ETSI TR 102 949 V1.1.1 (2014-09)

**Speech and multimedia Transmission Quality (STQ);
Wideband and Superwideband speech terminals;
Perceptually motivated parameters**

Reference
DTR/STQ-183

Keywords
loudness, speech, superwideband, terminal, wideband

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights 4
Foreword 4
Modal verbs terminology 4
Introduction 4
1 Scope 5
2 References 5
2.1 Normative references 5
2.2 Informative references 5
3 Definitions, symbols and abbreviations 7
3.1 Definitions 7
3.2 Symbols 8
3.3 Abbreviations 8
4 Sound levels and loudness 9
4.1 Loudness 9
4.2 Impact of signal level and spectrum (including pitch and frequency adjustment and balance) 10
5 Speech/Sound Quality and Intelligiblility 10
5.1 Speech intelligibility assessment 10
5.2 Impacts of impairments on speech intelligibility 11
5.3 Other quality parameters 11
5.3.1 Audio clarity 11
5.3.2 Naturalness 11
Annex A: Considerations about loudness assessment 12
Annex B: Objective and subjective tests: Influence of frequency bandwidth on loudness 16
B.1 Loudness depending on bandwidth and codec 16
B.1.1 Simulation process 16
B.1.2 Results presentation 19
B.1.2.1 Level depending on bandwidth 19
B.1.2.2 Level depending on codec 21
B.1.2.3 Loudness depending on bandwidth 22
B.1.2.4 Loudness depending on codec 23
B.2 Subjective Test results 25
B.2.1 Introduction 25
B.2.2 Selection and preparation of test signals 25
B.2.3 Description of the subjective test 28
B.2.3.1 Description of the response scale 28
B.2.3.2 Calibration of the sound reproduction chain 28
B.2.4 First stage of the subjective test: Measurement of individual loudness function 29
B.2.4.1 Dynamic range determination 29
B.2.4.2 Measurement of individual loudness function 30
B.2.4.3 Results for individual loudness functions 31
B.2.5 Second stage of the subjective test: Assessment of test signal loudness 31
B.2.5.1 Assessment of test signal loudness 31
B.2.5.2 Conversion from points to phons 32
B.2.6 Results for test signal loudness 33
B.2.6.1 Results averaged over all samples 33
B.2.6.2 Detailed results per sample 34
B.2.6.3 Results averaged over all samples, except Sample 4 36
Annex C: Bibliography 38
History 39

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**may not**", "**need**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

There are in practice a lot of factors that may affect the quality and usability of terminals in real use, including the users' behaviour, such as the real positioning of the terminal relative to ear(s), the influence of the distance and of the environment (noise, reverberation) the real voice level of the distant speaker, etc. The present document is intended to provide initial answers to questions raised:

- on the potential impact of speech spectrum and speech level on loudness;
- about differences perceived by the distant user when the local user uses alternatively different pick-up systems.

Technical reports on accessibility have shown that speech quality degradation may affect more strongly people with hearing impairments. Hence it appears that it is needed to consider other criteria than overall quality (e.g. intelligibility or clarity) and to consider the potential impact of loudness.

1 Scope

The present document investigates new perceptually motivated parameters defining more closely the audio quality, such as loudness, fidelity and intelligibility of the speech as perceived by the user, for wideband and superwideband speech terminals.

The annexes detail studies about loudness of received signals, depending on the transmission bandwidths, the codecs, the types of transmitted signals and compare results from different computation models.

The intention of the present document is to provide alternative or new quality parameters and test methods to be implemented in the relevant standards and specifications.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

Not applicable.

2.2 Informative references

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI ES 202 739: "Speech and multimediaTransmission Quality (STQ); Transmission requirements for wideband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user".

[i.2] ETSI ES 202 740: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

[i.3] ETSI TS 103 739: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".

[i.4] ETSI TS 103 740: "Speech and multimedia Transmission Quality (STQ);Transmission requirements for wideband wireless terminals (handsfree) from a QoS perspective as perceived by the user".

[i.5] ETSI ETS 300 807: "Integrated Services Digital Network (ISDN); Audio characteristics of terminals designed to support conference services in the ISDN".

[i.6] Recommendation ITU-T P.79: "Calculation of loudness ratings for telephone sets".

[i.7] Recommendation ITU-T P.58: "Head and torso simulator for telephonometry".

[i.8] Recommendation ITU-T P.581: "Use of head and torso simulator (HATS) for hands-free and handset terminal testing".

[i.9] Recommendation ITU-T P.501: "Test signals for use in telephonometry".

[i.10] Recommendation ITU-T P.863: "Perceptual objective listening quality assessment".

[i.11] Recommendation ITU-T P.10/G.100: "Vocabulary for performance and quality of service".

[i.12] ANSI 53.4-2007: "American National Standard procedure for the computation of loudness of steady sound".

[i.13] DIN 45631, 1991: "Procedures for calculating loudness level & loudness".

[i.14] ETSI EN 301 549: "Accessibility requirements suitable for public procurement of ICT products and services in Europe".

[i.15] ISO 532 B: "Method for calculating loudness", International standard (1975).

[i.16] ETSI TS 102 924: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for Superwideband/Fullband headset terminals from a QoS perspective as perceived by the user".

[i.17] ETSI TS 102 925: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for Superwideband/Fullband handsfree and conferencing terminals from a QoS perspective as perceived by the user".

[i.18] ISO TR 22411: "Ergonomics data and guidelines for the application of ISO/IEC Guide 71 to products and services to address the needs of older persons and persons with disabilities".

[i.19] Recommendation ITU-T G.711: "Pulse Code Modulation (PCM) of Voice Frequencies".

[i.20] Recommendation ITU-T G.722: "7 kHz audio-coding within 64 kbit/s".

[i.21] ETSI ES 203 038: "Speech and multimedia Transmission Quality (STQ); Requirements and tests methods for terminal equipment incorporating a handset when connected to the analogue interface of the PSTN".

[i.22] Recommendation ITU-T P.50: "Artificial voices".

[i.23] ANSI/ASA S3.5-1997 (R 2012) American National Standard: "Methods for Calculation of the Speech Intelligibility Index".

[i.24] Recommendation ITU-T P.862: "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs".

[i.25] Meunier S. and al.: "Calcul des indicateurs de sonie: revue des algorithmes et implémentation", 10ème Congrès Français d'Acoustique (2010).

[i.26] Zwicker E. and Fastl H.: "Psychoacoustics: Facts and models", 2nd Edition, Springer-Verlag, Berlin (1999).

[i.27] Glasberg B. R. and Moore B. C. J.:"A model of loudness application to time-varying sounds", J. Audio Eng. Soc, Vol. 50, n 5, 331-342 (2002).

[i.28] Sridhar Kalluri, Starkey Hearing Research Center (Berkeley, USA): "High frequency sound for the hearing impaired", ITU-T Workshop on "From Speech to Audio: bandwidth extension, binaural perception" Lannion, France, 10-12 September 2008.

[i.29] Ute Jekosch. TU Dresden: "Test on overall quality as perceived by high frequency hearing impaired subscribers", ITU-T SG12 - C101- September 2007.

[i.30] Cyril Plapous, Jean-Yves Le Saout, Jean-Yves Monfort: "Loudness depending on bandwidth and Codec". ETSI STQ(13)42-029r1.

[i.31] John Beerends, Ronald Van Buuren, Jeroen Van Vugt and Jan Verhave: "Objective Speech Intelligibility Measurement on the basis of natural speech in combination with perceptual modeling". JAES, Vol.57, N 5, 2009 May.

[i.32] Søren Jørgensen and Torsten Dau: "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing", J. Acoust. Soc. Am. Volume 130, Issue 3, pp. 1475-1487 (2011); (13 pages).

[i.33] Jianfen Ma, Yi Hu and Philipos C. Loizou: "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions", J. Acoust Soc Am. 2009 May; 125(5): pp. 3387-3405.

[i.34] Jean-Yves Monfort, JYMC.I.S.: "Status of Speech intelligibility studies and models for hearing impaired people. Plans for standards".

NOTE: Available at: http://docbox.etsi.org/Workshop/2014/201406_HFWORKSHOP/S02_Speech_Intelligibility/S02_Monfort_JYMLCIS.pdf

[i.35] Ewert and Dau: "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing", J. Acoust. Soc. Am. 108, pp. 1181-1196] (2000).

[i.36] ANSI S3.2-1989: "American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems".

[i.37] Recommendation ITU-T G.729.1 (Annex E): "G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729".

[i.38] Recommendation ITU-T G.722.1 (Annex C): "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss".

[i.39] Recommendation ITU-T G.719: "Low-complexity, full-band audio coding for high-quality, conversational applications".

[i.40] Recommendation ITU-T P.56: "Objective measurement of active speech level".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions given in Recommendation ITU-T P.10/G.100 [i.11] apply:

Definitions "generally used in psychoacoustics"

articulation index: A measure of the intelligibility of voice signals, expressed as a percentage of speech units that are understood by the listener when heard out of context. The articulation index is based on partially empirical, partially theoretical principles to predict the speech intelligibility under known signal-to-noise conditions.

loudness: Loudness belongs to a category of intensity sensations. Loudness is that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud. Loudness takes into account the spectral and temporal sensitivity of the human ear. Generally masking effects in time and frequency are taken into account. The loudness level measure according to Zwicker [i.26] was created to characterize the loudness sensation of tones. The loudness calculation procedures for stationary signals are defined in several standards such as [i.12], [i.13] and [i.15]. For the calculation of the loudness of time variant signal different models are known.

pitch: Pitch is an attribute of an auditory image that reflects listeners' impression on the location of the dominant spectral component along the frequency scale. In the case of complex harmonic tones, the pitch corresponds to a frequency close to the frequency difference between the harmonic components, i.e., the fundamental frequency.

roughness: The amplitude or frequency modulation of tones lead to different hearing events. A sound is perceived as rough if the envelope fluctuation is within the frequency range from 20 Hz to 300 Hz. The roughness perceived depends on the modulation frequency and the modulation depth.

sharpness (also used: thinness): Sharpness is the centre of gravity of the spectrum and gives information on the balance between high and low frequency energy in the sound. As more the centre of gravity (of the spectral envelope) is moved to higher frequencies, as sharper a sound is perceived.

spaciousness: Spaciousness is a multidimensional perception of the auditory image that reflects listeners impression of the location of a sound source and of the characteristics of the space in which the sound event exists. While the perception of loudness, pitch, duration and timbre is restricted to monotic hearing, the perception of spaciousness typically arises from dichotic stimulation.

timbre (sound colour): Timbre is that attribute of auditory sensation in terms of which a listener can judge to which extent two sounds, similarly presented and having the same loudness and pitch and duration, are dissimilar. Timbre depends primarily on the spectrum of the stimulus but also depends on the waveform, the sound pressure, the frequency location of the spectrum and the temporal characteristics of the stimulus.

tonality: Tonality is the logarithm of the ratio between the arithmetical and geometrical means of the spectrum and gives information on the presence of high peaks in the spectrum.

Definitions for transmission bandwidths

fullband telephony: Transmission of speech with a nominal pass-band wider than 50 Hz to 14 000 Hz, usually understood to be 20 Hz to 20 000 Hz.

narrowband telephony: Transmission of a signal (either speech or data) through a telephonic network with a nominal pass-band of 300-3400 Hz.

super-wideband telephony: Transmission of speech with a nominal pass-band wider than 100-7000 Hz, usually understood to be 50-14000 Hz.

wideband telephony: Transmission of speech with a nominal pass-band wider than 300-3400 Hz, usually understood to be 100-7000 Hz.

3.2 Symbols

For the purposes of the present document, the following symbols apply:

Son	Loudness is a subjective scale expressed in **sons**. By convention, the value of 1 son is attributed to the loudness of a pure tone of frequency 1 000 Hz at 40 dB SPL. Thus, a sound with loudness equal to 2 sons will be perceived as 2 times louder than a sound with a loudness of 1 son.
Phon	Loudness can also be expressed in **phons**, knowing that phon scale is equal to scale of dB SPL for a pure tone of 1 000 Hz.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

AI	Articulation Index
AMR	Adaptive Multi-Rate
CVC	Consonant-Vowel-Consonant
FB	Fullband
GAT	Group Audio Terminal
HATS	Head and Torso Simulator
IP	Internet Protocol
IRS	Intermediate Reference System
NB	Narrowband
RLR	Receive Loudness Rating

NOTE: See Recommendation ITU-T P.79 [i.6]).

RMS	Root mean square
SII	Speech Intelligibility Index
SPL	Sound Pressure Level
STI	Speech Transmission Index
STL	Short term loudness
SWB	Superwideband
WB	Wideband

4 Sound levels and loudness

This clause is mainly dedicated to normal hearing people. Additional data are needed for hearing impaired people (including ageing people) even if some guidances on loudness, pitch and frequency adjustments may be found in ISO TR 22411 [i.18].

The sound levels are currently expressed in dB SPL (reference 20 µPa) or dBPa (reference 1 Pa), and may be expressed using A-weighting. Loudness computation is more dedicated to characterize the level as perceived by the user,

4.1 Loudness

Several methods have been developed to compute the loudness. Annex A presents a practical proposal to assess loudness, compares the results provided by different computation models and summarizes an initial study and preliminary results for objective loudness assessment.

A first set of results for narrowband and wideband speech transmission, on a comparison between loudness ratings and loudness, shows that there is a rather good relationship between RLR and loudness for **handset mode**, independently of the input signal level and the speech bandwidth. This is mainly due to the fact that terminals in handset mode do not implement speech processing systems or implement systems with a limited impact on the level. A positive aspect is that for handset mode, and consequently from end-to-end transmission with handset terminals at both ends, the RLR calculation provides a good way to ensure the relevant loudness perceived by the users. Due to speech processing implemented in devices, the relationship between RLR and loudness is different for **handsfree** mode. The different behaviour in handset and handsfree modes relative to the signal level may explain the complaints of users when switching from handset to handsfree (or vice versa) and complaints about the loudness difference, as identified in the documents listed in the introduction of the present document.

The annex B of the present document provides data from TC STQ meeting documents.

Results on "Loudness depending on bandwidth and codec" [i.30] based on the model computations conclude as follows: *"loudness is sensitive to the bandwidth difference and there is a significant difference in loudness when switching from NB [300 Hz - 3,4 kHz] to WB [50 Hz - 7 kHz]. Loudness is also sensitive to the frequency range of codecs, especially for the one specifically designed for speech where the loudness increases from AMR (NB) to OPUS (FB) coding"*.

Another conclusion indicates that it would be possible for FB, SWB and WB signals to determine their perceived level by calculating the Loudness Rating in NB mode and appointing the offset in Phon between the considered bandwidth (WB, SWB or FB) and the NB. This solution would be compatible with the existing one in NB and would provide a way to get perceived levels in upper bandwidths too.

Subjective test results are provided, combining the four bandwidths (NB, WB, SWB and FB), the different coders and for different scenarios (speech only, speech mixed with music, speech mixed with background noise, musics,..). They may be compared with the available objective test results.

An overall conclusion is that the results of the subjective tests confirm that loudness increases with bandwidth extension, including when codec are applied. There is a significant gap between loudness in NB and WB conditions. There is also a smaller gap between WB and SWB conditions that is statistically significant in 7 conditions out of 9.

Between SWB and FB conditions the loudness differences are not significant.

As a conclusion on loudness computation, even if new objective loudness measurements are needed to enhance the potential correlations between subjective and objective test results, the results already available show that there is a significant interest to consider the loudness as an additional parameter to be used in future standards and specifications and to recommend objective loudness measurement methods.

As the references for the subjective tests are obtained with normal hearing subjects, it would be appropriate to investigate similar studies for hearing impaired people.

4.2 Impact of signal level and spectrum (including pitch and frequency adjustment and balance)

Based on the conclusions of "High frequency sound for the hearing impaired" [i.28], it can be said that:

- For Normal hearing people there is a Preference for extended bandwidth up to 16 kHz - The study has shown that the subjects have a preference for bandwidths greater than 10 kHz, proving the interest for superwideband transmission.
- The study has shown that hearing impaired people have shown the benefit of expanding the bandwidth from 4 kHz to 6 kHz, proving the interest for wideband transmission.

TS 102 924 [i.16] and TS 102 925 [i.17] include a clause dedicated to Equalization in the receive part of the terminal: "This type of terminal may be used for reproduction of signals other than pure speech (e.g. music) for which user's preference may be different in term of sound signature. So, the terminals (earphones, handsfree and GAT) may implement an equalization function adjusting frequency response according to user's preference." If such statements are agreed for the TS dealing with SWB and FB, it would be appropriate to consider if such parameters may also be implemented in standards of speech terminals, e.g. [i.1], [i.2], [i.3], [i.4] and [i.5].

In audio broadcasting, specific bandwidth enhancements are recognized to improve the audio signal for the listener:

- for wideband and superwideband, the Presence boost, currently between 4 kHz and 6 kHz, ensures vocal clarity and projection); and
- for superwideband the Brilliance boost, currently above 6 kHz, improves audio clarity.

Studies have also indicated that some signal equalization profiles may provide improvements for hearing impaired people, but no standardized values are currently available.

5 Speech/Sound Quality and Intelligiblility

EN301 549 [i.14] refers to "Audio clarity for VoIP". There is no standardized method to qualify "audio clarity", and consequently the EN proposes, as a first step, to assess this parameter in terms of MOS-LQO according to Recommendation ITU-T P.863 [i.10]. However it assesses the listening only quality, not intelligibility.

"Test on overall quality as perceived by high frequency hearing impaired subscribers" [i.29] describes listening quality tests with normal hearing subjects, hearing impaired subjects without hearing aids and hearing impaired subjects with hearing aids. This test was conducted in the context of the EC-funded project HearCom (www.hearcom.eu). It should be noted that the results are only for narrowband transmission, but give a significant set of conclusions and recommendations (e.g. the comfortable listening loudness).

There are very few results available for the time being. However anyone may observe that impacts of transmission impairments on speech intelligibility are more severe for people with hearing losses, but there is no model defining the levels of these impacts nor potential solutions to solve their. Future works are expected to provide a first set of solutions to the following clauses.

[i.34] provides a review of studies about intelligibility, in the context of hearing impaired people.

5.1 Speech intelligibility assessment

ISO TR 22411 [i.18] refers to STI (speech transmission index) which is mainly applicable for room acoustics and is not well adapted within the scope of the present document.

ANSI/ASA S3.5-1997 (R 2012) [i.23] defines a method for computing a physical measure that is highly correlated with the intelligibility of speech as evaluated by speech perception tests given a group of talkers and listeners. The measure is called the Speech Intelligibility Index, or SII. The SII is calculated from acoustical measurements of speech and

noise. This standard is not a substitute for ANSI S3.2-1989 (R 1995) [i.36] American National Standard Method for Measuring the Intelligibility of Speech over Communications Systems.

A few studies have been conducted in the last years to predict intelligibility in the context of speech transmission, but no methodology has been standardized. As an example:

"Objective Speech Intelligibility Measurement on the basis of natural speech in combination with perceptual modeling **[**i.31**]**": Abstract: *"The relation between subjective and objective speech intelligibility measurements is researched. For a large series of speech degradations, noise, linear and nonlinear distortions (speech codecs), intelligibility tests were carried out using short CVC words. In the subjective domain the percentage correctly identified words is taken as the intelligibility score for a certain type of degradation. In the objective domain Recommendation ITU-T P.862 [i.24] is used as the starting point to develop a perceptual model that allows predicting the perceived intelligibility of a speech fragment."*

As Predicting intelligibility is an important research area in room acoustics, studies are also conducted in this field, e.g.:

"Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing" [i.32]: the model described in this publication has been developed for room acoustics and is intended to predict the intelligibility of noisy speech.

"Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions" [i.33]: this paper also consider the prediction of the intelligibility of noisy speech, with stationary and fluctuating noises.

5.2 Impacts of impairments on speech intelligibility

The following impairments and their impacts on speech intelligibility should be investigated with the intention to obtain standardized measurement methods and requirements, for the benefits of both normal hearing users and hearing impaired users:

- Impacts of "network" impairments on speech intelligibility
- Impacts of noise on speech intelligibility
- Impacts of reverberation on speech intelligibility

Another Technical Report is developed within ETSI STQ, specifically dedicated to intelligibility matters. It completes the contents of the present document.

5.3 Other quality parameters

5.3.1 Audio clarity

This criteria should be based on several parameters/indicators, such as intelligibility, quality, noise reduction.

There is currently no standardized definition nor measurement method.

5.3.2 Naturalness

The implementation of SWB coders, as defined in TS 102 924 [i.16], TS 102 925 [i.17] provides the possibility to transmit a bandwidth covering the full speech spectrum, providing the possibility to ensure that the transmitted speech is almost similar to the original speech of the speaker. This Naturalness indicator should be investigated.

Annex A: Considerations about loudness assessment

Loudness ratings determined according to Recommendation ITU-T P.79 [i.6] are computed from the long term spectrum analyzed over about 30 seconds.

In a first step the objective is to make computation of the loudness of signals produced by telephones over this long term spectrum, using several computation models.

In a second step the loudness is computed on the varying signal (speech signal according to Recommendation ITU-T P.501 [i.9]).

The basic concepts of Recommendation ITU-T P.79 [i.6] algorithms are intended to compute the narrowband loudness rating, by analogy with the subjective reference adjusted on the IRS system. Implicitly also, loudness rating is intended to apply to linear systems and for a reference input level. At present, the algorithm has been updated for wideband without considering the impact of non linear and time variant systems and it is not intended to reconsider Recommendation ITU-T P.79 [i.6] for superwideband and fullband, in particular as the concept of loudness rating computation is based on speech signal only and not on other types of audio signals.

The use of loudness rating is a fundamental need for transmission planning, but from the user point of view the loudness of the speech or the sound really perceived by listeners is an important parameter. The user is expecting to listen the signals at a comfortable level, i.e. comfortable loudness, and to have almost similar loudness when commuting different functions (e.g. Handset and handsfree) during the same communication.

Annex A provides results of experiments on narrowband and wideband speech to compute loudness rating and loudness for speech terminals in the receive path.

Annex B provides results for narrowband, wideband, superwideband and fullband audio or speech signals.

Loudness computation models

Based on the article of S. Meunier [i.25], the following loudness computation models have been validated for stationary sounds and standardized:

- **Zwicker model** (first publication in 1958) which lead to an international standard (ISO 532B [i.15]) and a German standard (DIN 45631 [i.13]).
- **Moore and al. model** (published in 1996 with a revision in 1997) which lead to an US standard (ANSI 53.4-2007 [i.12]).

Regarding non stationary sounds, two main models exist to determine loudness but no one has been standardized yet:

- **Zwicker and Fastl model** [i.26] (published in 1999).
- **Model of Glasberg and Moore model** [i.27] (published in 2002).

In Zwicker and Fastl model [i.26] different indicators are recommended to estimate the overall loudness of a sound. They are statistical indexes such as the N7 (recommended for speech signals), the N5 (for ambient noises) or N4 (for traffic).

In Glasberg and Moore model [i.27], it is recommended to calculate the short term maximum loudness level (noted STLmax) to approach the overall loudness level of a sound varying with time.

Measurement process for loudness computation

Analysis chain implemented for this study is shown in figure A.1.

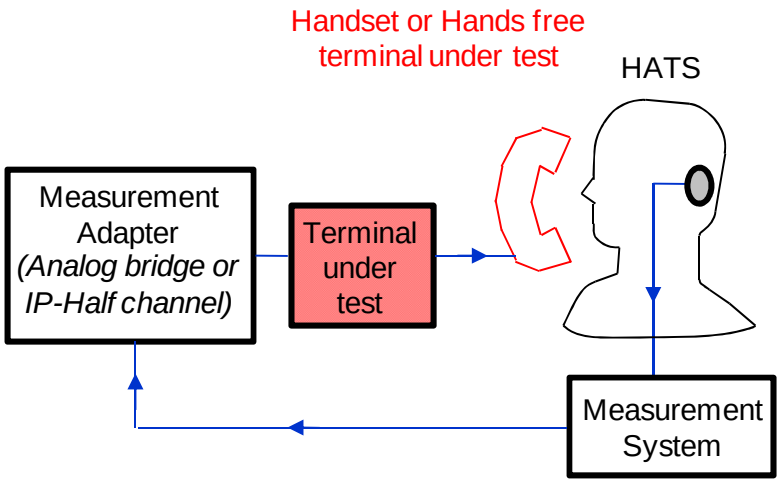

Figure A.1: Diagram of the measurement system used for the study

The receive channel was used for this first experiment. The measurement system sends the test signal to the measurement adapter which is connected to the terminal under test. For VoIP terminals, the measurement adapter is an IP half channel (VoIP reference point) whereas for analogue terminals, the measurement adapter is an analogue bridge (ES 203 038 [i.21] Circuit for measurement of transmission characteristics).

The tests were conducted for both handset or hands-free modes implemented in the terminals under test.

For the calibration, the reference signals are respectively:

- an electrical pure tone at 1 kHz and 50 mV RMS,
- an acoustical pure tone at 1 kHz and 97,1 dB SPL (+3,1 dB Pa),
- a P.501 British-English speech signal at 3 different levels (Nominal, Nominal +5 dB and Nominal -10 dB).

The receive signal is recorded by the artificial ear for both handset and hands-free modes of the telephones. DRP-ERP correction is applied in handset mode and free-field correction is applied in handsfree mode.

Note that the acoustical received signal is measured in dBSPL and in dBA, referenced to 20 µPa and not 1 Pa. The reason to use this acoustical reference instead of dBPa and dBPa(A) is related to the comparison with the phon scale (as the phon scale refers to dBSPL scale).

For the study conducted by Orange Labs, it was chosen to compute (for different speech signals) 4 loudness indicators:

- loudness from Zwicker model (noted **ISO** in the document),
- loudness from Moore and al. model (noted **ANSI**),
- loudness from Zwicker and Fastl model with N7 indicator (noted **N7**),
- loudness from Glasberg and Moore model (noted **STLmax**).

Figure A.2 shows the block diagram of operations performed to obtain the different values of loudness provided by the different algorithms.

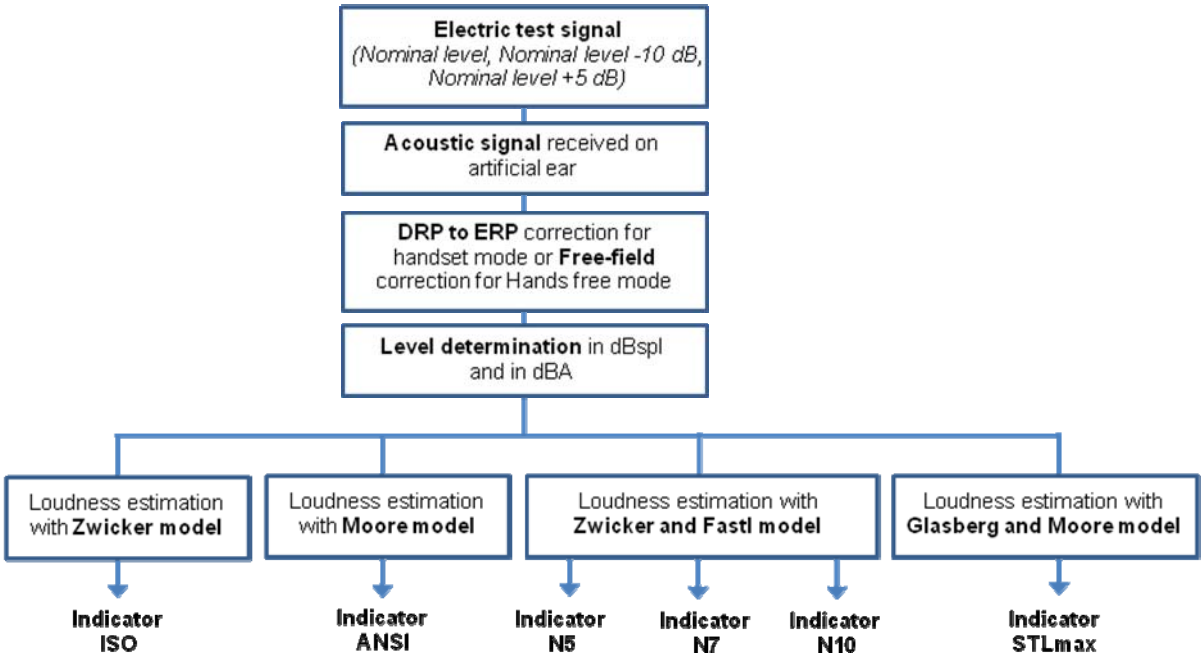

Figure A.2: Block diagram of the operations performed

Summary of the first results

Measurements are performed with a NB / WB terminal using Recommendations ITU-T G.711 [i.19] and G.722 [i.20] codecs, respectively. For this terminal, 12 configurations are used: 2 codecs (G.711 and G.722), 2 modes (Handset and Hands-free), 3 levels for test signal (nominal, nominal +5 dB and nominal -10 dB).

The test signal is Recommendation ITU-T P.50 [i.22] and for each configuration three samples are recorded.

For each of the 36 samples of this first series of measurements, the acoustic level in dB SPL is determined and 6 loudness values corresponding to the 6 initial indicators are derived.

Figure A.3 shows the values of these loudness indicators as a function of the acoustic levels.

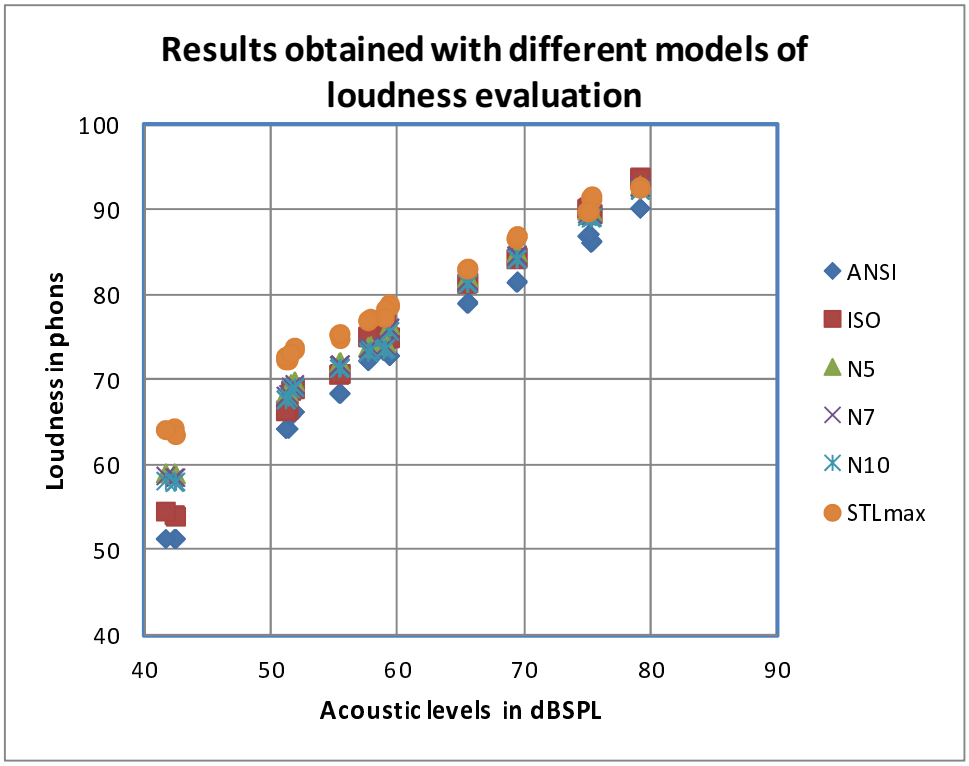

Figure A.3: First results for a device used in Recommendations ITU-T G.711 [i.19] and G.722 [i.20]

From this experiment it can be seen that loudness indicators (for long-term speech-like signal) appear to vary linearly with acoustical level in dB SPL. For the higher acoustical levels the loudness calculated with all the models are rather similar. The differences between models are larger for lower acoustical levels.

Other tests have been conducted, but need more investigation before being published.

Proposed configurations for additional studies

In order to give other laboratories the opportunity to perform similar investigations, the following configurations should be implemented in order to make comparisons of test results:

- several NB terminals using Recommendation ITU-T G.711 [i.19] codec
- several NB/WB terminals using Recommendations ITU-T G.711 [i.19] and G.722 [i.20] codecs
- 2 using modes at receive side: handset and hands-free
- test signal: Recommendation ITU-T P.501 [i.9] (British English speech signal)
- 3 levels for test signals: nominal (-16 dBm0), nominal +5 dB (-11 dBm0), nominal -10 dB (-26 dBm0)
- 1 volume level at the reception for the tested terminal: nominal for handset and maximum for hands-free

For each configuration, the signal is recorded at the receive acoustical output (measured through the artificial ear of the HATS [i.7] and [i.8]). In such conditions it is possible, for each speech sample to determine:

- acoustical level (in dB SPL and dB A)
- RLR (Receive Loudness Rating) associated to the configuration (terminal, codec, using mode, test signal level, test signal)
- 4 loudness indicators (indicators presented above)

Annex B:
Objective and subjective tests: Influence of frequency bandwidth on loudness

B.1 Loudness depending on bandwidth and codec

B.1.1 Simulation process

In this study, all the results are obtained from simulated signals. The simulations are defined in the diagram shown in figure B.1.1.

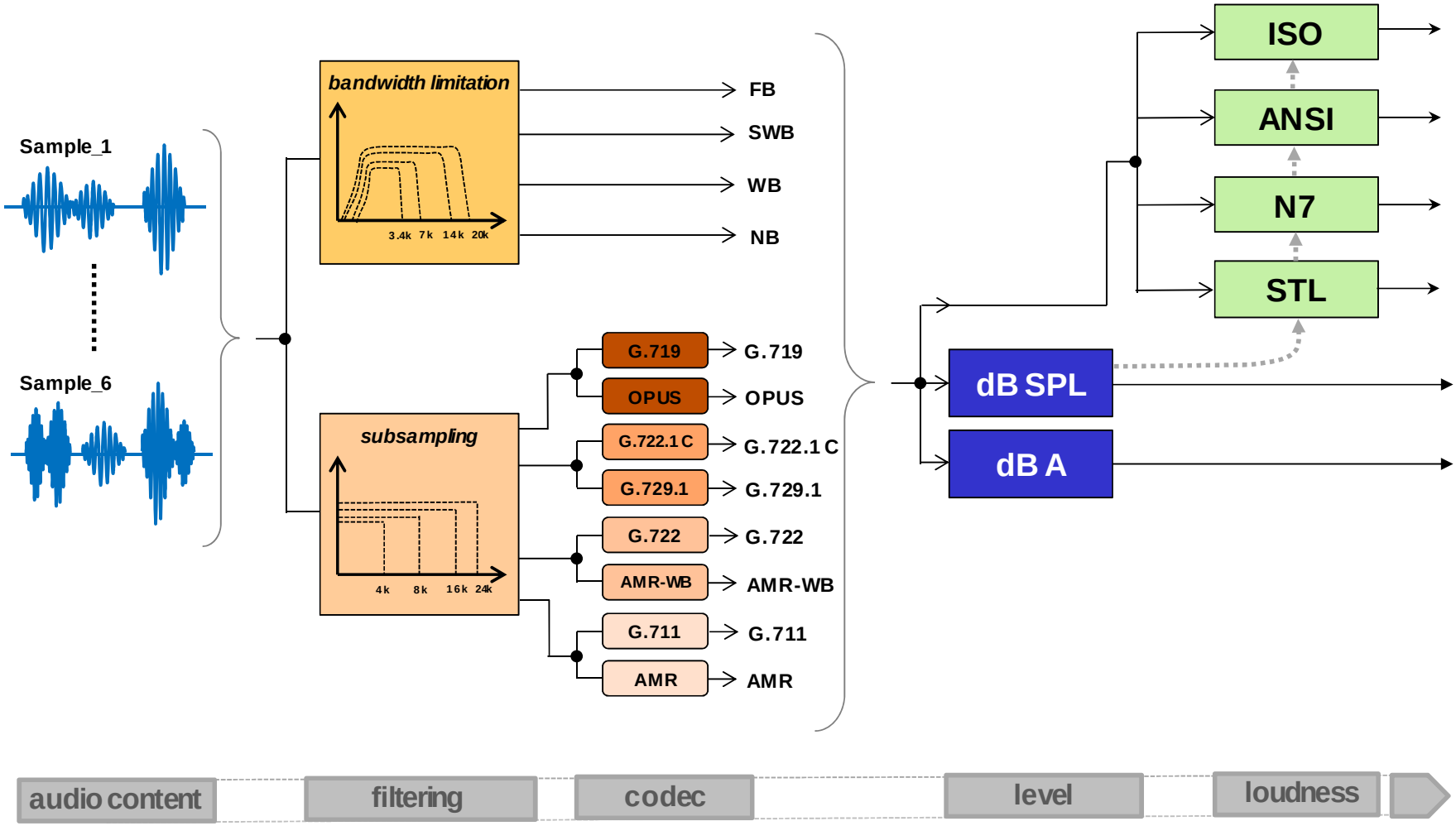

Figure B.1.1: Diagram of the simulation for loudness dependency to bandwidth and codecs

We selected 6 audio inputs with different contents. Hence, the 6 samples can be described as follows:

- Sample 1: Rock Music [7,8 s]
- Sample 2: Music then Speech mixed with Music [12,4 s]
- Sample 3: Speech (voice announcement) [7,6 s]
- Sample 4: Speech mixed with Noise [10,2 s]
- Sample 5: Speech (P.501 British English) [35,4 s]
- Sample 6: Speech then Speech mixed with Music [8,5 s]

The spectrograms of these 6 samples are available in [i.30] in order to illustrate their temporal evolution.

As already mentioned, it was decided to separate the effect of bandwidth limitation (filtering) on loudness from the effect of the codec itself. For instance, for a narrowband codec the resulting bandwidth is generally limited to [300 Hz to 3,4 kHz]. This limitation is due to both transducers and anti-aliasing filters, but it is not due to the codec itself. Thus, if we decide to code/decode an electric signal using G.711 the resulting signal will have a wider frequency range, i.e. [0 Hz to 4 kHz], than what is usually defined for a narrowband signal, i.e. [300 Hz to 3,4 kHz]. That is why the codecs will be applied without any prior filtering (other than the required subsampling).

The 4 bandwidths considered are:

- Full Band (FB) [20 Hz to 20 kHz]
- Super WideBand (SWB) [50 Hz to 14 kHz]
- WideBand (WB) [50 Hz to 7 kHz]
- Narrow Band (NB) [300 Hz to 3,4 kHz] (using flat receive-side modified IRS)

Additionally, we considered 2 codecs per bandwidth recalling that the bandwidth is not limited before coding and decoding, only subsampling is applied.

Thus, the 8 following codecs have been implemented in the experimental set-up:

- FB codecs, sampling rate at 48 kHz → OPUS (64 kb/s) and G.719 [i.39] (64 kb/s)
- SWB codecs, subsampled to 32 kHz → G.729.1 [i.37] (32 kb/s) and G.722.1 annex C [i.38](48 kb/s)
- WB codecs, subsampled to 16 kHz → AMR-WB (12,65 kb/s) and G.722 [i.20] (64 kb/s)
- NB codecs, subsampled to 8 kHz → AMR (12,2 kb/s) and G.711 [i.19] (64 kb/s)

It should be noted that for each sampling rate (codec bandwidth), 2 different families of codec have been chosen. The first family (in green) consists of codecs mainly designed for speech content whereas for the second one (in red), the codecs are not content dependent (they can work with any content as speech or music). These two different families are chosen because the coding is handled differently and it is interesting to know if it has an impact on loudness. In this annex B, the codecs designed for speech are identified as codec group 1, and the other ones are named as codec group 2.

Then, from the signals obtained from each bandwidth and each codec, we calculate their level in dBSPL and dBA (dBSPL with prior A-weighting filter). Finally, using the level in dBSPL as an input for loudness indicators, we calculate the loudness of each signal using the following indicators:

- loudness from Zwicker model (noted **ISO** in the document),
- loudness from Moore and al. model (noted **ANSI**),
- loudness from Zwicker and Fastl model with N7 indicator (noted **N7**),
- loudness from Glasberg and Moore model (noted **STL**).

The selected samples are all non stationary, thus the level in dBSPL and dBA results from an averaging over the full sample duration. On the same way, the calculation of ISO and ANSI indicators is supposed to be done on stationary signal, so the output of these indicators also result from an averaging over the full sample. The averaging is automatically done by considering the full sample as a single frame. For N7 and STL, there is no particular issue as these indicators are created to handle non stationary signals.

This whole simulation can be seen as a simulation of recordings on the receive side using a "perfect" terminal (with transparent frequency response) in handset mode. To do so, all the input samples are aligned to -26 dBoV (over the full [0 Hz to 24 kHz] frequency range) using Recommendation ITU-T P.56 [i.40] and a realistic nominal level on the receive side was simulated corresponding to -16 dBm for the equivalent electric input signal.

The experiment includes a total of 72 conditions corresponding to 6 samples, 4 bandwidths + 8 codecs and 1 receive level: 6 × (4+8) × 1 = 72.

B.1.2 Results presentation

The simulated signals obtained for different bandwidths and codecs are measured in dBSPL and in dBA, referenced to 20 µPa. The reason to use this acoustical reference instead of dBPa and dBPa(A) is the comparison with the phon scale (as a reminder: 0 dBPa equals to 94 dBSPL).

B.1.2.1 Level depending on bandwidth

Figures B.1.2 and B.1.3 respectively represent the level in dBSPL and in dBA depending on the bandwidth available (NB, WB, SWB and FB). Whatever the considered sample, there is a gap between the level in dBSPL in NB case and in the 3 other cases. This implies that on purely energetic point of view, there is a significant gap between NB and WB, but the additional power provided by SWB and FB is very small. But this does not mean that the human ear is not sensitive to these additional frequencies.

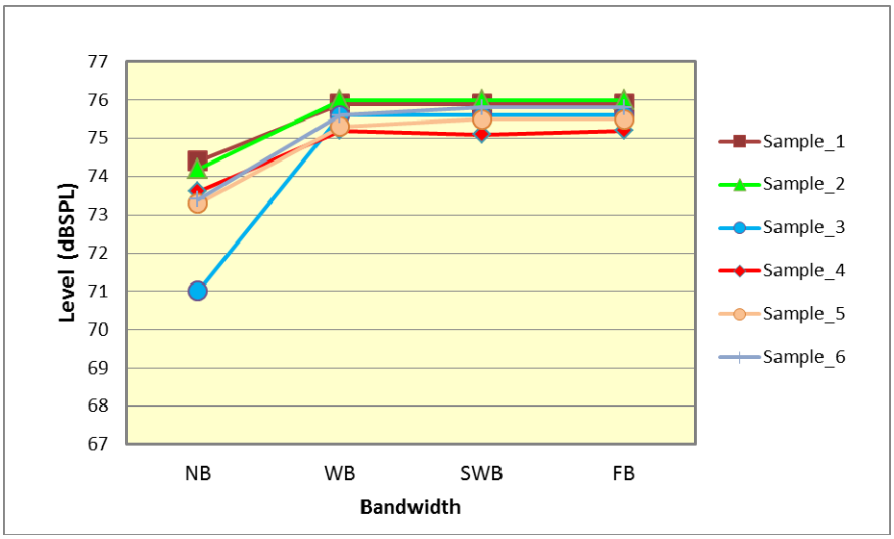

Figure B.1.2: Level in dBSPL depending on the bandwidth

When comparing the levels in dBA for all samples, we can see that the differences between WB and NB cases are much smaller than in dBSPL. This is of course because of the A-weighting ponderation that cuts off low and high frequencies. However, it is interesting to note that in dBA the results for the 6 samples are spread which is not the case in dBSPL. This spread can be explained by the frequency content of each sample. The power spectrum is quite different from a sample to another and each sample will not be affected in the same way by the A-weighting filter.

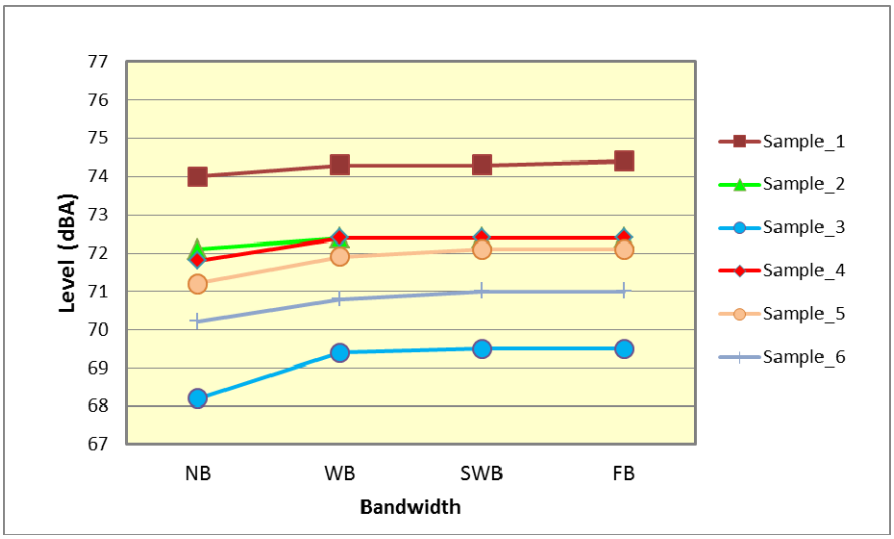

Figure B.1.3: Level in dBA depending on the bandwidth

The power spectrum of the 3 first samples is shown in figure B.1.4 and for the 3 last ones, in figure B.1.5.

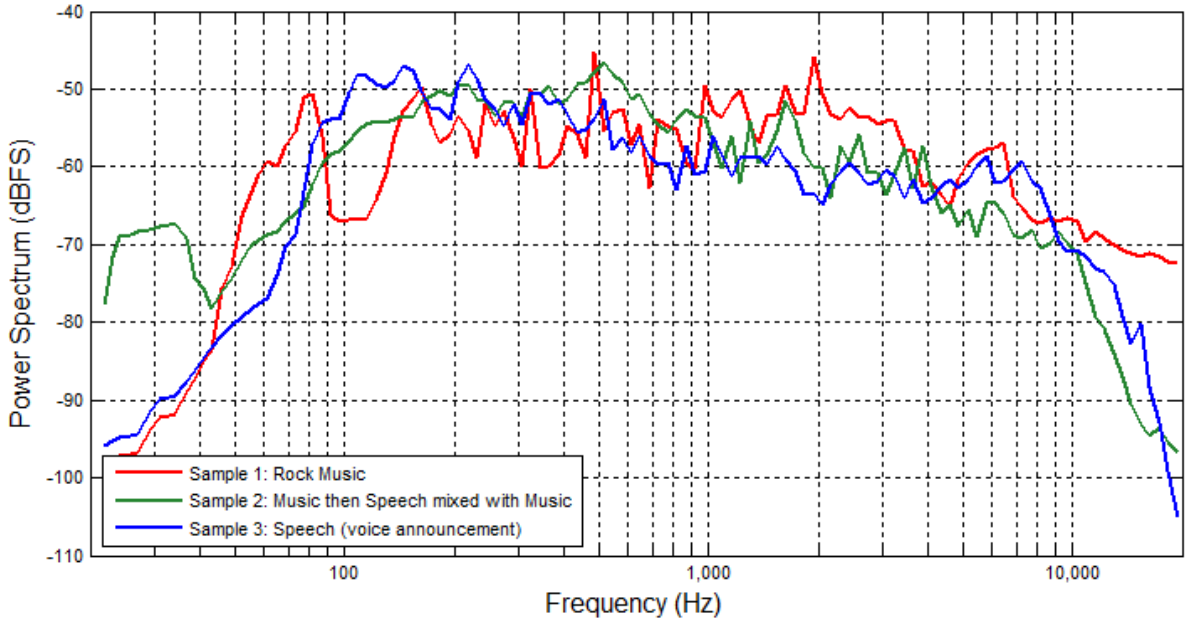

Figure B.1.4: Power spectrum of sample 1 (in red), sample 2 (in green) and sample 3 (in blue)

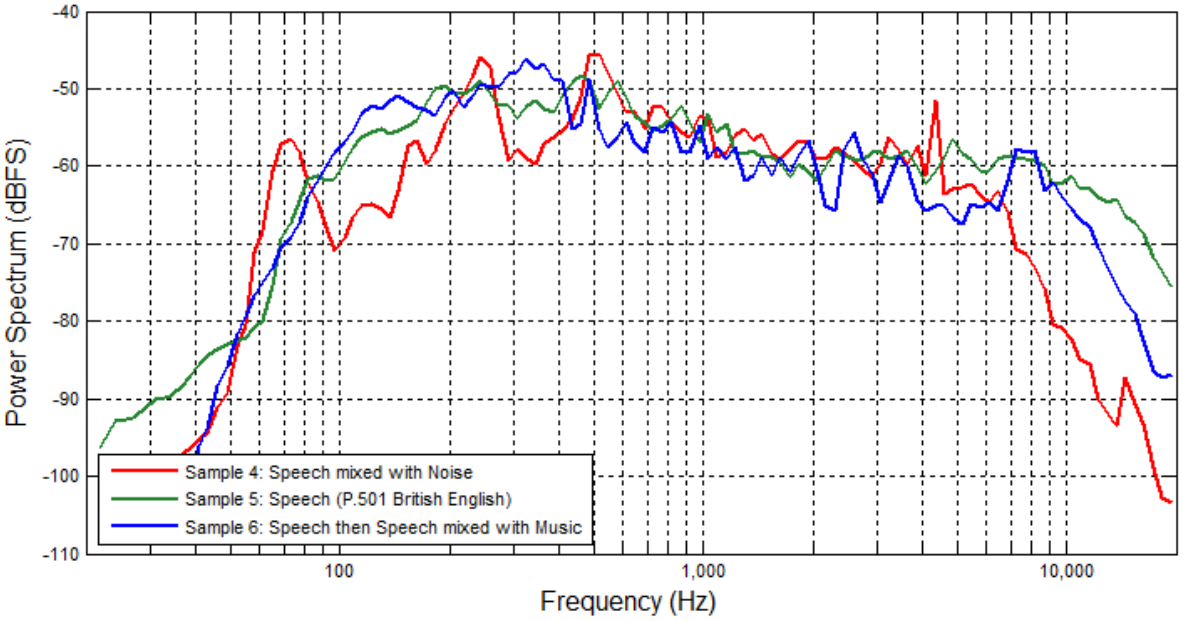

Figure B.1.5: Power spectrum of sample 4 (in red), sample 5 (in green) and sample 6 (in blue)

For instance, the power spectrum of sample 3 (in blue on figure B.1.4) exhibits more power in low frequencies than in high frequencies and it is more affected by the A-weighting ponderation than sample 1 (in red on figure B.1.4) which has a more balanced spectrum.

B.1.2.2 Level depending on codec

Figures B.1.6 and B.1.7 respectively represent the level in dBSPL for codecs in group 1 and in group 2. The level is rather stable for the group 1 of codecs. There is only a small difference of less than 1dB for AMR and AMR WB compared to G.729.1 [i.37] and OPUS. For group 2 of codecs, the level is constant through all the codecs and for all the samples. Thus, it is not observed the same effect as described for the bandwidth. It can be explained by the fact that in this experiment the NB codecs (for instance) operates on the complete [0 kHz to 4 kHz] frequency range, thus AMR and G.711 codecs do not suffer from the bandwidth limitation inherent to NB codecs (due to flat receive-side modified IRS). What is tested here is only the coding/decoding parts; the effect of the filtering is handled separately. Finally, it can be said that the codec has almost no effect on the level calculated in dBSPL, at least from a purely energetic point of view.

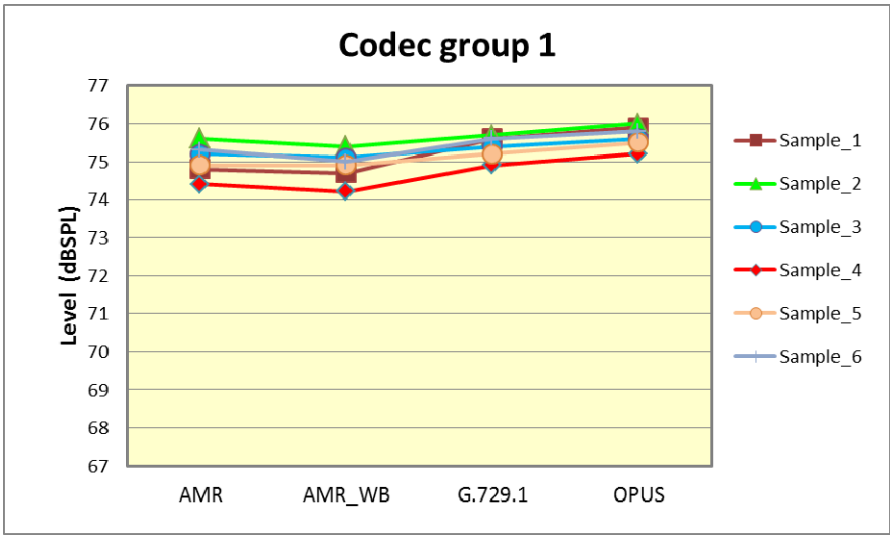

Figure B.1.6: Level in dBSPL depending on the codec from group 1 (speech codecs)

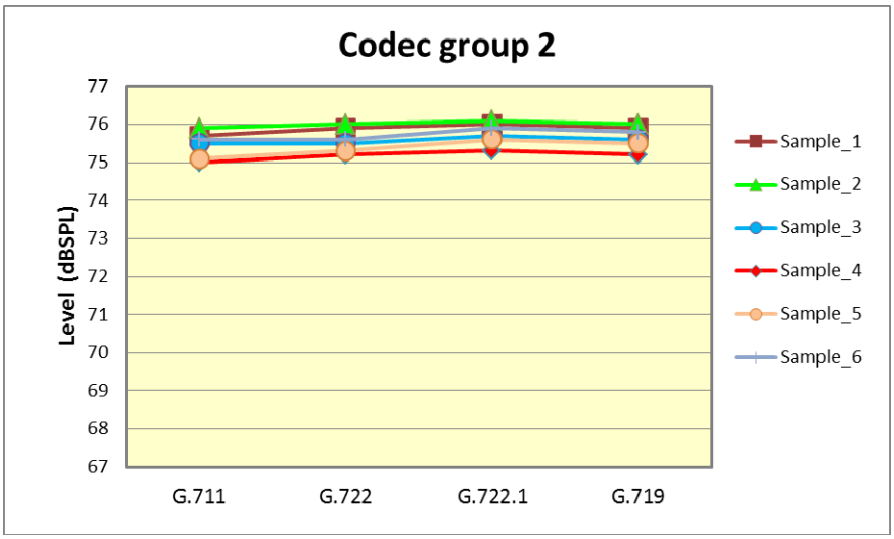

Figure B.1.7: Level in dBSPL depending on the codec from group 2 (not content dependent)

B.1.2.3 Loudness depending on bandwidth

Figure B.1.8 represents the loudness (in Phon) for ISO, ANSI, N7 and STL indicators relatively to bandwidth. For the ISO indicator, a result similar to the level in dBSPL was obtained: there is a gap between the loudness in NB and in the other 3 bandwidths. However, the ANSI indicator has a different behaviour and indicates that the loudness increases with bandwidth. This effect is also present with N7 indicator but is much smaller than for ANSI. However, the STL indicator seems to be insensitive to bandwidth increase (except for sample 4). It is also interesting to note that these 4 indicators react differently to the 6 selected samples. The ANSI and N7 indicators give narrowed results for all samples whereas ISO and STL indicators spread on larger range from one sample to another. It is not possible to infer from these results which indicator is closer to the reality. Thus, subjective tests are needed (see clause B.2) to confront them with the objective test results.

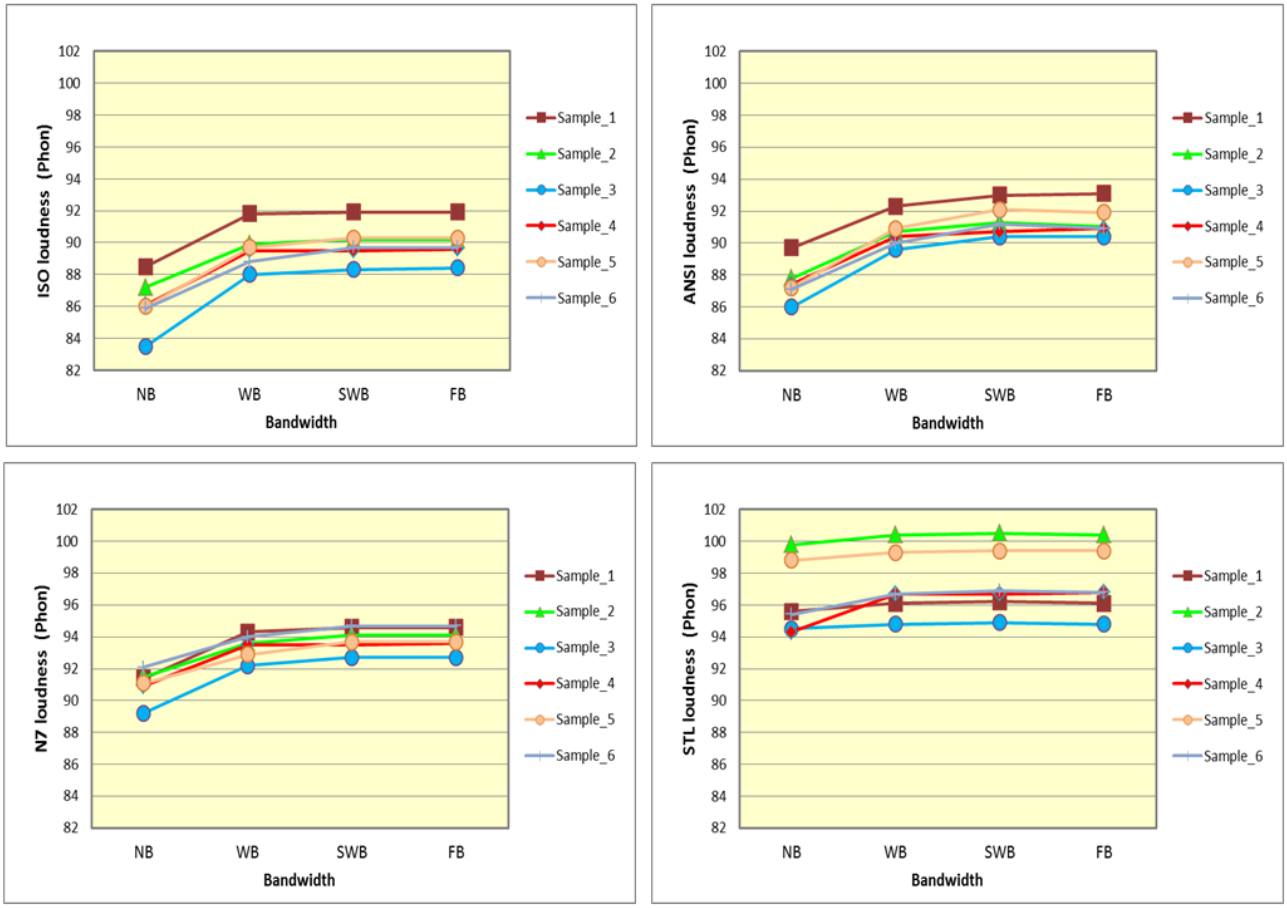

Figure B.1.8: Loudness indicators depending on bandwidth

B.1.2.4 Loudness depending on codec

Figure B.9 represents the loudness (in Phon) for ISO, ANSI, N7 and STL indicators relatively to codecs from group 1. As a reminder, these codecs have been designed mainly for speech. For this group of codecs, results are different from the one in dBSPL. The level in dBSPL is constant through all the codecs and all the samples. However, in terms of loudness a very interesting result is that the loudness (for ISO, ANSI and N7) increases with the increasing of the codec frequency range. This seems to indicate that the user perceives a noticeable difference in level between these codecs even if the energy remains constant. However also in this case the STL indicator seems to be insensitive to codec bandwidth increase (except for sample 4). As in the previous clause, it can also be noted a difference in the indicator dispersion from a sample to another. Again, results are more spread for ISO and STL indicators than for ANSI and N7. Subjective tests will be required to know which behaviour is the best fit to the reality.

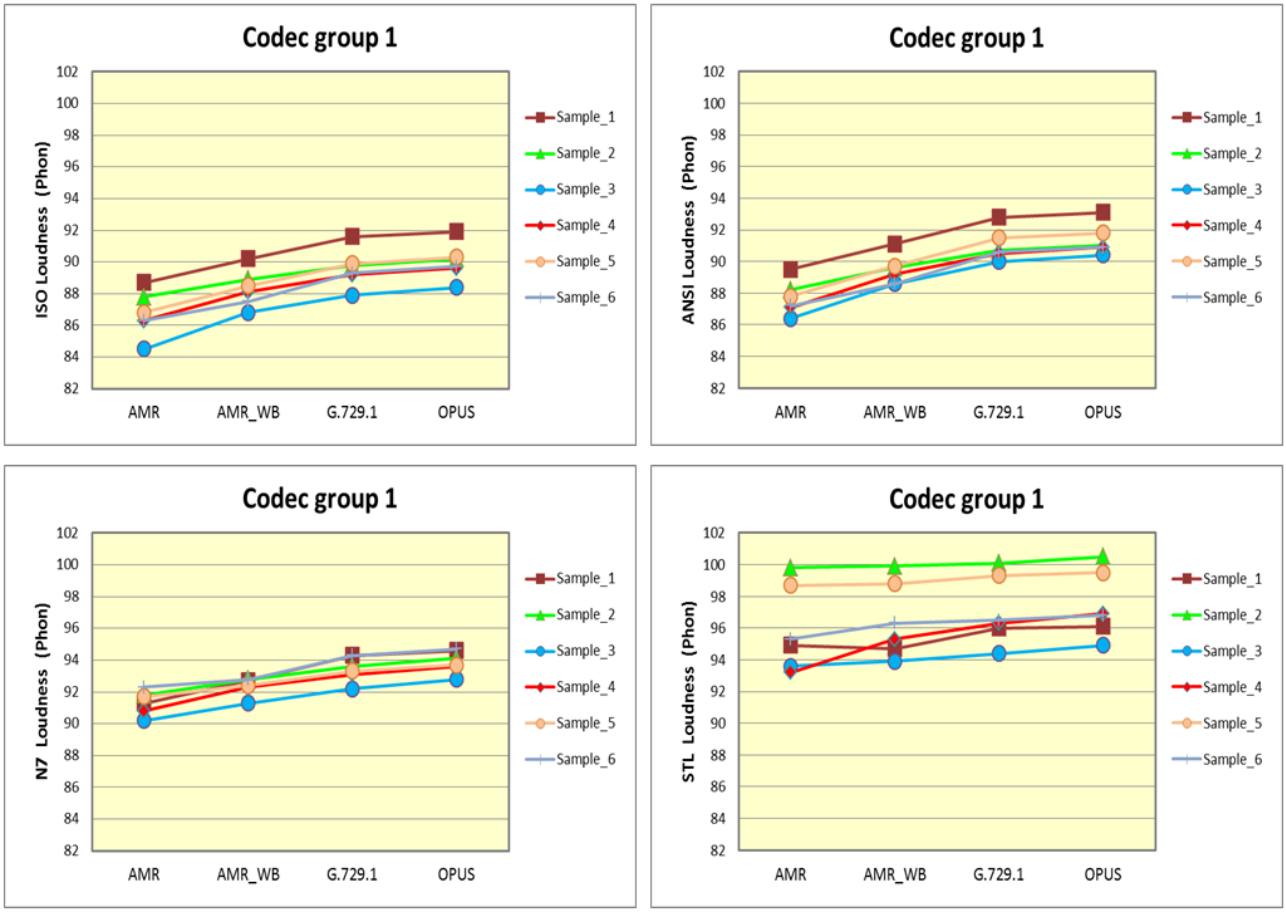

Figure B.1.9: Loudness indicators depending on codecs from group 1 (speech codecs)

Figure B.1.10 represents the loudness (in Phon) for ISO, ANSI, N7 and STL indicators relatively to codecs from group 2. As a reminder, these codecs are not content dependent and can be used to code speech, music or all type of contents. Results for group 2 are sensibly different from group 1, the dependency to codec bandwidth is less marked here. There is a gap between G.711 and G.722 but further increase of frequency range (with G.722.1 annex C [i.38] and G.719 [i.39]) does not bring much difference in level perception. Again in this case, the behaviour of the STL indicator is different as it is insensitive to codec bandwidth increase (except for sample 4). Results for group 2 are then quite close to the ones obtained when increasing the bandwidth (see figure B.1.9). As in the previous clauses it can also be noted that results are more spread for ISO and STL indicators than for ANSI and N7.

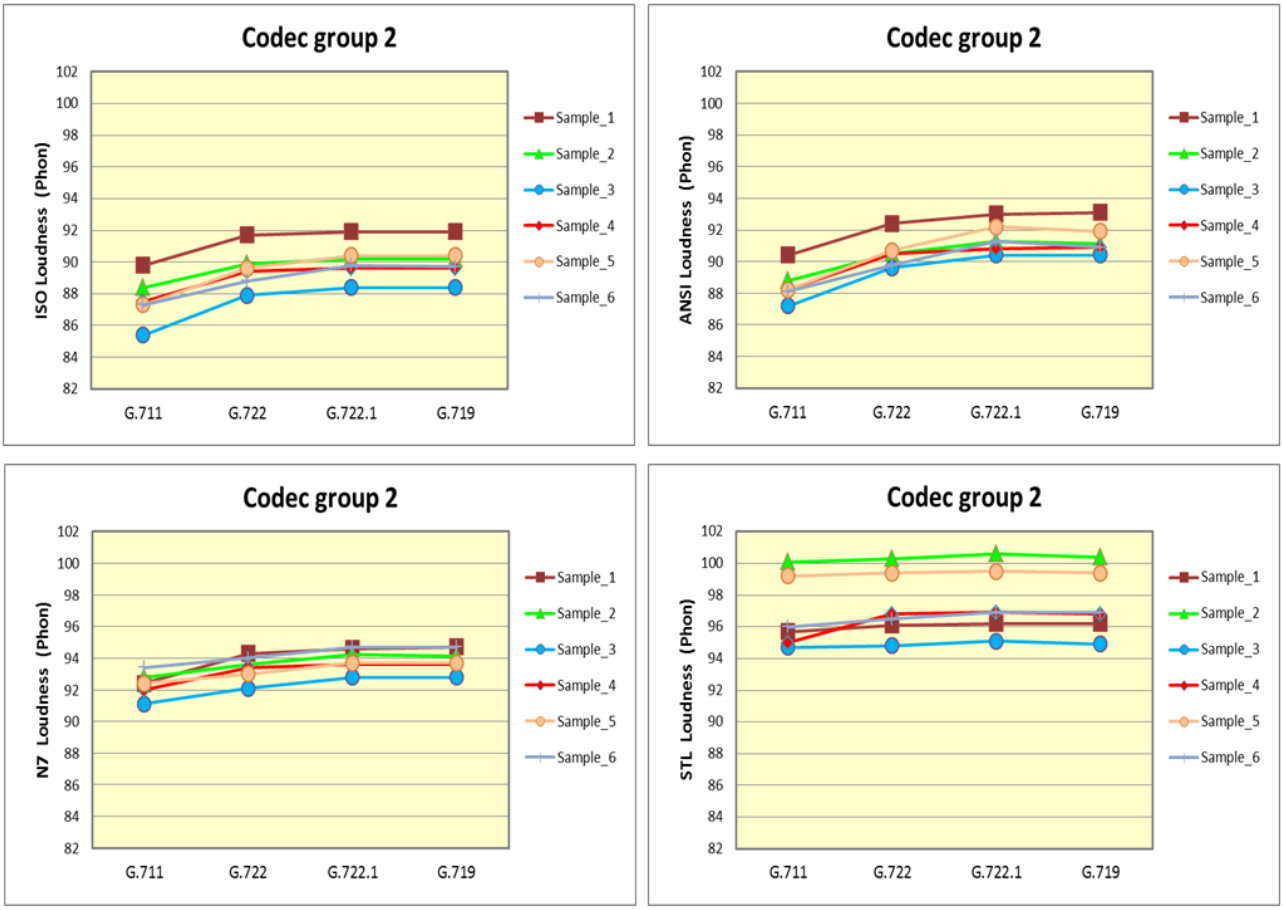

Figure B.1.10: Loudness indicators depending on codecs from group 2 (not content dependent)

B.2 Subjective Test results

B.2.1 Introduction

The goal of this subjective test is to investigate the influence of frequency bandwidth (from narrow band to full band) and the influence of different kinds of codecs on loudness of complex signals such as speech or music.

This subjective test includes two stages. In the first stage, the individual loudness function of each subject is estimated using a critical-band of noise signal. To do so, a special response scale of 100 points is used. In the second stage, each subject evaluates the loudness of the test signals using the same scale. The results are obtained in terms of points and thanks to the estimated individual loudness function it will be possible to convert the point scale to a phon scale. This subjective test will be described in this contribution.

The whole data summarized in this annex are available in TC STQ documents. Additional subjective tests with hearing impaired people should be conducted [i.34].

B.2.2 Selection and preparation of test signals

As our purpose is to investigate the influence of frequency bandwidth as well as the influence of different kinds of codecs on monaural loudness, some audio samples were selected and processed according to the diagram of figure B.2.1.

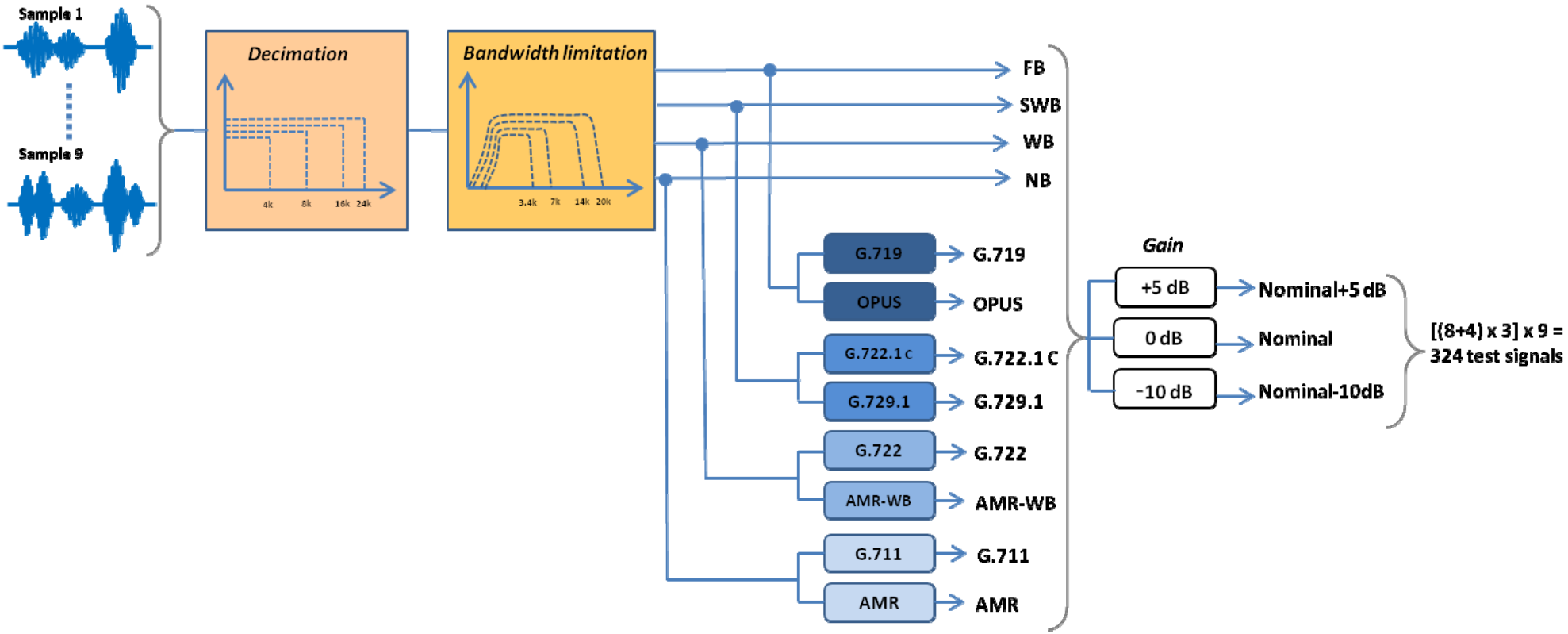

Figure B.2.1: Diagram describing the preparation of test signals for the subjective test

Hence, 9 audio samples with different contents are selected, ranging from speech in different contexts and languages to music. These 9 samples are described in table B.2.1.

Table B.2.1: Description of test signals

	Content description	Duration (seconds)	Speech language	Corresponding samples in annex B.1
Sample 1	Rock Music	7,8	X	Sample 1
Sample 2	Music then Speech mixed with Music	12,4	French	Sample 2
Sample 3	Speech (voice announcement)	7,6	French	Sample 3
Sample 4	Speech mixed with Noise	10,2	French	Sample 4
Sample 5	Speech (P.501) Part 1	8,3	British-English	Part of Sample 5
Sample 6	Speech (P.501) Part 2	9	British-English	Part of Sample 5
Sample 7	Speech (P.501) Part 3	9,2	British-English	Part of Sample 5
Sample 8	Speech (P.501) Part 4	10	British-English	Part of Sample 5
Sample 9	Speech then Speech mixed with Music	8,5	French	Sample 6
NOTE: Since P.501 signal (British-English single talk sequence) is too long (34.5 s) for the subjective tests, it was split into 4 parts; each part containing 3 male or 3 female speakers.				

First, these 9 samples, originally sampled at 48 kHz, are decimated (when required) and filtered out according to the following 4 usual bandwidths:

- Full Band (FB) [20 Hz to 20 kHz]
- Super WideBand (SWB) [50 Hz to 14 kHz]
- WideBand (WB) [50 Hz to 7 kHz]
- Narrow Band (NB) [300 Hz to 3,4 kHz] (using flat receive-side modified IRS)

Then, for each bandwidth, the filtered samples (FB, SWB, WB or NB) were coded/decoded using 2 different families of codecs (see figure B.2.1). The first family consists of codecs mainly designed for speech content whereas for the second one, the codecs are not content dependent. These codecs are described in table B.2.2. These two different families were chosen because the coding is handled differently and it would be interesting to know if it has an impact on loudness. In the rest of the present document, the codecs designed for speech will be referred as "Speech codecs" (named "Group 1" in clause B.1), and the other ones will be referred as "Generic codecs" (named "Group 2" in clause B.1).

Table B.2.2: Description of codecs

Bandwidth	Codec (bitrate)	
	Speech codecs	Generic codecs
FB codecs, sampled at 48 kHz	OPUS (64 kb/s)	G.719 (64 kb/s)
SWB codecs, decimated to 32 kHz	G.729.1 (32 kb/s)	G.722.1 C (48 kb/s)
WB codecs, decimated to 16 kHz	AMR-WB (12,65 kb/s)	G.722 (64 kb/s)
NB codecs, decimated to 8 kHz	AMR (12,2 kb/s)	G.711 (64 kb/s)

The signals directly obtained after filtering or "filtering + coding/decoding" lead to what is defined as the "Nominal" level (Gain at 0 dB in figure 1). An amplification of 5 dB is also applied to these signals which lead to "Nominal +5 dB" level and an attenuation of 10 dB which lead to "Nominal -10 dB" level. These two additional conditions are introduced to test a wider range of hearing levels. Finally, a total of 36 conditions were applied to 9 samples which results in a total of $[(8+4) \times 3] \times 9 = 324$ test signals.

B.2.3 Description of the subjective test

Eighteen **normal-hearing subjects** participated to the loudness subjective test. The subjects are seated in an acoustically treated room. Before the test, each subject is asked to read a set of instructions to understand how the test will be conducted. Each subject is also instructed verbally by the experimenter. During the instructions, the test application software is demonstrated and any questions are answered.

The test procedure includes two stages. In the first stage, the individual loudness function of the subject is estimated using a critical-band of noise (with center frequency at 1 kHz) at different levels. In the second stage, the listener evaluates the loudness of the 324 test signals. All evaluations are made on a specific response scale of 100 points. The results are obtained in terms of points and thanks to the estimated individual loudness function it is possible to convert the point scale into a phon scale.

B.2.3.1 Description of the response scale

After hearing a stimulus, the subject indicates how he/she perceives its loudness using a scale of 100 points that is reproduced in figure B.2.2. After each stimulus presentation, the subject has 5 seconds to provide his/her rating; passing to the next stimulus which was automated in order to push the subject to give a spontaneous evaluation. The subject can see the chosen numeric value displayed on the scale.

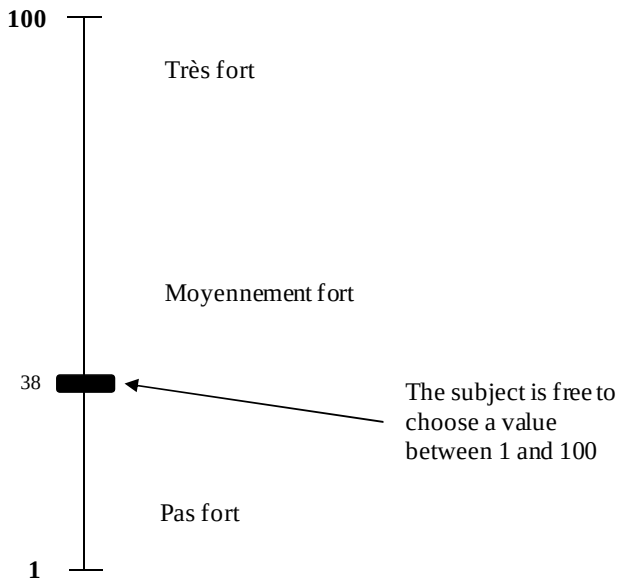

Figure B.2.2: Reproduction of the 100 points response scale

The three labels titled in French "Très fort" (very loud), "Moyennement fort" (averagely loud) and "Pas fort" (not loud) are used to help the subject to have three reference points. These labels are chosen as they are common French language expressions related to loudness. The term "fort" (loud) is used in the three labels since the loudness range covering all test signals is relatively high. This specific range of responses was used in order to give more precision to the subject responses. In fact, if the classical labels found in the categorical loudness scaling [1] had been used, ranging from "not heard" to "extremely loud", it would have been confusing for the subjects because the scale would have been too large compared to the tested range of signals.

NOTE: Throughout the rest of the document, the term "point" is used as a loudness unit for any loudness measured using the presented scale (see figure B.2.2); thus, loudness of all presented stimuli is comprised between 1 and 100 points.

B.2.3.2 Calibration of the sound reproduction chain

Before the beginning of the subjective test, the subject is asked about his/her preferred ear (left or right) when he/she makes a phone call. The test signals are then presented monaurally (left or right) to the subject via high-quality supra-aural headphones. All stimuli are digitally processed at a sampling rate of 48 kHz, D/A-converted and amplified.

The left and right side frequency response of headphones used for the tests are presented in figure B.2.3 in third octave (left side in black and right side in blue). Each curve results of an average of 5 measurements using pink noise.

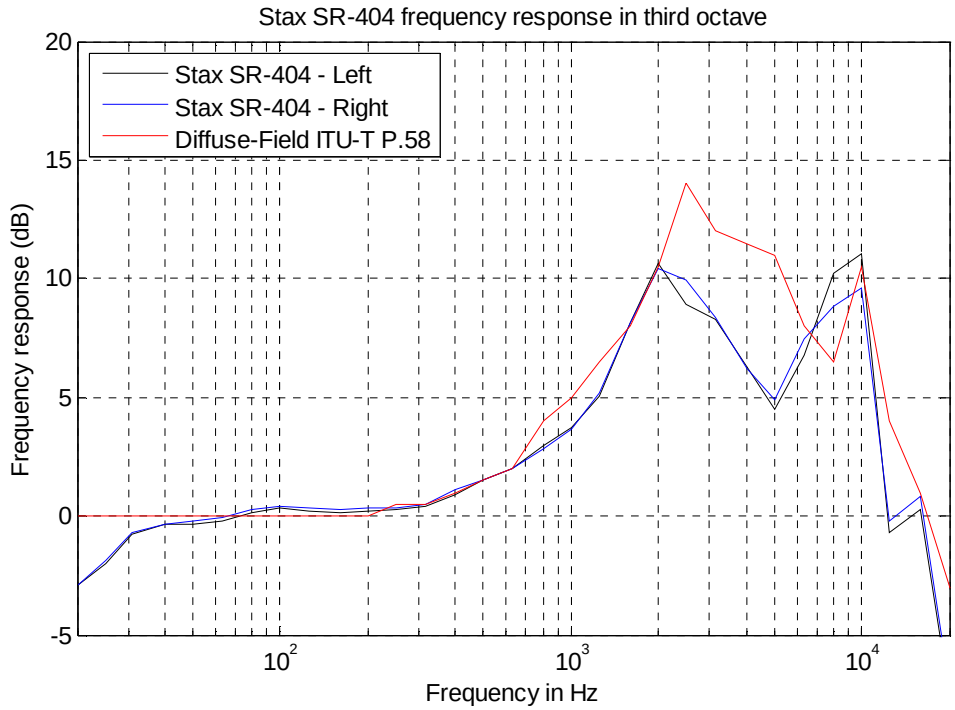

Figure B.2.3: High-quality Headphone Frequency response (left side in black and right side in blue) and Recommendation ITU-T P.58 [i.7] diffuse-field (red curve) in third octave

This headphone is chosen for its neutrality and fidelity over a wide range of frequencies. This seems to be the case as its frequency response is close the Recommendation ITU-T P.58 [i.7] diffuse-field (red curve). As this headphone is close to be diffuse-field calibrated, we decided that no additional equalization was required.

The listening level of the setup is calibrated using a Head And Torso Simulator (HATS), a measurement amplifier and a sound calibrator. It is calibrated to ensure a comfortable level of 77 dB SPL for FB signals at "Nominal" level.

B.2.4 First stage of the subjective test: Measurement of individual loudness function

The individual loudness function describes the relation between the signal level (in dB SPL) and the corresponding loudness (in phons) for each subject. To measure this function, stimuli are presented to the subject at different acoustical levels in a non-systematic way (pseudo-randomized). The stimuli are constructed based on a critical-band (Bark) of noise with center frequency at 1 kHz and duration of 1 second.

The range of presentation levels covered more than the loudness dynamic range of test signals (*i.e.* test signals that will be used in the second stage of the test, see figure B.2.1). Previous to this, a small test had been designed to determine this dynamic range.

B.2.4.1 Dynamic range determination

The determination of the dynamic range consists in making a loudness-balance test; which determines the sound levels at which a test signal and a comparison stimulus appear equally loud. In the specific case, the test stimuli are critical-band of noise (centred on 1 kHz) presented at different levels.

Over all test signals (see figure B.2.1), the ones with higher level in dB SPL come from the condition "FB and Nominal +5 dB" and the ones with lower levels come from the condition "NB and Nominal -10 dB". All these signals were tested in order to determine the maximum and the minimum of the dynamic range.

The critical-band of noise is presented in a large range of levels from 58 dB SPL to 91 dB SPL with a step of 3 dB. The subject has to select the stimulus that is as loud as the selected test signal as illustrated in figure B.2.4.

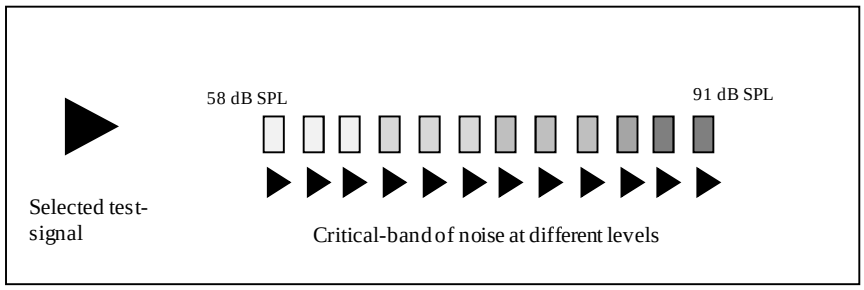

Figure B.2.4: The subject chooses the stimulus that is as loud as the selected test signal

At the end of this test, it was found that, in average, the test signals coming from condition "FB and Nominal +5 dB" correspond to a maximum of **85 dB SPL** and the test signals coming from condition "NB and Nominal -10 dB" correspond to minimum of **73 dB SPL**. In order to be sure that the full dynamic range was covered, it was decided to choose a larger dynamic range, i.e. [**61 dB SPL; 88 dB SPL**]. Therefore, the stimuli used for the determination of individual loudness function consisted of 10 critical-band of noise ranging from 61 to 88 dB SPL.

NOTE: This test was done before the start of the actual subjective test on loudness. It was conducted on ten colleagues working in the laboratory. For the actual loudness subjective test, the individual loudness function is measured for each subject over the pre-determined dynamic range.

B.2.4.2 Measurement of individual loudness function

The assessment of individual loudness function consists in two phases in which the subject rates the loudness using the scale described in figure B.2.2. The first phase is the training phase in which the subject hears a selection of samples covering the whole dynamic range of levels. This phase avoids biases caused by the first trials that do not cover the whole dynamic range. During the training phase, 4 stimuli are presented, one stimulus with the highest level, another with the lowest level and two stimuli with intermediate levels.

In the second phase, the 10 stimuli (critical-band of noise presented at different levels are presented 6 times each, using 6 pseudo-random orders. Attention is paid to keep the level difference between two successive stimuli not too high (smaller than half of the dynamic range).In such way, the context effects due to the tendency of many subjects to rate the current stimulus relatively to the previous one are reduced. All 64 trials (training plus 6 pseudo-random orders) are represented in figure B.2.5.

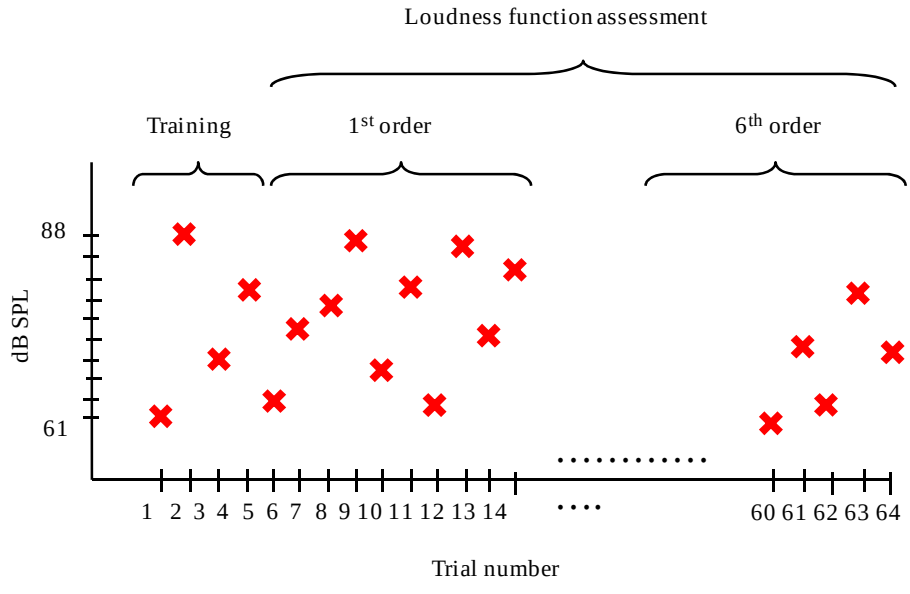

Figure B.2.5: Trials for the determination of individual loudness function

NOTE 1: All subjects hear the 64 trials in the same order.

NOTE 2: The assessment of the individual loudness function lasts about 8 minutes after which the subject is asked to take a break of around 3 minutes.

B.2.4.3 Results for individual loudness functions

Figure B.2.6 shows the individual loudness functions of the 18 subjects in term of points. The overall average is also displayed in dashed line. It can be observed that in general the curves are shaped like an "S" because of two saturation parts: the upper part [85 dB SPL to 88 dB SPL] and the lower part [61 dB SPL to 70 dB SPL]. These saturation parts are due to a saturation of the scale. In fact, the subjects always judge the sound as "very loud" when the signal level is higher than 85 dB SPL, and as "not loud" when the signal level is lower than 70 dB SPL. The interesting part is the middle part [70 dB SPL to 85 dB SPL] that is linear. In this linear part the responses of subjects are proportional to the presented acoustic level in dB SPL. Thus, in this range, i.e. [70 dB SPL to 85 dB SPL] the response scale (see figure B.2.2) is used efficiently.

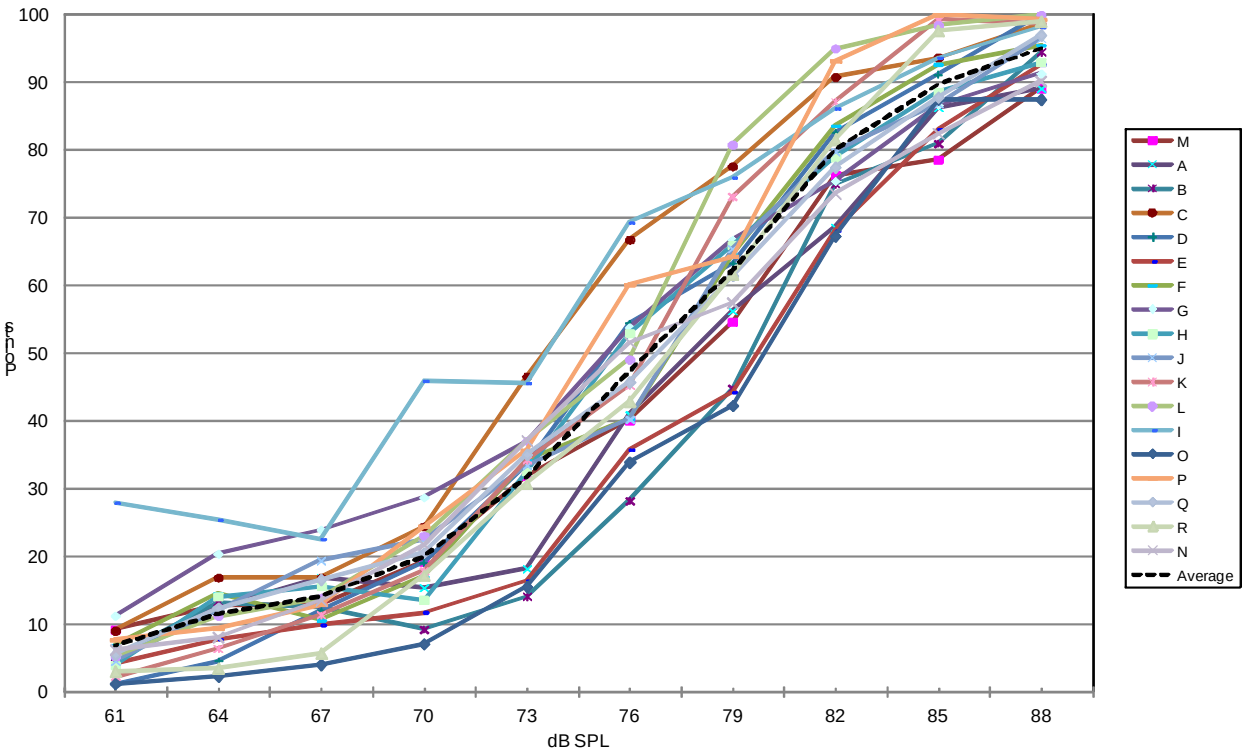

Figure B.2.6: Individual loudness functions (in term of points) obtained for the 18 subjects along with the overall average (dashed line)

Based on the linear part of these individual loudness functions, the results obtained (in term of points) in the second stage of the subjective test can be converted into phons. This is described in clause B.2.5.

B.2.5 Second stage of the subjective test: Assessment of test signal loudness

B.2.5.1 Assessment of test signal loudness

The test signal loudness assessment is composed by two phases in which the subject rates the loudness using the scale described in figure B.2.2. The first phase is the training phase in which the subject hears a selection of samples covering the whole dynamic range of levels. It also covers a wide range of conditions as sum up in table B.2.3. This selection contains the softest and loudest conditions (coloured box in table B.2.3). All 9 samples are used in the training so that the subject discovers them all before the second phase.

Table B.2.3: Test signals used for training phase

Sample	Condition
Sample 1	0 dB and SWB
Sample 2	-10 dB and G.729.1
Sample 3	0 dB and AMRWB
Sample 4	+5 dB and G.711
Sample 5	-10 dB and OPUS
Sample 6	-10 dB and AMR
Sample 7	+5 dB and G.722.1C
Sample 8	0 dB and G.722
Sample 9	+5 dB and FB

In the second phase, the 324 test signals (see figure B.2.1) are presented randomly. To do so, 6 random orders were created. Thus, each order is used for 3 subjects as sum up in table B.2.4. For the assessment of these test signals (including training), the subjects are asked to take into account the perceived level averaged over the full signal as they are relatively long (see table B.2.1).

Table B.2.4: Random order distribution

Subjects	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Random order number	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

At the end of this test, we obtained for each subject the loudness assessment for the 324 test signals in term of points. In the next clause it will be detailed how to transform the points into phons using the individual loudness functions.

NOTE: The assessment of the test signal loudness takes about 2 hours for each subject. A break of 3 minutes is requested after each 36 evaluations. The total duration of the subjective test is around 2,5 hours.

B.2.5.2 Conversion from points to phons

The estimated individual loudness function gives the relation between dB SPL and points for each subject (see figure B.2.6). The key to transform points to phons is that the phon scale is equal to dB SPL scale for a critical-band of noise with center frequency at 1 kHz. Thus, it is possible from estimated individual loudness function to infer the relation between points and phons. To do so, the dB SPL scale is simply replaced by the phon scale in figure B.2.6.

This relation is discrete as dB SPL (and then phon) scale is defined with a 3 dB step. In order to convert points to phons an interpolation is necessary. As a reminder, the individual loudness function is linear in the range [70 dB SPL to 85 dB SPL]. This is also the range that is of interest for these tests because all test signals are comprised in this range. Because the tests are performed in the linear part of the individual loudness function, it was decided to use a linear regression as fitting model. This is illustrated in figure B.2.7 with subject "B" for instance.

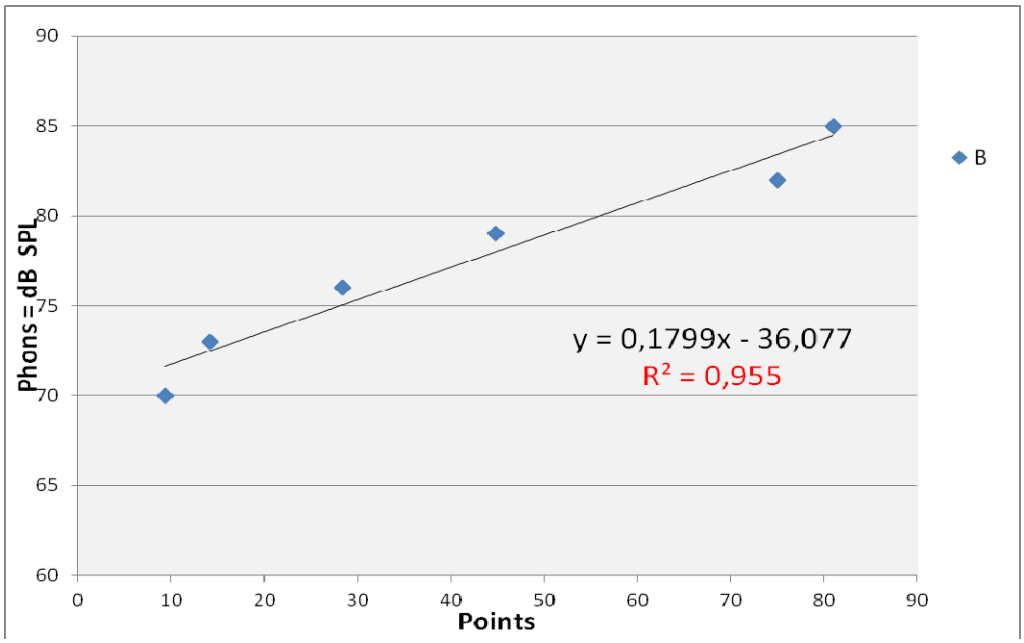

Figure B.2.7: Linear regression for point to phon conversion of subject "B"

Finally, all the results obtained in points for the assessment of test signal loudness can be converted to phons using this kind of fitting equation. For instance, for the subject "B", the following equation is used:

$$y(\text{phons}) = 0{,}1799 \times x(\text{points}) + 69{,}923$$

The point to phon conversion for each subject is based on his/her own individual loudness function, the reason being that each subject uses the response scale in his/her own way, the subject creates for him/herself an internal reference system which can vary largely from a subject to another. However, as long as the subject keeps the same internal reference system through the entire subjective test, the points can be converted to phons using his/her own individual loudness function.

B.2.6 Results for test signal loudness

Loudness results presented below are averaged over all 18 subjects. They are presented in term of Phons. The bandwidth and codec conditions (*see* figure B.2.1) are divided into three groups:

- "Bandwidth" including NB, WB, SWB and FB conditions,
- "Speech codecs" including AMR, AMR-WB, G.729.1 [i.37] and OPUS conditions,
- "Generic codecs" including G.711, G.722, G.722.1 annex C [i.38] and G.719 [i.39] conditions.

Loudness results are also presented for the three defined different levels, i.e. "Nominal +5 dB", "Nominal" and "Nominal -10 dB".

B.2.6.1 Results averaged over all samples

Figure B.2.8 gathers loudness results averaged over all samples. All conditions are represented in figure B.2.8, i.e. "Bandwidth", "Speech codecs", "Generic codecs" as well as the three levels, i.e. "Nominal +5 dB", "Nominal" and "Nominal -10 dB". These results are presented in term of Phons and come with confidence interval at 95 %. These results are consistent with what could be expected as loudness increases with bandwidth extension (after coding/decoding or not). There is a statistically significant gap between loudness in NB and WB conditions (including codec conditions) and for all levels. There is also a gap between WB and SWB conditions, but it is smaller and not significant from a statistical point of view. However this gap is a tendency that is found for all levels and hence should not be ignored. Finally, the gap is very small between SWB and FB conditions and is not significant. It can be also noted that results for "Bandwidth", "Speech codecs" and "Generic codecs" are rather similar.

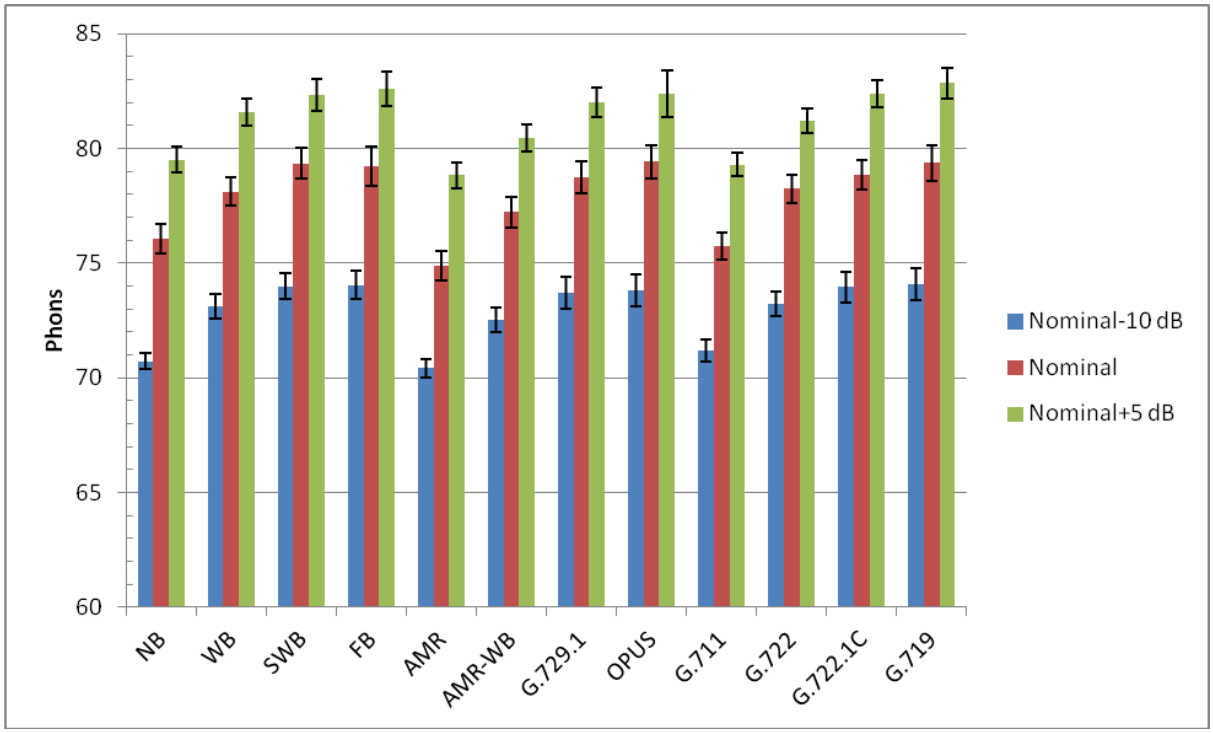

Figure B.2.8: Averaged results over all samples. All conditions.

It is interesting to have a closer look at loudness differences when switching from a frequency bandwidth to a higher one, e.*g.* when switching from NB to WB or, if there is coding/decoding, from G.711 to G.722. These results are gathered in table B.2.5.

Table B.2.5: Average loudness differences when switching from a frequency bandwidth to a higher one

	Bandwidth			Speech codecs			Generic codecs		
	NB →WB	WB →SWB	SWB →FB	AMR → AMR_WB	AMR_WB →G.729.1	G.729.1 → OPUS	G.711 → G.722	G.722 → G.722.1 C	G.722.1 C →G.719
Nominal +5 dB (phons)	2,39	0,87	0,05	2,09	1,19	0,12	2,02	0,73	0,15
Nominal (phons)	2,06	1,23	-0,14	2,33	1,53	0,66	2,49	0,63	0,49
Nominal -10 dB (phons)	2,07	0,75	0,26	1,62	1,53	0,40	1,91	1,18	0,45

B.2.6.2 Detailed results per sample

Figures B.2.9, B.2.10 and B.2.11 gather detailed loudness results obtained per sample. Figure B.2.9 corresponds to "Nominal +5 dB" condition, figure B.2.10 to "Nominal" and figure B.2.11 to "Nominal -10 dB". Results for samples 5, 6, 7 and 8 are averaged as they all correspond to P.501 signal (that have been cut into 4 parts for the subjective test). All the results are presented with the same loudness level range, i.e. 68 to 84 Phons, in order to make comparisons easier.

Results for "Nominal + 5 dB"

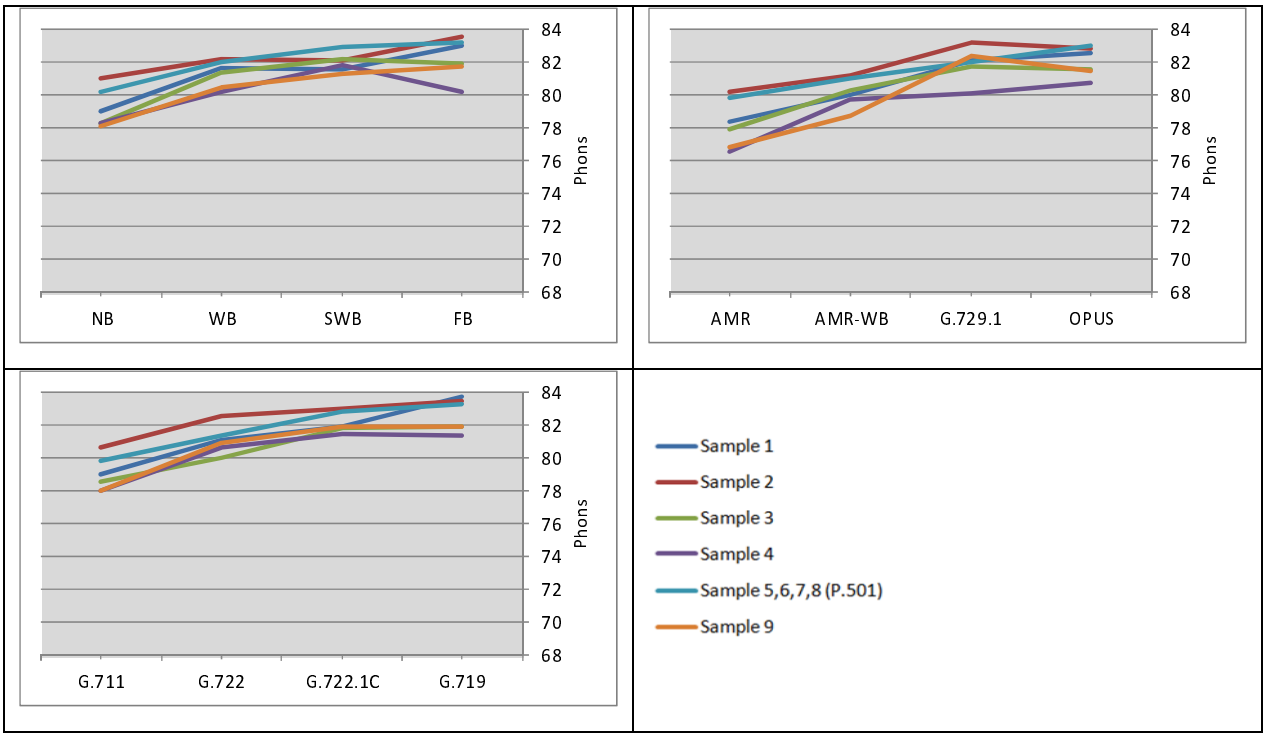

Figure B.2.9: Averaged results per sample for "Nominal +5 dB" condition. Results for samples 5, 6, 7 and 8 are averaged as they all correspond to P.501 signal.

Results for "Nominal"

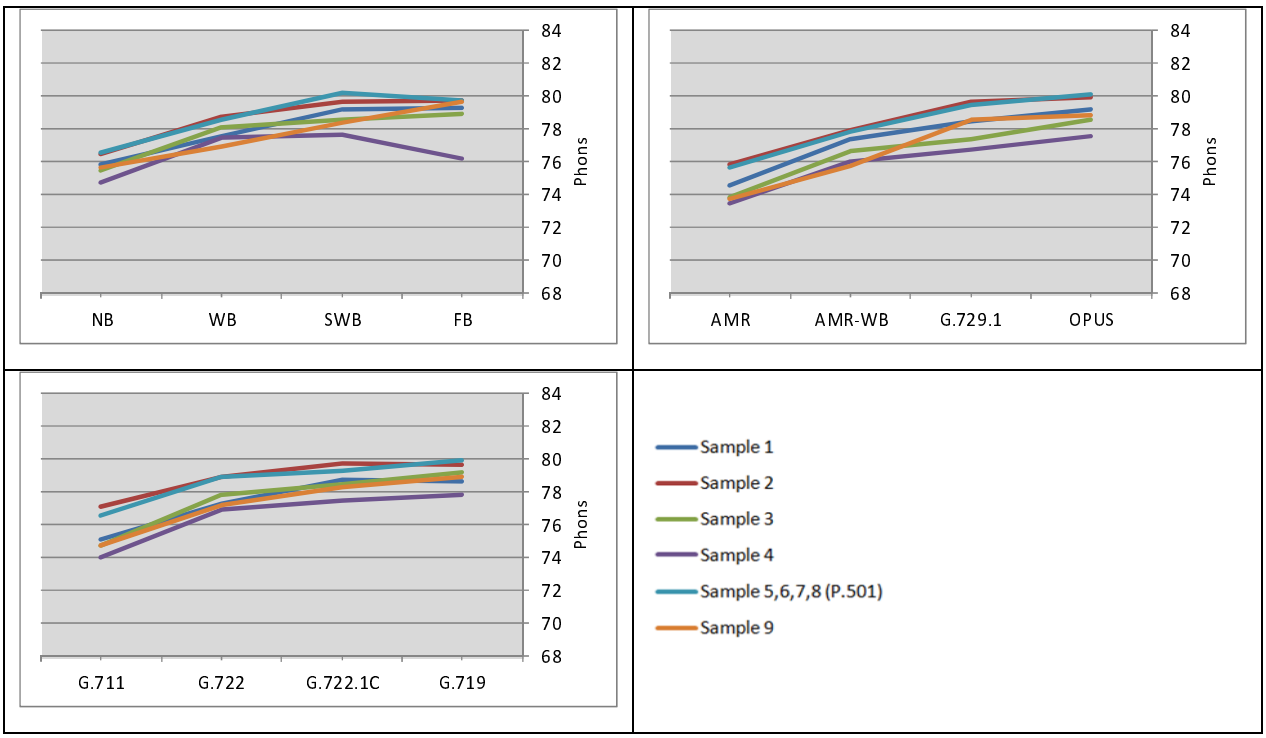

Figure B.2.10: Averaged results per sample for "Nominal" condition. Results for samples 5, 6, 7 and 8 are averaged as they all correspond to P.501 signal.

Results for "Nominal -10 dB"

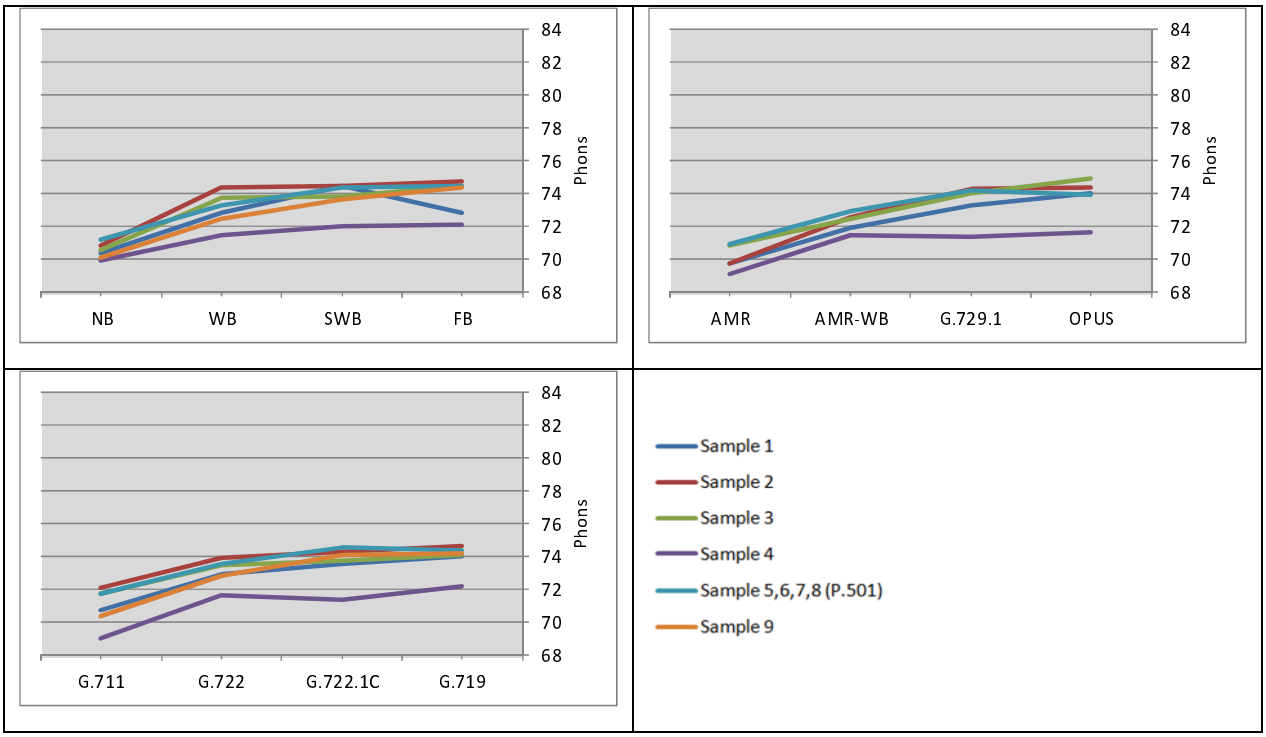

Figure B.2.11: Averaged results per sample for "Nominal -10 dB" condition. Results for samples 5, 6, 7 and 8 are averaged as they all correspond to P.501 signal.

As we can see in these three figures (B.2.9, B.2.10 and B.2.11), conclusions that can be drawn per sample are mainly consistent with conclusions drawn for averaged results. For the three defined levels ("Nominal +5 dB", "Nominal" and "Nominal -10 dB") loudness increases with frequency bandwidth extension. However, results for sample 4 are different as perceived level tends to decrease in some cases when switching from SWB to FB. The perceived level for Sample 4 is also noticeably lower than for the other samples, in particular for "Nominal -10 dB" condition. Sample 4 contains speech mixed with background noise and it is the only noisy sample. Probably noise has an influence on perceived loudness, may be because of noise masking effect; this behaviour should be checked in a future subjective test. However, other samples have consistent behaviour even though language and content are different (French, British-English, Music, and mixed contents) which is an encouraging result.

B.2.6.3 Results averaged over all samples, except Sample 4

As Sample 4 seems to have unexpected behaviour (probably because of background noise), the average over all samples was recalculated, but excluding Sample 4. Figure B.2.12 gathers all these loudness results in all conditions, i.e. "Bandwidth", "Speech codecs", "Generic codecs" as well as "Nominal +5 dB", "Nominal" and "Nominal -10 dB". The results and the associated conclusions are very similar to those obtained in figure B.2.8, however they are more accurate as Sample 4 introduces a bias in averaged results. In fact, confidence intervals slightly decreased in 34 cases over 36. As averaged and confidence intervals are modified when excluding Sample 4, it appears that the gap between WB and SWB conditions (including coding/decoding) becomes statistically significant in almost all conditions (i.e. 7 conditions out of 9).

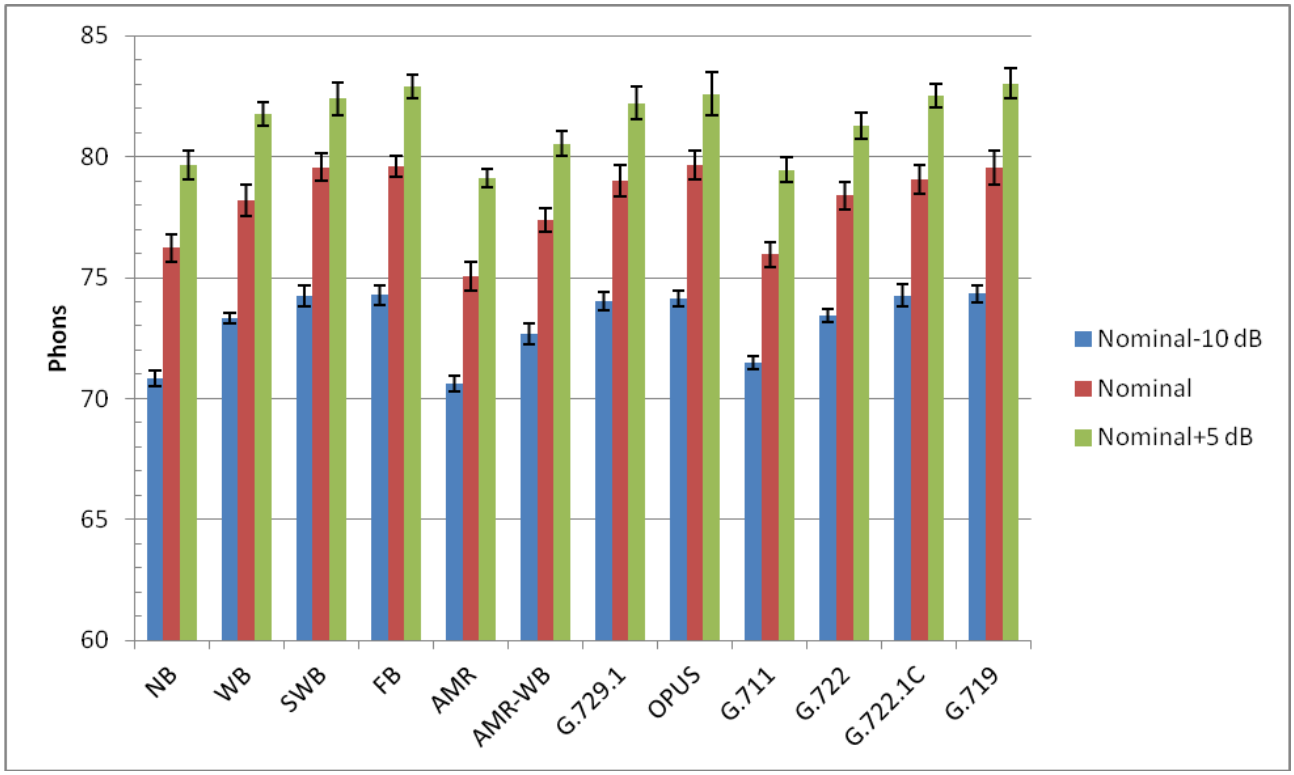

Figure B.2.12: Averaged results over all samples, except Sample 4. All conditions.

These results are gathered in table B.2.6 where Sample 4 was omitted to compute the averages.

Table B.2.6: Average loudness differences, except Sample 4, when switching from a frequency bandwidth to a higher one

	Bandwidth			Speech codecs			Generic codecs		
	NB →WB	WB →SWB	SWB →FB	AMR → AMR_WB	AMR_WB →G.729.1	G.729.1 → OPUS	G.711 → G.722	G.722 → G.722.1 C	G.722.1 C → G.719
Nominal +5 dB (phons)	2,50	0,91	0,05	2,06	1,37	0,10	1,95	0,85	0,06
Nominal (phons)	1,97	1,37	0,03	2,31	1,63	0,64	2,43	0,65	0,50
Nominal -10 dB (phons)	2,08	0,64	0,50	1,43	1,67	0,37	1,83	1,22	0,51

Annex C: Bibliography

Jean-Yves Le Saout, Jean-Yves Monfort. Proposed study on loudness assessment. ETSI STQ(12)39_019r1

Jean-Yves Le Saout, Cyril Plapous, Jean-Yves Monfort. Comparison between loudness ratings and loudness. ETSI STQ (12)40_26r1

Cyril Plapous. Subjective test: Influence of frequency bandwidth on loudness. ETSI STQ(13)44_026.

History

Document history		
V1.1.1	September 2014	Publication