16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

CHARACTERIZATION OF CURRENT CODECS DEGRADATIONS FOR SUBJECTIVE ASSESSMENT OF SPEECH QUALITY

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

(1) France Telecom R&D TECH/SSTP - 2 Av. Pierre Marzin, 22307 Lannion Cedex, France
(2) INSERM, U 642, Rennes, F-35000 France, (3) Université de Rennes 1, LTSI, Rennes, F-35000, France
phone: +33 (0)2 96 05 98 88, fax: +33 (0)2 96 05 35 30
{thierry.etameetame, laetitia.gros, catherine.quinquis}@orange-ftgroup.com, {regine.le-bouquin-jeannes, gerard.faucon}@univ-rennes1.fr

ABSTRACT

Our work aims to provide reference signals representative of the new techniques of speech coding, for subjective assessment tests of speech quality. At first, we determine a four-dimensional perceptive space of the degradations resulting from 20 wideband codecs. Then, from a verbalization task, it appears that the identified attributes are clear/muffle, background noise, noise on speech and hiss. Finally, these dimensions are characterized with correlates such as spectral centroid, power spectrum for the period of silence, Mean Opinion Score and correlation coefficient.

1. PURPOSE AND MOTIVATION

The quality of speech signals transmitted by a telecommunication system may be impaired at different levels of the end-to-end transmission chain, especially by coding. Subjective assessment with naive listeners is the most reliable way to determine overall perceived speech quality of digital codecs. In the Absolute Category Rating procedure (described in [1]) extensively used for the assessment of codecs audio quality, the listeners are asked to listen to short speech samples processed by the codecs under study and to give their opinion on the speech quality on an appropriate scale. The Modulated Noise Reference Unit (MNRU) is extensively used in subjective evaluations of digital processes, both in conventional telephone bandwidth and in wideband, to introduce controlled degradations into speech signals [2]. These degradations are characteristic of the quantization noise of the codecs like PCM (Pulse Code Modulation) or ADPCM (Adaptive Differential Pulse Code Modulation). The resulting signals are used as reference conditions so that experiments made in different laboratories or at different times can be sensibly compared [1].

The improvements in audio coding technologies modified the types of degradation and so the MNRU is not representative any more of the current degradations.

Therefore, our work aims to provide reference signals representative of the current coding degradations. The first step of the work consists in identifying the perceptive characteristics of the degradations resulting from the new audio coding techniques by listening experiments. The comparison of these perceptive characteristics with various coding techniques should help to artificially generate these degradations and to introduce them into speech signals to get new reference signals.

Codecs	Technical Specifications
ITU-T G722 at 64, 56, 48 kbps (WB)	Sub-Band Adaptive Differential Pulse Code Modulation (SB-ADPCM)
ITU-T G722.1 at 24, 32 kbps (WB)	Modulated Lapped Transform (MLT) and bit allocation by categorization
ITU-T G722.1 C at 24 kbps (S-WB)	Low-complexity extension mode to G.722.1
ITU-T G722.2 at 8.85, 12.65, 15.85, 23.85 kbps (WB)	Algebraic Code-Excited Linear Predictive (ACELP) and bandwidth extension (BWE, 6400 - 7000 Hz)
ITU-T G729.1 at 14, 20, 24, 32 kbps (WB)	- CELP (50-4000 Hz) - Time-Domain Bandwidth Extension (TD-BWE) (4000 - 7000 Hz) - Time-Domain Aliasing Cancellation (TDAC) (50 - 7000 Hz)
High Efficiency-AAC at 16, 24, 32 kbps (FB) MPEG-4 standard for low bit rate coding	MDCT (Modified Discrete Cosine Transform) with Spectral Band Replication (SBR), stimulus is a stereo-to-mono down-mixing signal
MPEG-1 layer III (FB) (ISO/IEC 11172-3) at 32, 64 kbps	MDCT and psychoacoustic model: constant bitrate encoding with standard MP3, built by Cool Edit Pro

Table 1 - Technical specifications of the 19 codecs

Codecs under evaluation are presented in section 2. Section 3 justifies the choice of the MultiDimensional Scaling (MDS), technique used in this study to determine a multidimensional perceptive space which underlies the perception of current impairments. The listening experiment and its results analyzed by MDS are given in section 4. In section 5, these dimensions are interpreted and characterized with correlates which are all further discussed.

2. TECHNICAL SPECIFICATION OF CODECS

As it will be detailed in paragraph 4.1, we selected WideBand (WB, 16 kHz input signal sampling rate), SuperWideBand (S-WB, 32 kHz) and Full Band (FB, 48 kHz) codecs which are similar in terms of global speech quality. Most of the quality assessment tests on speech consider test signals in the same bandwidth. Since there is no on-going work on Narrow Band (NB) codecs, only wider bandwidth codecs are considered. Each codec output signal is filtered in order to limit its bandwidth to wideband, so that the subjective judgements do not relate too much to the bandwidth of the codecs. In order to have a maximum of possible degradations, we considered a maximum of codecs presenting varied coding techniques. Table 1 briefly presents the technical specifications of 19 codecs, some of them being considered in the dissimilarity test.

3. MULTIDIMENSIONAL SCALING TECHNIQUE

In [3] we chose MultiDimensional Scaling (MDS) technique to determine a multidimensional perceptive space which underlies the perception of current codecs degradations, in order to take into account the multidimensional nature of speech quality. This technique presents the advantage to not have presumption on dimensions contrary to the methods using semantic descriptors. The multidimensional scaling consists in studying the perceptive structures which underlie the judgements of similarities given for pairs of stimuli, by translating them into a matrix of distance. This matrix is used to project all stimuli or objects in a multidimensional space according to a mathematical model [5].
It is assumed that the order induced by the judgements of dissimilarities in human listening experiments is more reliable than the numbers given by listeners. The nonmetric multidimensional scaling ([7] [8]) takes into account this point by computing a dissimilarity matrix in which values are a rank ordering of the distance between objects. In addition, the inter-individual variability between subjects is taken into account in the weighted multidimensional scaling also referred as INDSCAL (INDividual differences SCALing) ([9][10]). In our study, we chose a non-metric INDSCAL MDS which takes into account the characteristics of perceptive evaluations of audio quality.
Two examples (also reported in [11]) of perceptive spaces underlined by Non-metric INDSCAL MDS from the domain of speech quality evaluation are summarized in Table 2. Hall's study data [6] contain 10 speech processing systems such as low bit-rate codecs, MNRU, ADPCM (32kbps). As reported in [11], Hall's study [6] delivered a proof based on MDS that MNRU is perceptually different from low bit-rate coded speech: the 3 dimensions determined (listed in Table 2) and the respective loadings for different conditions show that MNRU is the only degradation in the test strongly positioned on "noisiness" dimension [11]. Wältermann [12] studied 14 different VoIP/PSTN network conditions including packet loss. To compare results in [6] and [12] to ours, it is important to mention that:
- "Naturalness" dimension does not present physical correlate, but is the most highly correlated dimension with MOS,
- "Noisiness" and "Frequency content" dimensions are shared by [6] and [12],
- "Continuity" dimension in [12] is the perceptive effect due to packet loss conditions.
The effects of some disturbances influencing speech quality during transmission over telecommunication networks have been studied in [6] and [12], in order to develop speech objective quality measures. Our approach aims at understanding the impact of coding techniques degradations to provide reference signals representative of those degradations. Hence, in this paper, we will study separately coding degradations before to combine them with other forms of impairments (like packet loss for example). Moreover, our study includes the recently normalized WB ITU-T G729.1 hybrid codec and so other possible new degradations.

Study	**Hall [6]**	**Wältermann [12]**
System	**codecs(10)**	**VoIP/PSTN (14)**
Naturalness	x	
Noisiness	x	x
Amount low frequencies / Frequency content	x	x
Continuity		x

Table 2 – Summary of two tests for the MDS speech signals

4. EXPERIMENTS

4.1 Selection of codecs

In order to introduce degradation at different magnitudes, tandem speech coding was applied to the nineteen codecs described in Table 1. Tandem speech coding, where two codecs operate on a signal in cascade, as in the case of mobile communication, can amplify the degradation. Our study will consider the cases where one (_x1), two (_x2), or three (_x3) identical codec(s) are present (see Table 3 for details).
In a MDS analysis, it is desirable to include practically as many stimuli as possible in an experiment, since the number of dimensions which can be extracted increases with the number of stimuli. Ideally, Kruskal recommends nine stimuli for two dimensions, thirteen for three and seventeen for four. These recommendations are, however, for a single matrix of data. When more than about ten matrices are to be analyzed, the recommendations can be weakened [5]. Therefore, around twenty codecs with more than twenty subjects seem appropriate for a perceptive space of four or five dimensions, which is a reasonable number of dimensions from a perceptive point of view.
In [13], the naive listeners did not differentiate between higher quality conditions. To avoid inter- or intra-individual inconsistencies in the formation of the perceptive space, an ACR (Absolute Category Rating) test (P.800 [1]) was run to select 20 codecs with medium speech quality so that the judgements of dissimilarity do not relate too much to global quality, but only to the type of defect. For this reason, the high quality-original speech signal is not included in the test to see its behaviour as a reference. The 20 tandem codecs finally selected for the dissimilarity test are given in Table 3.

	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

Table 3 - The 20 tandem codecs for the test of dissimilarity (the symbols of the first column correspond to codec families)

4.2 Dissimilarity test procedure

A 6-second speech sample uttered by one male (in french, "La vanille est la reine des arômes. Fragile, il ne résiste pas à l'air glacé."), originally a full band signal sampled at 48 kHz, was downsampled as input signal for wideband and super-wideband codecs. It was then filtered with a [50 - 7000 Hz] band-pass filter for wideband codecs and a [50 - 14000 Hz] band-pass filter for super-wideband versions. The resulting stimuli were processed by the 20 selected codecs with output limited to wideband. Finally, these coded versions were upsampled to 48 kHz for compatibility with stimulus presentation equipment.

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 perceptive distance between coded speech samples. The dissimilarity between the samples is rated on a continuous line scale varying between 0 (similar) and 100 (different). For each pair, the two coded versions of the speech sample were presented so as the subject could freely switch from one version to another to better detect differences. A single subject participated in each session and was asked to listen to the test sample pair at least once, after which the overall similarity was to be rated on the scale. Scores were typically collected in two sessions around one hundred trials each (90 minutes) in two different days. Each session included a ten-pair preliminary session. Twenty five subjects participated in the experiment.

4.3 Results

Analyses were carried out by the SPSS (Statistical Package for Social Sciences) INDSCAL algorithm that is based on the ALSCAL (Alternating Least Square SCALing) algorithm.

The stimulus space derived for the twenty speech samples used with the male talker is shown in Figure 1. Several criteria, including residual stress, variance accounted for, and average weight given by the listeners to each dimension, indicate that four is an appropriate number of dimensions (stress = 0.21). Dimension 1 (DIM1) accounts for 20.73 % of the total explained variance, dimension 2 (DIM2) for 17.36 %, dimension 3 (DIM3) for 11.94 % and dimension 4 (DIM4) for 10.81 %. The results reveal a regrouping of the codecs according to technical specifications.

In order to study the talker effect on the perception of impairments, another experiment was carried out using the same dissimilarity test procedure with a speech sample generated by a female talker and 25 other listeners. Although processes often affect male and female voices differently [1], the comparison of the two multidimensional spaces obtained with the two talkers show that these two spaces are similar and highly correlated: the correlation coefficient for dimension 1 is 0.94, 0.98 for dimension 2, 0.93 for dimension 3 and -0.88 for dimension 4. Therefore, it seems that the same perceptive dimensions underlie the perception of impairments generated by current codecs.

5. INTERPRETATION OF DIMENSIONS

The results shown in section 4 establish that weighted multidimensional scaling can be used to fit speech samples processed by the codecs used in this experiment into a four-dimensional stimulus space. In this section, we try to identify what perceptive attributes and physical correlates correspond to the four dimensions.

5.1 Perceptive attributes

In order to identify the perceptive attributes corresponding to the four dimensions, we had a verbalization task at the end of the dissimilarity test: after listening to the twenty tested codecs, subjects were invited to describe in their own words the degradations they had perceived during the test. As illustrated in Figure 1, the G722 codecs degradation is the only degradation strongly positioned on dimension 2. For the four tested G722 codecs, the verbalization experiment evinces attributes like noisy, background noise, blowing, and crackling (generic term: background noise). In the same way, dimension 4 separates only G729.1 codecs from all the other codecs. With verbalization qualifications, it appears that the most often identified attribute for dimension 4 is hiss. When listening to different objects along the dimension 1 and dimension 3, the identified attributes are clear/muffle with dimension 1 and noise on speech with dimension 3.

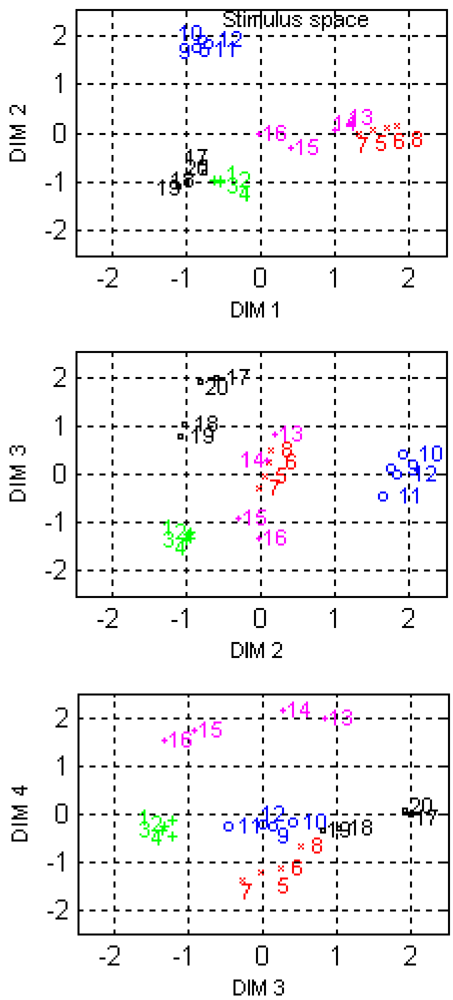

Figure 1 - Plot of object space

5.2 Physical correlates

At first, we try to find a link between energy in different bands (which may reveal the attribute clear/muffle) and dimension 1. So, we computed the average power spectrum of each of the 20 speech samples, using a 1024-point Fast Fourier Transform (FFT), in order to determine correlation with DIM1. Then, we subtracted (on a dB scale) the original signal power spectrum from each of the individual spectrum. The resulting deviations of some codecs are represented in Figure 2 for male talker.

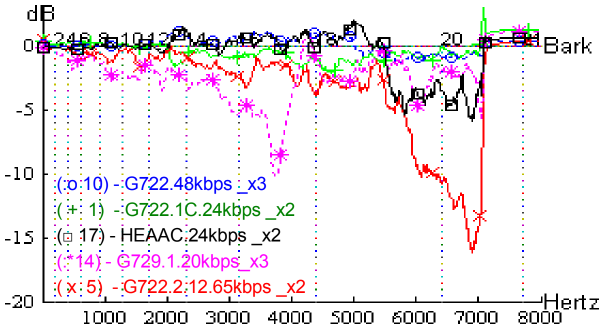

Figure 2 – Deviations from original spectrum of some codecs

Figure 2 illustrates the energy change between G722.2 (x 5) (ACELP codecs) at positive extremity of dimension 1 (Figure 1) on the one hand, and on the other hand G722.1, G722, HEAAC, MP3 at negative extremity of this dimension.
WB CELP speech coding typically fails to match well the spectrum in the high frequency band and then generates bandwidth limitation distortion. For the G722.2, the loss of its energy mainly in the high frequencies removes brightness. Therefore, we also calculated the spectral centroids (SC) of each of the 20 speech samples as one of the indicators of timbral brightness or sharpness. The SC feature describes the center of gravity of the power spectrum and 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 P_i. The number N is the total number of frequencies of the SC feature *i.e.* equal to the number of Fourier coefficients which is also the frame size in number of samples. The SC explains if the spectrum is dominated by low or high frequencies. It is related to the perceptive dimension of timbre denoted as the sharpness of the signal [14]. The SC has been applied in classification of different audio samples [15] and for monophonic instrument recognition [16]. The SC feature was used, among other features, for hierarchical music genre classification in [17]. In [18], SC and bandwidth parameters are correlated with the attribute "Frequency content" of speech stimuli. The SC descriptors are calculated per frame and averaged over all frames for each of the 20 voiced samples. The Pearson correlation between coordinates along dimension 1 and the 20 averaged SC of tandems/codecs is -0.94 for male talker and -0.91 for female talker (correlations significant at the 0.05 level, in bold character in all tables). As in [12][18], SC descriptor explains dimension 1 characterized in our study by the attribute clear/muffle and in [12][18] by "frequency content".

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

Table 4 – Correlation coefficient between DIM2 coordinates and the power spectrum on silence for male (m) and female (f)

Then, we computed the power spectrum of each of the 20 resulting coded samples on the period of silence. Table 4 lists, for low band (LB, 50 – 4000 Hz) and high band (HB, 4000 – 7000 Hz), Pearson correlations values for both male and female talkers, between coordinates along dimension 2 and the power spectrum of each individual wideband sample. The high correlation values on higher band indicate that the SB-ADPCM G722 codecs that are at the "background noise " end of dimension 2 tend to have background noise which is more perceived on silence than other codecs that are at the opposite end. This audible noise generated by G722 codecs is due to SB-ADPCM coding technique of only 2 bits/sample quantization on higher band. Since dimension 2 separates only G722 ADPCM codec from all the other codecs, like Hall in [6] for low bit-rate coding, our study shows that ADPCM G722 codecs degradation is perceptually different from degradations of coding techniques such as ACELP (G722.2), transform-coding (MP3, HE-AAC), hybrid-coding (G729.1).
Subjects tend to associate dimension 3 with noise on speech signal. Table 5 lists Pearson correlation between coordinates and MOS scores. Each MOS score is an average on 24 other listeners for each sentence (female and male speakers) performed by the 20 wideband codecs. As "Naturalness" dimension in [6], this dimension 3 named "noise on speech signal" does not clearly present physical correlate, but is correlated with MOS (see Table 5). The same perceived quality is noted in [19] in MDS analysis of 20 female voices rated by speech pathologists and untrained listeners. This procedure resulted in a five-dimensional solution including voice quality.

	MOS (m)	MOS (f)
DIM1	-0.13	-0.36
DIM2	-0.19	-0.29
DIM3	**-0.69**	**-0.56**
DIM4	-0.15	-0.28

Table 5 – Correlation coefficient between coordinates and MOS scores for male (m) and female (f)

According to results presented in Table 5, MOSs decrease with dimension 3. As mentioned in [6], we know that, other things being equal, MOS would be expected to decrease with increasing noisiness. The configuration of the codecs reveals that, in each gathering of the codecs according to technical specifications, the factor bitrate is monotonic along dimension 3 (if we also take into account the tandem speech coding). This is confirmed when listening to different objects along this dimension 3 (HE-AAC or MP3 codecs seem more noisy than G722.1 codecs, for example). So, our "noise on speech signal" dimension seems similar to the dimension named "noisiness" in [6] and [12].

In Figure 1, dimension 4 is characterized by the fact that G729.1 codecs tend to be far apart from the other codecs. Subjects tend to associate dimension 4 with hiss. Hiss and reverberation are assumed to be related to occurrences of depth and width of peaks and notches of the power spectrum. Therefore, we try to find a link between the energy change and dimension 4 which may reveal hiss or reverberation. So, we estimate the maximum R_{xy} of the cross-correlation between the power spectra of the original signal x and each coded version y on speech periods, of which the occurrences would indicate correlation with DIM4. These maxima are calculated per 512 frame size, using a 1024-point FFT. In Table 6, Pearson correlation coefficients between coordinates along dimension 4 and the maxima R_{xy} of the cross-correlations are listed. $R_{xy}mean$, $R_{xy}max$, $R_{xy}std$ are respectively the mean, maximum and standard deviation over all frames.

	$R_{xy}mean$	$R_{xy}max$	$R_{xy}std$
DIM4 (m)	**-0.83**	**-0.90**	**-0.90**
DIM4 (f)	**0.59**	**0.56**	**0.67**

Table 6 – Correlation coefficient between coordinates along DIM4 and the correlation of power spectra for male (m) and female (f)

The significant Pearson correlations values for male and female talkers seem to indicate that the structure of spectra changes along the dimension 4. The configuration of G729.1 codecs on one extremity of dimension 4 suggests that this dimension tends to be associated with hiss.

6. CONCLUSION

The aim of this study was to link the current coding technologies with their potential perceptive degradations, in order to be able to artificially and quantitatively generate these degradations, and so to create a reference quality system for subjective tests.

We presented a brief review of multidimensional scaling techniques, including INDSCAL, weighted multidimensional scaling technique that can be applied to pair dissimilarity judgements to determine a multidimensional perceptive space in which auditory stimuli can be represented.

We presented results from a listening experiment in which subjects gave dissimilarity judgements on pairs of speech samples coded by 20 wideband tandem codecs. The resulting dissimilarity matrices were processed by INDSCAL MDS to generate stimulus. Coded speech samples for one male talker are therefore represented in a 4-dimensional perceptive space characterized by different attributes such as clear/muffle, background noise, noise on speech, hiss. Similar analysis for the female talker leads to similar conclusions.

The first two perceptive dimensions are characterized by physical correlates: spectral centroid for dimension 1 and the power spectrum on silence period for dimension 2. The cross-correlation between the spectrum of the original signal and this one of the coded versions characterizes dimension 4. Dimension 3 can be hardly characterized using physical correlates, but is correlated with MOS. These results are compared to other perceptive spaces underlined by MDS from the domain of speech quality evaluation.

We related these perceptive characteristics with coding techniques. This result is consistent with our objective to link these perceptive characteristics to various coding techniques. These degradations should now be modeled, and calibrated to artificially generate these degradations and to introduce them into speech signals.

REFERENCES

[1] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality". International Telecommunications Union, 08/96.
[2] ITU-T Recommendation P.810, "Modulated Noise Reference Unit (MNRU)". International Telecommunications Union, 02/96.
[3] T. Etame, G. Faucon, R. Le Bouquin Jeannes, L. Gros and C. Quinquis, "Characterization of the multidimensional perceptive space for current speech and sound codecs," AES 124th convention, Amsterdam, The Netherlands, May 17-20, 2008.
[4] C. Osgood, "The nature and measurement of meaning," Psychological Bulletin, 49 (197-237), 1952.
[5] V.V. Mattila, "Ideal point modelling of the quality of noisy speech in mobile communications based on multidimensional scaling," AES 114th convention, Amsterdam, The Netherlands, March 22-25. 2003.
[6] J.L. Hall, "Application of multidimensional scaling to subjective evaluation of coded speech," J. Acoust. Soc. Am. 110 (4), Oct. 2001.
[7] J.B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika, Vol. 29, pp. 1-27, 1964.
[8] J.B. Kruskal, "Nonmetric multidimensional scaling: a numerical method," Psychometrika, Vol. 29, pp. 115-129, 1964.
[9] J.D. Carroll, "Individual differences and multidimensional scaling," R.N. Shepard, A.K. Romney & S.B. Nerlove (Eds.) Multidimensional Scaling: Theory and Applications in the Behavioral Sciences (Vol.1, pp.105-155). New York and London: Seminar Press, 1972.
[10] J.D. Carroll and J.J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," Psychometrika, 35, pp. 283-319, 1970.
[11] A. Raake, *Speech Quality of VoIP - Assessment and Prediction*. Pub, U.K.: Wiley, 2006.
[12] M. Wältermann, K. Scholz, A. Raake, U. Heute, S. Möller, "Underlying quality dimensions of modern telephone connections," Interspeech, paper 1089-Wed3FoP.11, 2006.
[13] D. Pascal, "Comparative Performances of Two Subjective Methods for Improving the Fidelity of Speech Signals," Preprint 84th Audio Engineering Society (AES) Convention, Paris, March 1988, (Paper L-3, Preprint 2639).
[14] ISO/IEC FDIS 15938-4:2002(E), *MPEG-7*. Information technology - multimedia content description interface, part 4 : Audio, 2003.
[15] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio," *IEEE*, 3(3):27–36, 1996.
[16] K. Martin, "Towards automatic sound source recognition: Identifying musical instruments," in *NATO Computational Hearing Advanced Study Institute*, 1998.
[17] J.J. Burred and A. Lerch, "A hierarchical approach to automatic musical genre classification," in *Proceedings of the 6th International Conference on Digital Audio Effects*, London, Sept. 2003.
[18] K. Scholz, Kirstin, M. Wältermann, L. Huo, A. Raake, S. Möller, and U. Heute, "Estimation of the quality dimension "directness/frequency content for the instrumental assessment of speech quality," *Interspeech*, paper 1219-Wed1A3O.6, 2006.
[19] B. Geiser and al., "A multidimensional scaling study of voice quality in females," *Phonetica*, 50(1), 15-27, 1993.