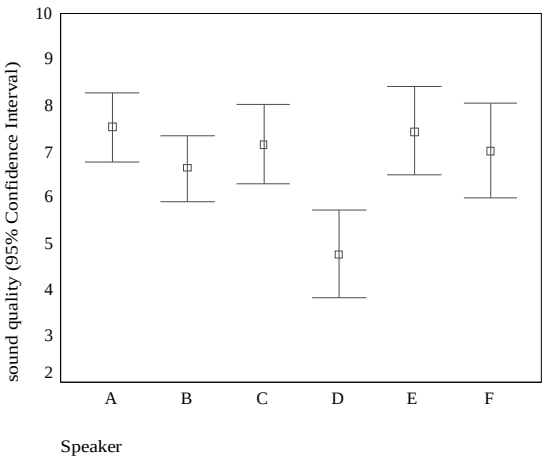

Figure 5: Non-French listeners: Ratings for SUS-text and wideband transmission (50-7000 Hz) as a function of the speaker.

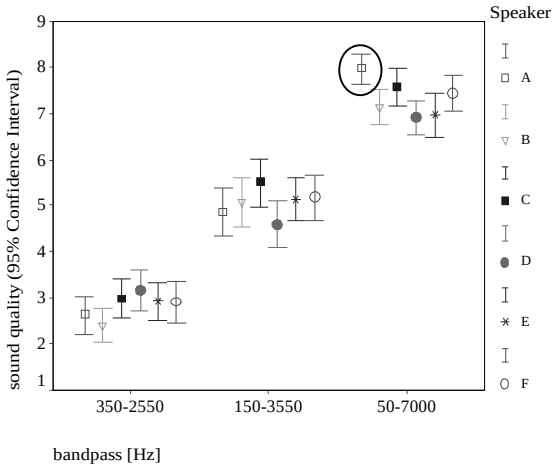

Figure 6: French listeners: Ratings as a function of the speaker, averaged over all text types.

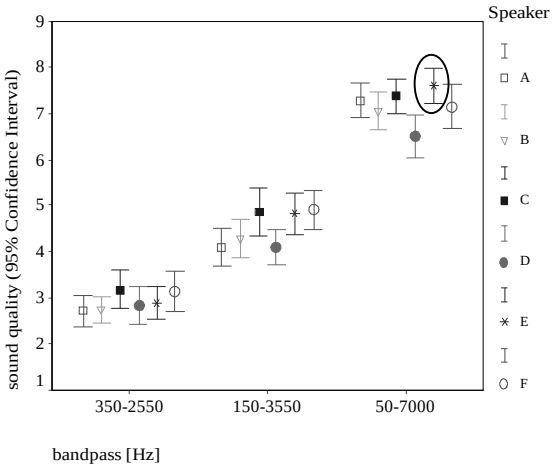

Figure 7: Non-French listeners: Ratings as a function of the speaker, averaged over all text types.

As stated above, the excerpts from “Le Petit Prince” and Pascal were rated higher for wideband than the other two text types by the non-French listeners. This can probably be explained by the fact that this material is read in a literary way by the speakers and may correspond more to the expectation of how French should sound like (if the language is not understood).

The difference between text types found in the ratings of the French listeners can have various reasons:

Either, the listeners are effected by the sense of the speech material, so that the speech sign is perceived as being incomplete, as in case of SUS (cf. section 2, case (a)). Consequently, the speech material is rated more positively, if the sense or content corresponds to the expectation raised by the telephone-typical situation and by the sound of the transmitted speech. This is supported by the results shown in Figure 3, upper graph: For the two lower bandwidths of 350-2550 Hz and 150-3550 Hz, all sentence material which is untypical for a telephone context (Saint-Exupéry and Pascal; SUS) was rated lower than the Eurom sentences, which could occur in an actual telephone conversation.

Another reason for lower ratings in case of SUS could be that also other speakers than speaker ‘D’ expressed the senselessness of the SUS-material with a particular form, which concerns features that were not accessible to the non-French listeners (section 2, case (b)).

If lower intelligibility or higher listening effort played a role in the more critical rating of the SUS-material by the French listeners, this effect should be lower for wideband than for the other two bandwidths. As this is not the case (see Figure 3), it can be assumed that the sound ratings were not or little influenced by the intelligibility of the speech samples, independently on the underlying text material. In particular, the French listeners had specifically been informed that also senseless sentences would be presented. Hence, it is not very likely that the lower ratings of the French listeners for SUS were caused by the lower intelligibility of this material or by the potentially higher listening effort only. Instead, the expectation of a meaningful content certainly played a role, although only ratings of the form were aimed at. For the lowest bandwidth of 350-2550 Hz, a slightly higher listening effort may have

been required, but the negative offset for the SUS material as compared to the other three text types is the same as for the two higher bandwidths.

In the wideband case, the speakers were found to be rated differently by the two groups of listeners, independently on the text type. Here, the two groups had a preference for different speakers (see Figures 6 and 7). In a broad tendency, the French listeners preferred speaker 'A', a young female speaker with higher pitch and very clear articulation. The non-French listeners, on the other hand, had a preference for speaker 'E', a male, older speaker of lower pitch, who read the text material similarly to telling a fairytale. The articulation of this speaker seems less accurate than that of most of the other speakers (which cannot be perceived when the language is not understood), which may have caused that he was rated lower by the French listeners.

It has to be noted that the differences between the sound ratings for the three bandpass conditions are independent on the type of speech material used. The speech material (if the language is understood) mainly introduced a constant shift on the sound quality scale. Consequently, if the aim of a study is the subjective comparison of different sound-related channel characteristics, in principal both low predictability sentences (SUS) and everyday sentences (Eurom) can be used. Other, more specific text types such as the literary "Le Petit Prince" and the excerpts from Pascal, may introduce a conflict with the telephone-related expectation (cf. Figure 3).

6. Conclusions and Outlook

If the sense of the sign is essentially understood, form cannot be separated from content by asking listeners for one-dimensional ratings of vocal sound quality. Instead, several factors potentially contribute to the judgments, e.g.:

- In case of SUS, the sign is perceived by the listener as incomplete because of its unexpected content or lack of sense, and thus the form is rated less positive. Because of the listeners' expectation, the sound quality of speech material rather untypical for the telephone-context is rated slightly lower than more typical material, when the transmission bandwidth corresponds to that of a normal telephone link.
- Since some speakers are unable to read SUS material in a natural way, the unpredictability is reflected in an uncommon form which is unexpected and irritating to the listener.
- The unpredictability and thus lower intelligibility of the SUS sentences leads to a higher listening effort at reduced bandwidths. As only one rating scale is used, the SUS material is rated more critically (the quality dimension of "intelligibility" cannot clearly be separated from the target dimension of "speech sound").

Quality rank orders between different transmission conditions do not seem to be affected by the content of the underlying text material. However, the study described here shows that the sense of the underlying text material has to be taken into consideration, when the development of a method for absolute auditory measurement of formal aspects of speech is aimed at. Obviously, there is no prototypical speech material, which possesses specific sound features that are exclusively measured by such a method. In speech quality research, the separability of the acoustic signal and the speech sign as a whole is often assumed. The study described here proves that this assumption has to be reconsidered depending on the measurement task.

For future tests aimed at sound-related quality dimensions such as timbre or naturalness, an exclusion of sense-related features can best be achieved by choosing a small number of meaningful sentences, using the same sentence for each condition. This corresponds to the praxis of past analytical tests. E.g., an appropriate set of sentences is a reduced version of the Eurom material (Gibbon, Moore and Winski, 1997). However, this approach will only allow analytical, relative tests to be carried out. The results of these tests have to be compared to transmission conditions that are already well quantified with respect to their absolute impact on quality by using appropriate scaling experiments. Only this way, the test results can be applied to network design and monitoring.

In order to investigate the impact of content on speech sound quality in a more multidimensional way, further auditory tests have to be carried out, focusing on a combination of form-related listener judgments with judgments of intelligibility.

7. Acknowledgements

This work has been carried out at the Institute of Communication Acoustics, Ruhr-University Bochum (Prof. J. Blauert, PD. U. Jekosch). It was performed in the framework of a PROCOPE co-operation with the LMA, CNRS, Marseille, France (Dr. G. Canévet). The author would like to thank U. Jekosch, S. Möller and S. Schaden for fruitful discussions and S. Meunier (CNRS Marseille) for her help in auditory test organization at the LMA.

8. References

Bappert, V. and J. Blauert, 1994. Auditory Quality Evaluation of Speech Coding Systems. *Acta Acustica*, 2:49-58.

Benoît, C., M. Grice, and V. Hazan, 1996. The SUS Test: A Method for the Assessment of Text-to-Speech Synthesis Intelligibility using Semantically Unpredictable Sentences. *Speech Communication*, 18 (4):381-392.

Gibbon, D., R. Moore, R. Winski, eds., 1997. *Handbook of Standards and Resources for Spoken Language Systems.* D-Berlin: De Gruyter.

Huggins, A., R. Nickerson, 1985. Speech quality evaluation using "phoneme-specific" sentences. *J. Acoust. Soc. Am.*, 77 (5):1896-1906.

IEC Publ. 268-13 (date unknown). *Listening Test on Loudspeakers*. International Electrotechnical Commission, CH-Geneva.

Institut de la Communication Parlée (ICP), F-Grenoble, 1994. Eurom 1 Multilingual Database; 40 blocks of 5 thematically linked sentences: http://www.icp.grenet.fr/

ITU-T Recommendation G.712, 1992. *Transmission Performance Characteristics of Pulse Code Modula-*

tion. CH-Geneva: International Telecommunication Union.

ITU-T Recommendation P.56, 1993. *Objective Measurement of Active Speech Level*. CH-Geneva: International Telecommunication Union.

ITU-T Recommendation P.79, 1999. *Calculation of Loudness Ratings for Telephone Sets*. CH-Geneva: International Telecommunication Union.

ITU-T Recommendation P.800, 1996. *Methods for Subjective Determination of Transmission Quality*. CH-Geneva: International Telecommunication Union.

Jekosch, U., 2000. Sprache hören und beurteilen: Ein Ansatz zur Grundlegung der Sprachqualitätsbeurteilung. Habilitation thesis. D-Essen: University Essen.

Krebber, W., 1995. *Sprachübertragungsqualität von Fernsprech-Handapparaten*, Vol. 357. Series 10. D-Düsseldorf: VDI-Verlag GmbH.

McGee, V.E., 1964. Semantic Components of the Quality of Processed Speech. *J. of Speech and Hearing Res.*, 7:310-323.

Möller, S. and A. Raake, 2002. Telephone Speech Quality Prediction: Towards Network Planning and Monitoring Models for Modern Network Scenarios. *Speech Communication*, in press.

Möller, S., 2000. *Assessment and Prediction of Speech Quality in Telecommunications*. USA-Boston: Kluwer Academic Publ.

Nöth, W., 1990. *Handbook of Semiotics*. USA-Bloomington: Indiana University Press.

Ogden, C. and I. Richards, 1923. *The Meaning of Meaning*. USA-New York: Harcourt.

Pascal, D. and M. Boyer, 1990. Multidimensional Perceptive Measurement of Quality: Comparative Performance of Two Methods. In: *Proc. 13th Int. Symp. on Human Factors in Telecommunications*, I-Turin, 519-520.

Peirce, C., 1986. Semiotische Schriften, vol. 1. D-Frankfurt/Main: H. Suhrkamp.

Raake, A. (2000). Perceptual Dimensions of Speech Sound Quality in Modern Transmission Systems. In: *Proc. Int. Conf. Spoken Language Processing. (ICSLP 2000)*, CHN-Beijing, 4:744-747.

Stickney, G. and P. Assmann, 2001. Acoustic and Linguistic Factors in the Perception of Bandpass-Filtered Speech. *J. Acoust. Soc. Am.*, 109 (3):1157-1165.