International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.85

Amendment 1

(03/2013)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Methods for objective and subjective assessment of speech quality

---

A method for subjective performance assessment of the quality of speech voice output devices

Amendment 1: New Appendix I – Evaluation of speech output for audiobook reading tasks

Recommendation ITU-T P.85 (1994) – Amendment 1

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	**Series**	**P.80**
		P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.85

A method for subjective performance assessment of the quality of speech voice output devices

Amendment 1

New Appendix I – Evaluation of speech output for audiobook reading tasks

Summary

Amendment 1 to Recommendation ITU-T P.85 introduces a test methodology that addresses the evaluation of speech output for audiobook reading tasks.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T P.85	1994-06-21	12
1.1	ITU-T P.85 (1994) Amd. 1	2013-03-28	12

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T P.85

A method for subjective performance assessment of the quality of speech voice output devices

Amendment 1

New Appendix I – Evaluation of speech output for audiobook reading tasks

(This appendix does not form an integral part of the Recommendation.)

Whereas the method and the questionnaires given in the main body of this Recommendation are adequate for applications providing vocal answers related to telephone directory inquiries, weather forecast, mail order and similar tasks, they are less adequate for applications where longer text paragraphs and potentially literature are read through synthetic speech output, as is the case in audiobook reading tasks. For such applications, the task of the voice output is not pure information provisioning, but rather an expressive, emotion-triggering vocal output, putting more emphasis on different aspects of speech output quality than in traditional telecom services [9].

To address these aspects, a test methodology is presented in this Appendix which addresses mainly two perceptual dimensions of such speech output, namely its *listening pleasure* and its *prosody*. These dimensions were extracted with the help of principal axis factor (PAF) analysis from auditory test results involving typical examples of TTS systems and typical (prosaic and poetic) text passages, as summarized in Table I.1. Details on the derivation of the methodology are given in [8] and [10].

Table I.1 – Examples of text categories for audiobook reading tasks

ID	Category	Author	Book
1	Long sentences	Sven Regener	Der kleine Bruder[1]
2	Direct speech, incomplete sentences	Douglas Adams	The Hitchhiker's Guide to the Galaxy
3	Higher level of lexis, complex sentence structure	Charles Dickens	The Adventures of Oliver Twist
4	Poetic, picturesque	Antoine de Saint-Exupéry	Wind, Sand and Stars
5	Direct speech, basic language	Tommy Jaud	Resturlaub[1]
6	Action, short sentences	Thomas Harris	Hannibal
7	Children's book	Astrid Lindgren	Pippi Longstocking
8	Thriller	Ken Follett	Code to Zero

In the following paragraphs, the differences from the standard procedure described in the main body of this Recommendation are outlined. All other characteristics of the test methodology (selection of test participants, listening environment, etc.) should remain as described previously.

1 No English translation available.

I.1 Speech material

Passages from standard audio books should be used as material for the listening test. The passages should cover a wide variety of writing styles and book categories, potentially including thrillers, funny books, action-packed passages, books for children, books with very long sentences, and passages containing almost only direct speech, in case that these styles are considered as representative for the later application. Exemplary text categories are given in Table I.1. The passages should have an approximate length of one minute when read aloud.

I.2 Rating scales

For assessing speech output in audiobook reading tasks, a questionnaire with eight items is proposed, see Table I.2. Four of these items are modified versions of items in the main body of this Recommendation, and four items are added which are specific to audiobook reading tasks. These additional items were selected based on the review of current literature, and address prosodic elements like communicative, structuring, aesthetic and emotional aspects that can be seen as the most important factors for reading and interpreting books [11].

Table I.2 – Questionnaire items for audiobook reading tasks

Standard items (modified)	Additional items related to audiobooks
Overall impression	Speech pauses
Voice pleasantness	Intonation
Listening effort	Emotion
Acceptance	Word stress

The scales address the following aspects of quality:

- Overall impression

This scale evaluates the overall quality of the synthesized signal and was adopted with slight modifications.

- Voice pleasantness

This scale measures the degree of voice pleasantness from unpleasant to pleasant and was adopted with slight modifications.

- Listening effort

This scale describes the effort a listener is required to make when listening to this voice over a longer period of time. It was adopted from this Recommendation with slight modifications.

- Acceptance

The acceptance scale from the main body of this Recommendation was transformed into a continuous scale.

- Speech pauses

This scale evaluates if punctuation marks (e.g., period, comma, question mark, exclamation mark, colon, etc.) have been converted into appropriate speech pauses between words, sentences and paragraphs.

- Intonation

This scale determines if the produced pitch curve fits with the sentence type, e.g., the pitch of interrogative sentences usually increases at the end of a sentence whereas the pitch of declarative sentences decreases.

- Emotion

Variation of emotion is achieved by variations of sound pressure, intonation, speech pauses and volume [11]. To ensure an authentic reading experience, the voice should reflect the atmosphere of the scene and the moods of the characters.

- Word stress

Unnatural stress and accentuations often result in very annoying voices and thus also cause problems in text comprehension [11]. This scale is used to determine if the stress was perceived as unnatural or confusing.

The scales are presented in a continuous scale layout with overflow areas on both extremities, in order to avoid saturation effects. Such effects might occur if a characteristic of one stimulus is already rated in an end category of a scale, and the subsequent stimulus is considered as having an even more extreme characteristic of that kind. The scales may be presented on paper (by collecting responses with a pencil), or on a graphical user interface (by collecting responses with sliders on a computer screen).

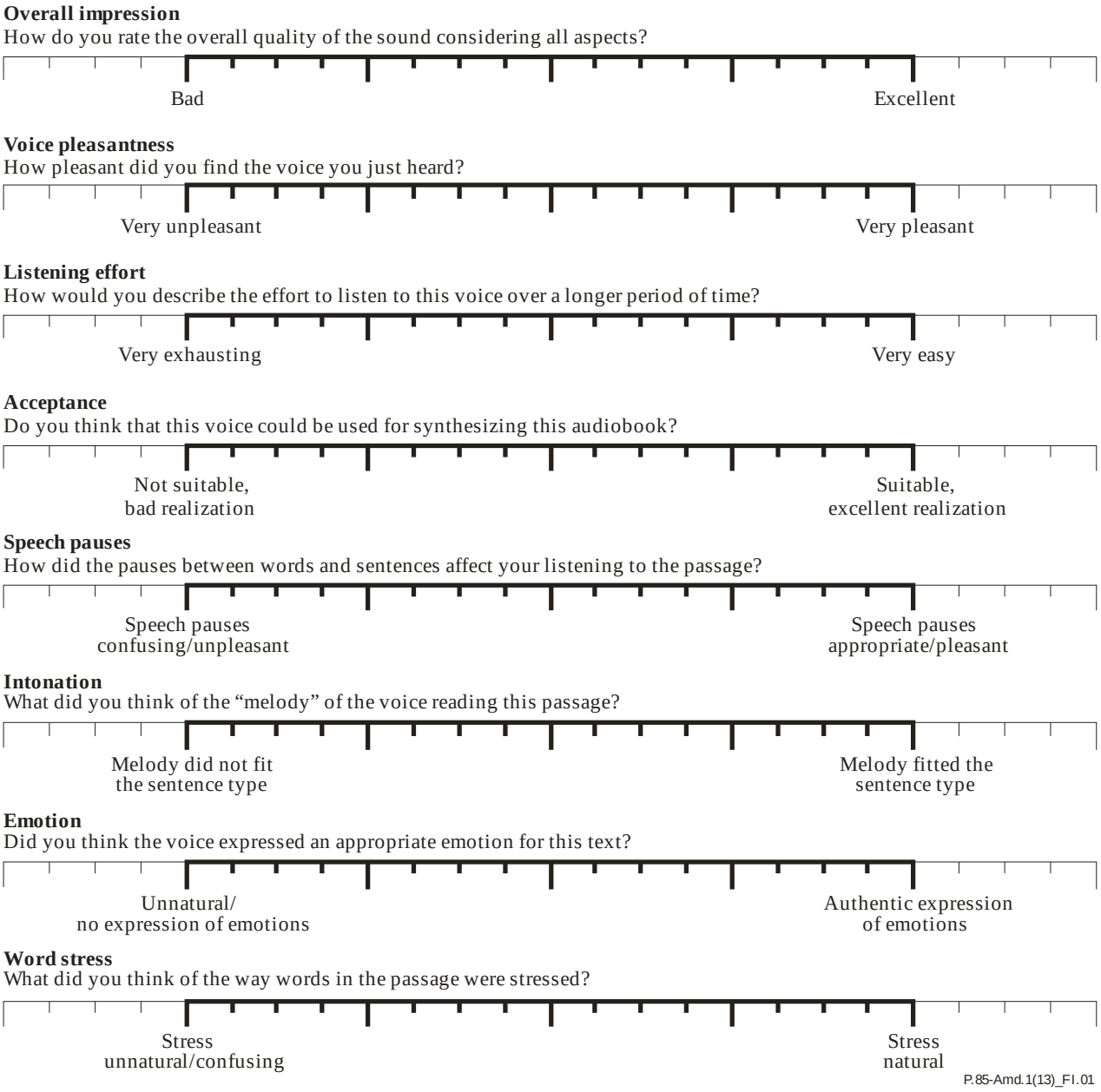

Figure I.1 – Graphical layout of rating scales for audiobook reading tasks

I.3 Test procedure

Test participants are instructed to first rate their overall impressions of the stimulus on the continuous rating scale ranging from bad to excellent. Subsequently, quality estimates for the other eight scales are solicited. To avoid any impact with regard to the order, the sequence of these subsequent eight scales (except MOS) should be randomized between participants. In order to familiarize themselves with the test procedure, participants first have to pass a training phase with approximately two stimuli that are not included in the main test.

Bibliography

[8] F. Hinterleitner, G. Neitzel, S. Möller, C. Norrenbrock, "An Evaluation Protocol for the Subjective Assessment of Text-to-Speech in Audiobook Reading Tasks", in: *Blizzard Challenge 2011 Workshop*, Sept. 2, Torino, 2011.
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.227.8997

[9] F. Hinterleitner, S. Möller, C. Norrenbrock, and U. Heute, "Perceptual Quality Dimensions of Text-to-Speech Systems", in: *Proceedings of the 12th Annual Conference of the ISCA (Interspeech 2011)*, International Speech Communication Association (ISCA), 2011.
http://www.interspeech2011.org/conference/programme/sessionlist-static.html

[10] F. Hinterleitner, C. Norrenbrock, and S. Möller, "Perceptual Quality Dimensions of Text-to-Speech Systems in Audiobook Reading Tasks", in: *Proceedings of the 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013)*, Bielefeld, 2013.

[11] U. Rautenberg and T. Schnickmann, Das Hörbuch – Stimme und Inszenierung. Harrassowitz Verlag, Wiesbaden, 2007, Chapter "Die Stimme im Hörbuch: Literaturverlust oder Sinnlichkeitsgewinn?", pp. 21-54.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Terminals and subjective and objective assessment methods**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2013