ITU-T Workshop on
"From Speech to Audio: bandwidth extension, binaural perception"
Lannion, France, 10-12 September 2008

CHARACTERIZATION OF CURRENT CODECS DEGRADATIONS FOR SUBJECTIVE ASSESSMENT OF SPEECH QUALITY

T. ETAME [1, 3], C. QUINQUIS [1], L. GROS [1],
G. FAUCON [2, 3] and R. LE BOUQUIN JEANNES [2, 3]

[1] France Telecom R&D, Lannion, France
[2] INSERM, Rennes, France,
[3] University of Rennes, Rennes, France

Overview

- Introduction
- Technical specifications of codecs under evaluation
- Multidimensional scaling technique (MDS)
- Experiments: selection of codecs, dissimilarity test and results
- Degradation indicators: what correlates correspond to the perceptive dimensions?
- Conclusion

Overview

- Introduction
- Technical specifications of codecs under evaluation
- Multidimensional scaling technique (MDS)
- Experiments: selection of codecs, dissimilarity test and results
- Degradation indicators: what correlates correspond to the perceptive dimensions?
- Conclusion

Introduction

- Context:
  - subjective assessment is the most reliable way to determine overall perceived voice quality of network equipment, as digital codecs
  - reference conditions are useful in subjective tests to provide anchors so that results from different tests can be compared
  - the MNRU[(1)] is extensively used in subjective evaluations to provide a simulated and calibrated degradation qualitatively similar to quantization distortion of waveform codecs
  - the improvements in audio coding technologies modified the types of degradations and so the MNRU[(1)] is not representative any more of the current degradations
- Goal:
  - the purpose of our work is to produce a reference system that can simulate and calibrate degradations of speech and audio codecs which are currently used on telecommunications networks
- Our approach:
  - produce the multidimensional perceptive space underlying the perception of current degradations
  - characterize these perceptive dimensions
  - simulate and calibrate similar degradations

[(1)] MNRU: Modulated Noise Reference Unit (ITU-T P.810)

Overview

- Introduction
- Technical specifications of codecs under evaluation
- Multidimensional scaling technique (MDS)
- Experiments: selection of codecs, dissimilarity test and results
- Degradation indicators: what correlates correspond to the perceptive dimensions?
- Conclusion

Technical specifications of codecs under evaluation

- ADPCM codecs (Adaptive Differential Pulse Code Modulation)
  - **ITU-T G722** at 64, 56 and 48 kbps
- ACELP codecs (Algebraic Code-Excited Linear Predictive)
  - **ITU-T G722.2** at 6.6 - 23.85 kbps
- Transform codecs
  - MLT (Modulated Lapped Transform) with Scalar Quantized Vector Huffman Coding and bit allocation by categorization (**ITU-T G722.1** Wideband at 24, 32 kbps, **ITU-T G722.1C** Super-Wideband at 24, 32, 48 kbps)
  - MDCT (Modified Discrete Cosine Transform) with scalar Huffman coding and rate loop, based on psychoacoustic model (**MP3** format which is the standard MPEG Layer-3 (MP3 / ISO/IEC 11172-3))
  - MDCT with Spectral Band Replication (SBR) (**HE AAC** Full band at 16 - 48 kbps)
- Hybrid codecs (CELP + Time-Domain Aliasing Cancellation (TDAC))
  - **ITU-T G729.1** at 14 - 32 kbps

Overview

- Introduction
- Technical specifications of codecs under evaluation
- Multidimensional scaling technique (MDS)
- Experiments: selection of codecs, dissimilarity test and results
- Degradation indicators: what correlates correspond to the perceptive dimensions?
- Conclusion

Multidimensional scaling technique (MDS)

- Limits of methods using semantic descriptors
  - people lack of vocabulary to describe most of the auditory sensations
  - the choice of descriptors is biased by the experimenter

- Multidimensional scaling technique
  - MDS consists of studying the perceptive structures which underlie the judgments of similarities given for pairs of stimuli, by translating them into a matrix of distance
  - the matrix of distance is used to project the whole of the stimuli or objects in a multidimensional space according to a mathematical model
    - there is no presumption on dimensions

- Which MDS?
  - nonmetric multidimensional scaling (the ranking induced by the judgments of dissimilarities in human listening experiments is more reliable than their values)
  - INDSCAL (INDividual differences SCALing) (subjects differently weight the perceptive dimensions when they elaborate an overall judgment)

Overview

- Introduction
- Technical specifications of codecs under evaluation
- Multidimensional scaling technique (MDS)
- Experiments: selection of codecs, dissimilarity test and results
- Degradation indicators: what correlates correspond to the perceptive dimensions?
- Conclusion

Experiments: selection of codecs (1/2)

- Tandem speech coding was applied to the nineteen following codecs in order to introduce degradation at different magnitudes
  - G722 at 64, 56, 48 kbps;
  - G722.2 at 8.85, 12.65, 15.85, 23.85 kbps;
  - G722.1 at 24, 32 kbps;
  - G729.1 at 14, 20, 24, 32 kbps;
  - HEaac at 16, 24, 32 kbps;
  - G722.1 C at 24 kbps;
  - MP3 at 32, 64 kbps.
  - 58 conditions (19 codecs x 3 tandem-levels + original signal)
- An ACR (Absolute Category Rating) test was run on to select around twenty codecs with medium speech quality (2 < MOS < 3.5)
  - The samples used in the ACR test consisted of pairs of sentences spoken by two male and two female talkers, two samples per talker
  - Each condition output signal is filtered in order to limit its bandwidth to wideband
  - Thirty-two subjects participated in this ACR listening test

Experiments: selection of codecs (2/2)

	Description		Description
+ 1	G722.1C_24kbps_x2	°11	G722_56kbps_x2
+ 2	G722.1C_24kbps_x3	°12	G722_56kbps_x3
+ 3	G722.1_24kbps_x2	*13	G729.1_14kbps_x3
+ 4	G722.1_24kbps_x3	*14	G729.1_20kbps_x3
x 5	G722.2_12.65kbps_x2	*15	G729.1_24kbps_x2
x 6	G722.2_12.65kbps_x3	*16	G729.1_32kbps_x3
x 7	G722.2_15.85kbps_x2	□17	HEAAC_24kbps_x2
x 8	G722.2_8.85kbps_x2	□18	HEAAC_32kbps_x2
° 9	G722_48kbps_x2	□19	MP3_32kbps_x1
°10	G722_48kbps_x3	□20	MP3_32kbps_x2

The 20 tandem codecs selected for the test of dissimilarity

Experiments: dissimilarity test

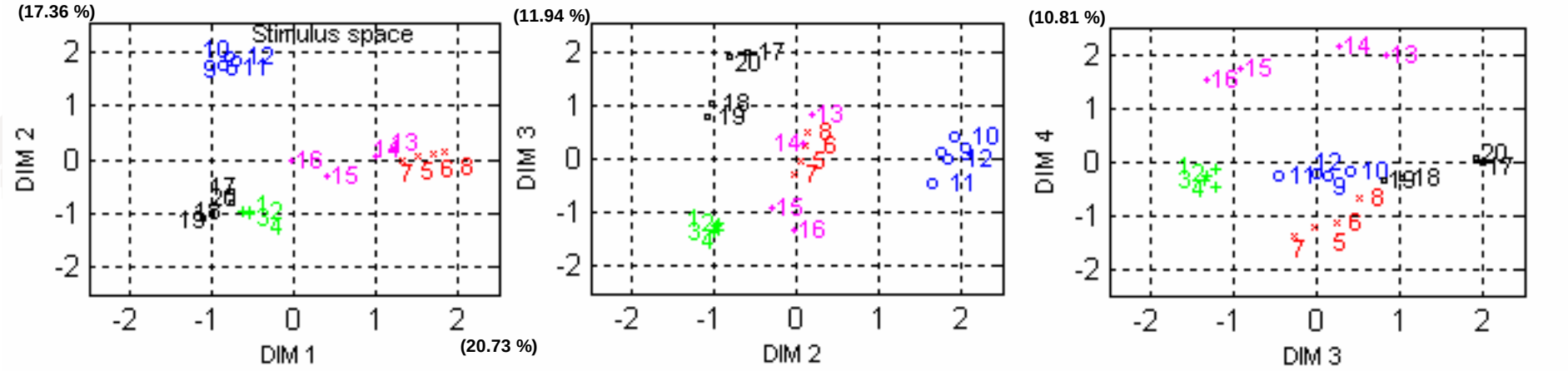
- dissimilarity test
  - A 6 s speech sample uttered by one male was processed by the twenty selected codecs with output limited to wideband
  - All in all, 210 pairs (190 + 20 null pairs) were presented in random order to subjects, with a different randomization for each subject
  - For each pair, the subject was asked to evaluate similarity coded speech samples on a continuous line scale varying between similar(0) and different(100)
  - A single subject participated in each test. Scores were collected in two sessions around one hundred trials each (90 minutes) in two different days
  - Twenty-five subjects participated in the experiment
- verbalization task
  - In order to identify the perceptive attributes corresponding to the four dimensions
  - After listening to the twenty tested codecs, subjects were invited to describe in their own words the degradations they had perceived during the test
  - dissimilarity test

Overview

- Introduction
- Technical specifications of codecs under evaluation
- Multidimensional scaling technique (MDS)
- Experiments: selection of codecs, dissimilarity test and results
- Degradation indicators: what correlates correspond to the perceptive dimensions?
- Conclusion

Experiments: dissimilarity test results

- MDS: 4-dimensional solution interpretable (stress = 0.21)



	Description		Description
+ 1	G722.1C_24kbps_x2	◦11	G722_56kbps_x2
+ 2	G722.1C_24kbps_x3	◦12	G722_56kbps_x3
+ 3	G722.1_24kbps_x2	*13	G729.1_14kbps_x3
+ 4	G722.1_24kbps_x3	*14	G729.1_20kbps_x3
x 5	G722.2_12.65kbps_x2	*15	G729.1_24kbps_x2
x 6	G722.2_12.65kbps_x3	*16	G729.1_32kbps_x3
x 7	G722.2_15.85kbps_x2	□17	HEAAC_24kbps_x2
x 8	G722.2_8.85kbps_x2	□18	HEAAC_32kbps_x2
◦ 9	G722_48kbps_x2	□19	MP3_32kbps_x1
◦10	3	□20	MP3_32kbps_x2

- Perceptive attributes
  - DIM 1: "clear/muffle"
  - DIM 2: "background noise"
  - DIM 3: "noise on speech"
  - DIM 4: "hiss "
- Study of talker effect with 25 other listeners
  - Male and female spaces are similar and highly correlated
    - DIM 1: 0.94
    - DIM 2: 0.98
    - DIM 3: 0.93
    - DIM 4: -0.88

Degradation indicators: what correlates correspond to the 4 dimensions?

- The **spectral centroid** (SC), indicator of timbral brightness is linked with DIM1 (clear/muffle)

	SC (for male talker (m))	SC (for female talket (f))
DIM 1	**-0.93***	**-0.91***
DIM 2	0.01	0.17
DIM 3	-0.07	0.23
DIM 4	0.06	-0.35

- The SC feature can be formulated as: $SC = \frac{\sum_{i=1}^{N} f_i \times P_i}{\sum_{i=1}^{N} P_i}$
  - where f_i (expressed in Hz) is the frequency of the *i*th frequency coefficient with power Pi,
  - *N* = number of frequencies of the *SC* feature

- The **energy on the period of silence** explains DIM2 (background noise) only on the higher band [4-7 kHz]

	LB (50 – 4000 Hz)	**HB (4000 – 7000 Hz)**
DIM2 (m)	0.45	**0.71**
DIM2 (f)	0.39	**0.76**

- DIM3(noise on speech) is significantly correlated with **MOS**

	MOS (m)	MOS (f)
DIM 1	-0.13	-0.36
DIM 2	-0.19	-0.29
DIM 3	**-0.69***	**-0.56***
DIM 4	-0.15	-0.28

- The **cross-correlation between the power spectra** of the original signal x and of each coded version y characterizes DIM4 (hiss)

(m)	R_{xy} mean	R_{xy} max	R_{xy} std
DIM 1	-0.24	-0.15	-0.09
DIM 2	-0.05	0.00	-0.04
DIM 3	0.00	0.10	0.05
DIM 4	**-0.86***	**-0.90***	**-0.91***

(f)	R_{xy} mean	R_{xy} max	R_{xy} std
DIM 1	-0.34	0.40	-0.11
DIM 2	0.10	-0.02	0.09
DIM 3	-0.03	-0.30	-0.08
DIM 4	**0.62***	**0.78***	**0.74***

(*) Pearson correlations significant at the 0.05 level

Overview

- Introduction
- Technical specifications of codecs under evaluation
- Multidimensional scaling technique (MDS)
- Experiments: selection of codecs, dissimilarity test and results
- Degradation indicators: what correlates correspond to the perceptive dimensions?
- Conclusion

Conclusion

- The aim of this study is:
  - to link the current coding technologies with their potential perceptive degradations
  - to produce a reference system that can simulate and calibrate these degradations for subjective tests
- Results
  - Coded speech samples for one male talker are represented in a 4-dimensional perceptive space (clear/muffle, background noise, noise on speech, hiss)
  - Similar analysis for the female talker leads to similar conclusions
  - Perceptive dimensions are characterized by correlates: spectral centroid (dimension 1), energy on the period of silence (dimension 2), cross-correlation between the spectrum of the original signal and this one of the coded versions (dimension 4), MOS (dimension 3)
- Further works
  - to simulate and calibrate these degradations for subjective tests

Thank you for your attention

- acknowledgments
  - The present study was carried out at France Telecom R&D , TECH/SSTP, Lannion, France
  - It was supported by:
    - INSERM, U 642, Rennes, France
    - Université de Rennes 1, LTSI, Rennes, France