Aalto University
School of Electrical
Engineering

Communication acoustics
Ch 17: Sound quality

Ville Pulkki and Matti Karjalainen

Department of Signal Processing and Acoustics
Aalto University, Finland

November 22, 2016

Sound quality

We want to *evaluate*: good quality is more desirable, valuable, positive, appealing, and useful to us

- In technical sciences, the quality of products is of interest in R/D
- How can we know the quality of products producing sound?
- Audio / HiFi
- Concert halls
- Noise annoyance
- Speech intelligibility / Public address
- Hearing aids / cochlear implants
- Razors / Car engine sounds / Mechanical products

Two meanings of term ’quality’

- Quantity vs Quality
  - Categorization by type or class of objects
  - When two observations or entities cannot be compared on the same (metric) scale they are said to be qualitatively different
- Quality as synonym for ’excellence’
  - To grade or rank objects on a subjective scale of preferability such as ‘good–poor’,
  - This definition is used here

Sound quality and psychoacoustics

- Classical psychoacoustics sees human as simple metering device
- Basic results are of course very useful
- In sound quality the expectations, mood, and other cognitive factors are important
- Judged sound quality for devices or communication channels depends heavily on expectations
- Basic psychoacoustic experimentation techniques may not be used

History of sound quality

- Evolution of musical instruments and concert halls
- 1920's: serious development of telephone technologies and sound quality therein
- 1950's–: HiFi hobby
- 1980's–: audio coding, applying human frequency resolution in audio coding, related sound quality studies
- 2000's: spatial aspects of sound

Evaluation and measurement of sound quality

- Sound quality is a fundamentally subjective (perceptual) concept but it can be approximated by objective and computational criteria
- Subjective quality can be evaluated by listening experiments, for example:
  - Compare to 'perfect quality' reference to find out if any degradation can be noticed
  - Compare two or more sounds and sort then by quality preference
  - Characterize sound quality by conceptual description (such as not annoying, slightly annoying, annoying, very annoying)
  - Give an overall quality rating on a numerical scale
  - Give a rating for a specific quality factor (numerical scale)
  - Give quality ratings for several different quality factors (multidimensional scaling)
- Based on subjective experimentation, a computational (objective) measure and model can be derived to simulate the perceived quality
  - Objective measures are less laborious and yield high repeatability
  - It is important to check the validity range of a model

Systemic framework for sound quality

Sounds of the problem domain => Subjective listening experiments => Data analysis (= quality ratings) => Sound quality models and theories

The development of sound-quality models and theories.

Sound source — signal analysis => — Signal and auditory features — mapping to quality factors => — sound quality measures

A general structure of a computational sound quality model

Subjective sound quality measurement

- Mean opinion score (MOS)
- Principle: ask opinion of some aspect of sound quality in numerical scale with anchors, take an average + other statistical measures
- Standardized methods to measure, e.g., (ITU-T P.800 (1996))

Value	Quality (MOS)	Impairment (DMOS)
5	Excellent	Imperceptible
4	Good	Perceptible, not annoying
3	Fair	Perceptible, slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Subjective sound quality measurement

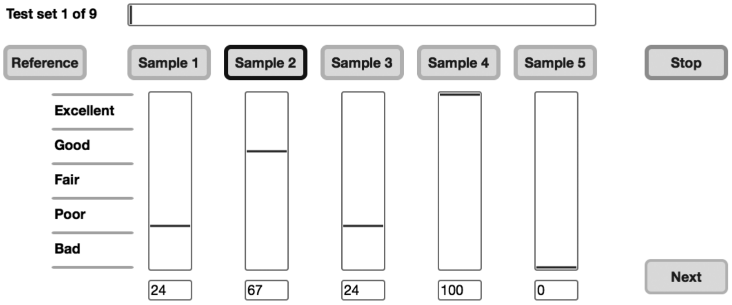
- MOS scales have been also defined for measurement of improvement or degradation of quality

Value	Categories (CMOS)
3	Much better
2	Better
1	Slightly better
0	About the same
-1	Slightly worse
-2	Worse
-3	Much worse

Subjective sound quality measurement

Multiple-stimulus hidden reference with anchors (MUSHRA)

- ITU-R BS.1534-1 (2003), ITU-R BS.1534-1 (2014)
- Originally: fast testing of audio codecs
- Reference (clean audio), test items, hidden reference, low-passed signals as anchors



Audio quality

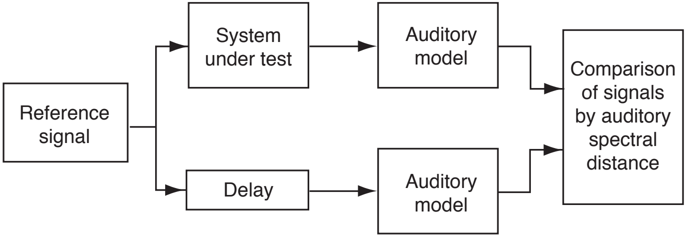
- Audio content production
- Product is finalized in mastering studio
- Consumer should perceive the audio content in the same way as the audio engineering in the mastering studio
- How to measure similarity?
- Typically the degradations in transmission channel are measured, not the differences in perceptions

Monaural audio quality

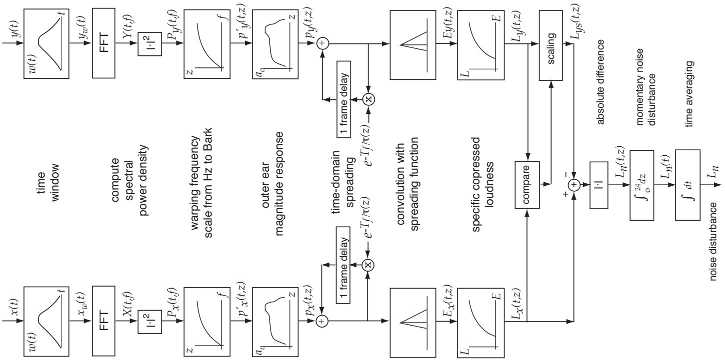
- Degradations of audio present by listening only one channel
- Deviations in:
  - Magnitude response
  - Phase and group delays
  - Non-linear distortions
  - Signal-to-noise ratio

Perceptual model for monaural audio quality

- Simple distance measures between original and reproduced audio do not tell much about perceptual difference
- Could we use auditory models to estimate perceived difference
- Principle:



Perceptual audio quality measure (PAQM)



Adapted from Beerends and Stemerdink (1992)

- Evolved to PEAQ (Perceptual audio quality) standardized measure
- Valid in limited applications

Spatial audio quality

- Tests with 5.1 surround: Spatial quality corresponds to 30% of total quality
- If sound is degraded by colorations, spatial quality looses its value
- How to measure spatial quality subjectively?
  - Comparison with reference, if exists and can be brought to listening room
  - Not simple solutions exist
- How to measure spatial quality objectively?
  - Binaural auditory models under research, standardization has failed

Quality of speech communication

- Speech intelligibility
- Speaker recognizability
- Speech naturalness
- Subjective and objective measurement

Speech quality: subjective measurement

- Articulation tests and articulation score
  - /CV/ or /CVC/ sequences used to measure recognition percentage
- Intelligibility test and intelligibility score
  - recognition percentage using meaningful words or sentences
- Rhyme tests (RT)
  - using 'rhyme' words or syllables (in Finnish: /patti/, /tatti/, /katti/)
- Diagnostic rhyme tests (DRT)
  - modifying single distinctive feature at a time (nasality, voicing, etc.) in RT
- Speech interference tests (find a disturbing noise level of 50% articulation)
- Quality comparison method, including pairwise comparison methods
- Other methods
  - Indirect judgement tests (PARM, QUART)
  - Communicability tests (communicate a drawing task, measure the difficulty)
  - Task recall tests (memorizing ability)
  - Noise suppression tests

Speech quality: objective measurement

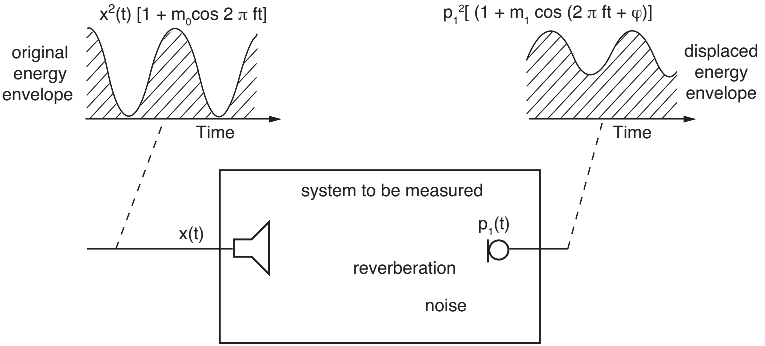
- Articulation index (AI)
  - for measuring a (linear) speech transmission channel with additive noise
  - articulation loss is assumed to be additive from 20 frequency bands
- Percentage articulation loss of consonants (%ALcons)
  - measure of speech intelligibitly, estimated from room acoustic parameters
- Room acoustical indices, see below
- Speech transmission index (STI, STIPA)
  - based on modulation transfer function, see below
- Signal-to-noise ratio (SNR)
  - ratio of speech vs. noise (power) level (in dB)
  - segmental SNR (SNRseg) based on short-time segmental SNRs
- Spectral distance measures
- Auditory sound quality measures (based on auditory modeling)
- Other methods
  - weighted spectral slope distance
  - LPC (linear prediction) distance measure

Modulation transfer function

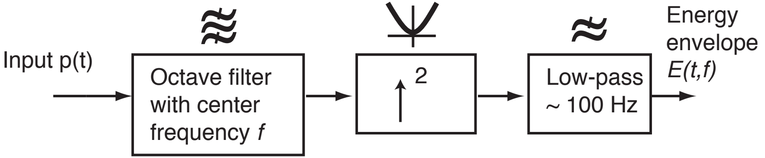
- The auditory system analyzes signals by critical bands
- Each band is analyzed by signal level, i.e., modulation envelope
- More important than the exact transfer function is modulation transfer function, i.e., how signal modulations in each critical band are transmitted
- The auditory system is most sensitive to modulations of about 4 Hz
- Modulation transfer is degraded by:
  - Reverberation (lowpass of modulation)
  - Background noise (reduction of relative modulation)
  - These effects are multiplicative (cascaded)
- Modulation transfer function is a mathematically motivated approximation of auditorily relevant signal transfer analysis

Effect of reverberation and noise on modulated sound

- Depth of modulation changes due to reverberation and noise



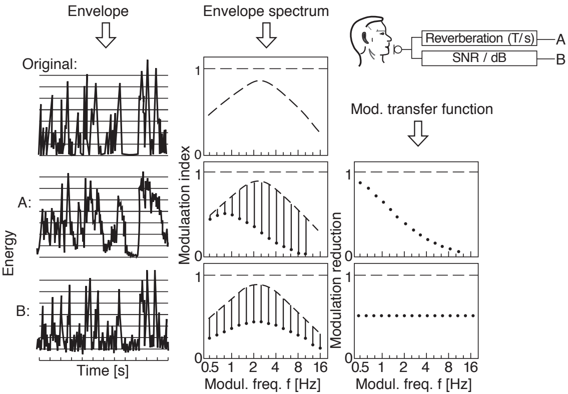
Measuring the envelope



- Similar to auditory modeling
- m_1 can be estimated:

$$m_1(f, f_\mathrm{m}) = 2\,\frac{\sqrt{|\int_t E_1(t,f)\sin(2\pi f_\mathrm{m} t)\,dt|^2 + |\int_t E_1(t,f)\cos(2\pi f_\mathrm{m} t)\,dt|^2}}{\int_t E_1(t,f)\,dt}$$

Effect of noise and reverberation on modulation transfer function



Adapted from Houtgast and Steeneken (1985)

Speech Transmission Index STI

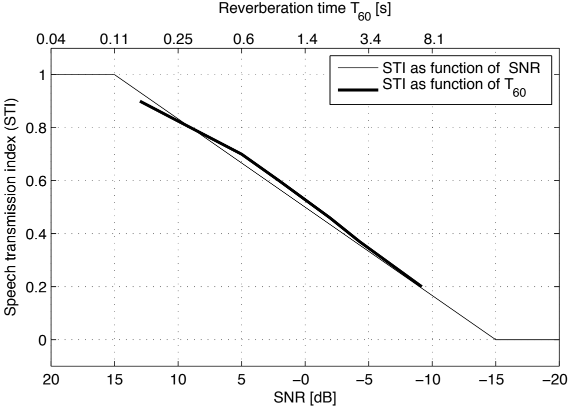
- Measure m with many carrier frequencies and modulation frequencies, take logarithm-like-measure and take a weighted average (Eq. 17.8)
- Speech transmission index (STI) optimally reflects the intelligibility of speech over measured channel
- STI measurement requires presentation of all carrier-frequency-pairs at separate times – slow
- STIPA uses limited number of combinations (gray boxes), and presents at the same time

Carrier-frequency and modulation pairs

- STI: all boxes
- STIPA: gray boxes

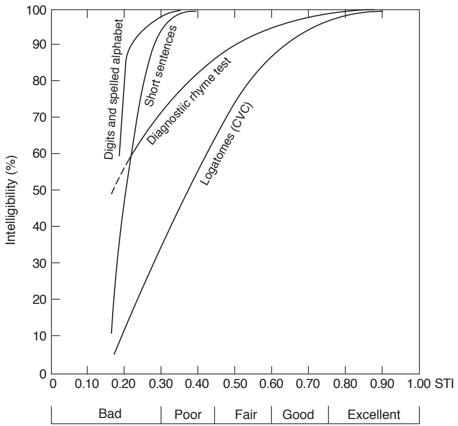
	Octave band						
Modulation frequency	125	250	500	1k	2k	4k	8 kHz
F_1 = 0.63 Hz							
F_2 = 0.8 Hz							
F_3 = 1.0 Hz							
F_4 = 1.25 Hz							
F_5 = 1.6 Hz							
F_6 = 2.0 Hz							
F_7 = 2.5 Hz							
F_8 = 3.15 Hz							
F_9 = 4.0 Hz							
F_{10} = 5.0 Hz							
F_{11} = 6.3 Hz							
F_{12} = 8.0 Hz							
F_{13} = 10 Hz							
F_{14} = 12.5 Hz							

Dependence of STI on noise and reverberation time



Adapted from Houtgast and Steeneken (1985)

Correlation between STI and speech intelligibility



Adopted from (Houtgast and Steeneken, 1985 and Steeneken and Houtgast, 2002).

Objective speech quality measurement for telecommunication

- STI does not estimate e.g. naturalness of speech
- Mobile telecommunication codecs may make speaker unidentifiable, although intelligibility is good
- More advanced methods needed
  - Models for general speech quality are expected to give a high MOS value only for natural-sounding and intelligible speech
  - Methods for measuring the perceptual effect of background noise suppression
  - Measures for echo suppression

Objective speech quality measurement for telecommunication

- When a mobile device enters the market, sound quality is to be tested
- Listening tests would be tedious
  - Basic quality
  - Effects of network problems delays, echos, background noise, background noise suppression
- Listening tests are not conducted, instead a standardized auditory-model-based evaluation tool is used
- 3GPP

Techniques for objective speech quality evaluation

- Perceptual speech quality measure (PSQM) 1998, withdrawn, e.g., temporal streching and compression of sound produced way too high error values
- Perceptual evaluation of speech quality (PESQ) 2002, tolerant to temporally varying jitter
- Telecommunications objective speech quality assessment (TOSQA) 2003, input through acoustical channel
- Perceptual objective listening quality assessment) POLQA 2011, also broad-band codecs
- Hearing-aid speech quality index (HASQI) 2014, sound-quality of hearing aids

The devices entering the markets have to produce enough high score with some of the mentioned techniques

Techniques for evaluation of the effect of background noise

- Mobile phones often involve algorithms to suppress background noise using non-linear DSP methods with time-variant processing
  - Non-linear noise-suppression algorithms try to reduce noise
  - How much does the algorithm reduce the problem
  - Does the algorithm introduce some new artifacts?
- Standardized MOS scales for listening tests
- Objective measurement 3QUEST needs recordings in standardized noise field with and without processing

Similar techniques exist also for measurement of the effect of echoes in two-way communication channel

Sound quality in auditoria and concert halls

- Concert halls have been built for music
- Music has been composed for concert halls
- Evolution of acoustics based on trial-and-error
- Certain type of music needs certain acoustics

Perceptual attributes of concert halls

Defined by Beranek by personal listening,

- Intimacy or presence / Reverberation or liveness / Spaciousness: Apparent source width(ASW) / Spaciousness: Listener envelopment (LEV) / Clarity / Warmth / Loudness / Acoustic glare / Brilliance / Balance / Blend / Ensemble / Immediacy of response / Texture / Freedom from echo / Dynamic range and background noise level / Extraneous effects on tonal quality / Uniformity of sound
- List is subject to debate and further studies

Objective measures of concert hall acoustics

- Measure few impulse responses in the hall
- Compute the values from the responses
- Subject to criticism, correspondence to actual perceptual quality is questionable

Subjective level of sound	Sound strength G in decibels
Perceived reverberance	Early decay time (EDT)
Perceived clarity of sound	Clarity C_{80} in decibels
Apparent source width (ASW)	Early lateral energy fraction, J_{LF}
Listener envelopment	Late lateral sound level, L_J in decibels

Objective measures of concert hall acoustics

Examples of the measures

- Strength

$$G = 10\ \log_{10} \frac{\int_0^\infty p^2(t)\,dt}{\int_0^\infty p_A^2(t)\,dt}$$

- Clarity (energy ratio between the early and late response)

$$C_{80} = 10\ \log_{10} \frac{\int_0^{80\text{ ms}} p^2(t)\,dt}{\int_{80\text{ ms}}^\infty p^2(t)\,dt}.$$

- Lateral fraction (reflects the ratio of lateral sound in the overall response)

$$J_{\mathrm{LF}} = \frac{\int_{5\text{ ms}}^{80\text{ ms}} p_8^2(t)\,dt}{\int_0^{80\text{ ms}} p^2(t)\,dt}$$

Noise quality

- Sound that is disturbing or annoying $\rightarrow$ purely subjective measure
- Annoyance: general concept of noise quality, also, how noise may upset an operation or activity
- Disturbance is connected to negative feelings where the functioning of the subject is not necessarily disrupted
- Annoyance depends on signal and on context
    - Speech and laughter are disturbing in open-plan office, ventilation humming is ok
    - Speech intelligibility over distance is not desired in open-plan office, just opposite to theatres and auditoria
- Objective model for noise annoyance or disturbance explain only certain cases

Product sound quality

- Minimize negative effects and maximize positive effects of product sound
- Mechanic devices communicate their functioning to to the user
- Electronic devices may have loudspeakers to do the same (car turn signal earlier relay, nowadays loudspeaker)
- Examples:
  - Cars and work machines
  - Home appliances
  - Office equipment
  - Personal devices

References

These slides follow corresponding chapter in: Pulkki, V. and Karjalainen, M. Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics. John Wiley & Sons, 2015, *where also a more complete list of references can be found.*

References used in figures:

Beerends, J.G. and Stemerdink, J.A. (1992) A perceptual audio quality measure based on a psychoacoustic sound representation. J. Audio Eng. Soc., 40, 963–978.

Houtgast, T. and Steeneken, H.J.M. (1985) The modulation transfer function in room acoustics. B&K Techn. Rev., 3, 3–12.

Steeneken, H.J. and Houtgast, T. (2002) Validation of the revised STIr method. Speech Commun. 38(3), 413–425.