ELL 788
Computational Perception & Cognition

July – November 2015

Module 9

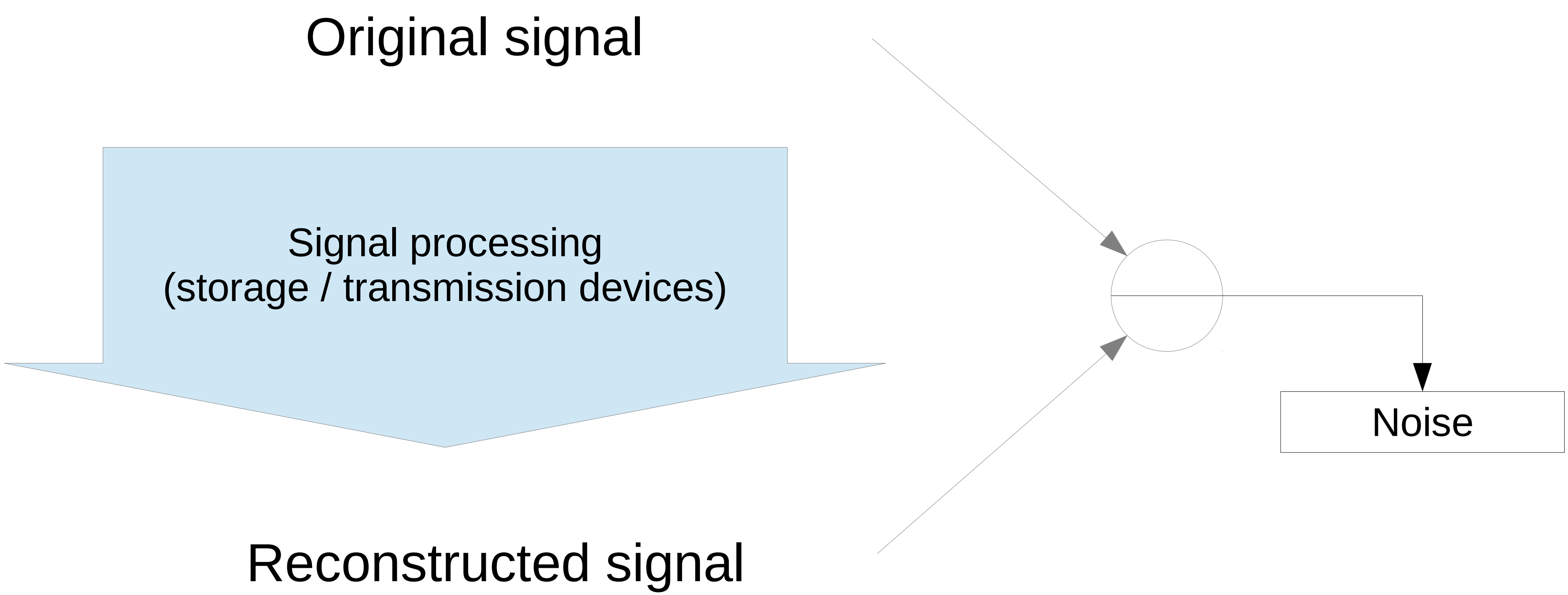
Audio Engineering: Quality Assessment

Audio quality:
why measurement is important?

- To assess quality of devices storing / transmitting audio, e.g.
  - Music systems
  - Telephone instruments / networks
  - Earphones
  - Cochlear implants ...
- Requirements are different for
  - Speech
  - Music (Hi-Fidelity)

Assessing audio quality

A signal processing approach



Signal to noise ratio

$$SNR = \frac{P_{Signal}}{P_{Noise}}$$

Sound event *vs.* Audio event

- Signal processing approach attempts to reproduce sound event accurately
  - Neither necessary, nor desired

For example,

- *Pause suppression in telephony to save bandwidth deteriorates SNR, but improves perceptual sound quality*
- *Perceptual coding (e.g. MP3) aims at reproducing an audio event and not a sound event.*

Perceptual approach to measure audio quality

- An appraisal of the perceived composition of an audio service with respect to its <u>desired</u> composition
  - Is contextual
    - What is good for telephony may not be good for hi-fdelity headphones for music
  - Depends on expectation on technology / prior experience
    - What is perceived good today may not be perceived good enough tomorrow.

Psycho-accoustic measurements

- Based on subjective judgement of users
  - Utilitarian: Overall quality in context of use
    - *Can you clearly understand what is spoken (telephony)*
  - Individual subdimensions
    - Intelligibility
    - Sound color (power spectrum / timbre)
    - Noisiness ...
  - Compute Mean Opinion Score (MOS)
    - *5-point scale (5:Excellent, ... 1: Bad)*

Approaches to judge audio quality

Multi-dimensional scaling (MDS)

- Sound is produced to listener in pairs
  - before and after signal processing
- Listener asked to differentiate auditory events in different perceptual dimensions
  - *Intelligability, Noisiness, ...*
- The disparities are presented in lower dimensional space

Semantic Differential (SD)

- Sound events are presented individually
- listeners rate each auditory event on a number of bipolar scales
  - *Loud–quiet, noisy–not noisy, ...*
- Comparison with an implicit reference
  - *Subjective: based on user's experience / expectation*
- Results presented in a reduced dimensional space

Which one is more appropriate for real-life quality judgment ?

Measurement criteria

- **Validity:**
  - Should measure what is intended to be measured
- **Reliability:**
  - Results should be stable within and between measurement instances
- **Objectivity:**
  - Independence of results from the assessor (not the subject)
- **Realistic test situation:**
  - Perceptual expectation depends on
    - Context (e.g. Speech vs. music)
    - Past experience
    - Updated by technology changes

Economy and feasibility: Need for instrumental measurements

- Psycho-accoustic experiments are expensive and time-consuming
  - Cannot be done in real-time
    - Dynamic assessment of voice quality in a telephone network (dynamic re-routing)
- Attempt to replace psycho-accoustic measurements by instrumental mesaurements
  - The perception and judgment processes triggered by the sound event can be described by algorithms which are trained to produce estimations or predictions of judged quality
  - Economic and less time-consuming
  - Repeatable, reliable and objective
  - Can be performed in real-time

Dimensions of perceptual quality prediction algorithms

- **Media:**
  - Audio (+ visual),
    Speech / music
  - Channel, e.g. Music system / Telephone network
- **Time-frame of prediction:**
  - Instantaneous audio quality during transmission
  - Average audio quality of a recorded song / speech
  - Overall audio quality produced by a system, e.g. an amplifier system
- **Interaction Situation:**
  - Listening only / Conversational
  - Studio / Home Enviromnet, Public place

... more

- **The predicted target variable:**
  - Overall quality or individual quality features
    - Intelligibility, noisiness, etc.
- **The types of signal degradations:**
  - effects of codecs / impact of channel degradations
    - noise, attenuation, echo, delay
- **The input information used for the prediction:**
  - Single-ended or double-ended
- **Application scenario:**
  - On-line prediction (monitoring / immediate corrective action)
  - Off-line prediction (planning)

Objective sound quality measurement methods

Table 1: Technical classification of objective sound quality measurement methods

		Input signal	Main purposes
Media layer models	Full-reference	Original sound, processed sound (signal from device under test)	· Ascertaining performance of equipment, etc. · Optimizing system parameters
	Non-reference	Processed sound (signal from device under test)	· In-service quality management
	Reduced-reference	Processed audio (signal from device under test), features of original sound	· In-service quality management
Packet layer models		Packet header information (RTP etc.)	· In-service quality management
Parametric models		Quality design & management parameters	· Network quality design · In-service quality management
Bitstream layer models		Coded bitstream (before decoding)	· In-service quality management
Hybrid models		Combination of the above	· In-service quality management

For, Hi-Fidelity audio, work is primarily restricted to Full-reference method.

A couple of points

- We need quantitative measures
  - Continuous scale / discrete values / labels
  - The experience (possibly, infinitely many dimensions) needs to be quantized into finite 1-D scale
- Temporal integration
  - Perception of quality and judgment are instantatneous
    - 4 – 8 sec clips are used
  - Temporal integration of quality takes place when the experience gets longer (1 - 2 minutes)
    - Negative events count more than positive ones
    - Persistence: negative quality perception persists for a few seconds
    - Recency effect: Events happening close to the judgment point-in-time are more important than previous ones

Models for instrumented audio quality analysis

- Signal comparison approach
  - Quality prediction as a comparison between perceived and expected characteristics
- Parametric Approach
  - Integrating different quality dimensions (system parameters)
- Temporal Integration Models
  - Aggregating the instantaneous experiences into a single rating for a longer duration

Signal comparison model

- Comparison between expected and actual signals
  - Assumes availability of original (ideal) signal
- Compare the signals (sound-events) in spectral domain (perceptual space)
  - Time-align the signals; normalize in amplitude
  - Transform both signals to perceptual space (spectrogram)
  - Extract perceptual features: Pitch, Loudness, etc. and compare
    - *Assumption: perceptual distance between the (clean) input and the output signal of the transmission channel is inversely related to quality*
- Integrate over time to produce overall quality judgment
- Convert to a MOS-like score

Parametric model

- Mouth to ear quality (telephony)
- Does not depend on availability of signals
- Uses 18 scalar system parameters that describe the perceptual effects associated with different terminal and transmission equipment. e.g.
  - Loss of loudness (with respect to a reference path)
  - Phase distortions (Delays for different frequencies)
  - Parameters affecting noise, echo, etc.
- Integration of different types of degradations onto a single quality scale
  - “Impairment factors” for talking, listening and conversation are calculated ffrom the input parameters (details in next slide)
  - Impairments are then subtracted from the optimum quality of the system
  - Finally, a MOS score is computed

Impairment factors

- Degradations resulting from a too low SNR
- Degradations occurring simultaneously with the speech signal
  - *Too loud or too quiet connection, bad side tone, etc.*
- Degradations occurring delayed with respect to the speech signal
  - *Echo, conversational impact of delay, etc.*
- Degradations resulting from nonlinear and time-varying processing
  - *Codecs, packet loss, etc.*

Temporal integration models

- Temporal integration to assess the overall quality perception (over a call)
- Integration of MOS over successive time samples
  - Simple average can be a first approximation
  - To improve the results
    - Higher weights for extremely negative ratings
    - More weight to samples near the judgment (end) time

Diagnostic prediction

- MOS is not very useful for diagnosis
  - *How to improve the quality ?*
- Signal comparison or parametric models provide more insights
- Two possible approaches for diagnosis
  - Technical causes can be identified which provoke such problems
    - Several technical causes may lead to similar perceptual defects
    - Algorithms / models may be too specific to the technology
  - Perceptual dimensions can be estimated which indicate the related perceptual effects
    - That tells what problem diimensions to address

Perceptual dimensions for speech

- Perceptual qualities for speech
  - ***Intelligibility***
  - Coloration, Discontinuity, Noisiness (Orthogonal)
- Prediction of speech intelligibility (Articulation Index)
  - Compute SNR within several frequency bands; Normalize; subject to masking effects
  - Combine with a preceptually weighted average $\rightarrow$ Articulation Index
- Predicting other speech qualities
  - Coloration is associated with frequency response
  - Noise in silence / Noise in speech
  - Discontinuity: a non-linear combination of an interruption rate, an artefact rate, and a clipping rate (derived from spectogram)
- Comining different qualities to compute MOS
  - Use cognitive model – trained *kNN* classifier

Some applications

- Telephone network planning
  - Equipment to use; Routing
- On-line adaptation of routing
  - In case of a node failure
- Online Intelligibility Improvement of Speech
  - Use of filters based on perceptual noise model
  - Changing consonant-vowel ratio

References

- Moller and Heusdens. Objective Estimation of Speech Quality for Communication Systems. Proc. IEEE, Sept 2013
- Objective perceptual audio quality measurement methods