ELECTRONIC VISION

MDS codec evaluation based on perceptual sound attributes

Marcelo Herrera Martínez[*]
Edwar Jacinto Gómez[**]
Edilberto Carlos Vivas G.[***]

Abstract:

The present paper deals with the subjective evaluation of audio coding technologies using the "Similarity Rating" psychometric method. Compressed audio excerpts are presented to a group of experienced listeners by pairs of stimuli. Each pair represents a different type of distortion, with compression included. These types of distortion correspond to three psychoacoustic attributes, sharpness, roughness, and fluctuation strength. Listener's task is to evaluate the degree of similarity between the pairs. The final output is the localization of each measured codec in this psychoacoustic space. Multidimensional scaling is the statistic technique, belonging to Multivariate Statistical methods, which allows processing the results obtained from this method.

Key words:

Psychometric method, Similarity Rating, Psychoacoustic attributes Sharpness, Roughness, Fluctuation Strength, and Multidimensional Scaling (MDS).

1. Introduction

The subjective evaluation of audio compression systems has become an area of great interest in recent years. The purpose is to find a tradeoff between transmission bit rates and Quality of Service "QoS". Objective tests fail to evaluate these systems since they base their models of evaluation on the same prin-

---

[*] B.Sc. In Electronic Engineering and PhD. Degree at the Czech Technical University in Prague. Current position: professor Universidad San Buenaventura, Bogotá (Colombia). E-mail: mherrera@usbbog.edu.co

[**] B.Sc. In Electronic Control and Instrumentation Engineering: Universidad Distrital Francisco José de Caldas, Bogotá (Colombia). Current position: professor Universidad Distrital Francisco José de Caldas, Technological Faculty, Bogotá (Colombia). E-mail:edwarjg@udistrital.edu.co

[***] Msc. Industrial Automation, Universidad Nacional de Colombia (Colombia). Current position: professor Universidad San Buenaventura, Bogotá (Colombia). E-mail: evivas@usbbog.edu.co

ciples as the models of the systems which are under evaluation. For that reason, audio institutions have carried out tests based on the ITU-Recommendation, namely using DBTS (Double-Blind Triple Stimulus with hidden reference method) belonging to the successive categories methods [1]. This evaluation has also been carried out by the author in order to supply target values to implement an objective method with the help of ANN Networks. The present paper proposes a new system of evaluation based on annoyance rating, which is in fact the phenomenon evaluated in DBTS and provides no complex evaluation of the compressed signal. An evaluation taking into account that the audio-perceptual task is a complex one, has several dimensions of a multidimensional space, which seems to be more approximate to reality. In the present paper there are several psychoacoustic attributes which have not been evaluated, such as tonality, tonalness and noisiness, among others.

2. Perceptual space formed from typical distortion types

The investigated psychoacoustic space is built from three psychoacoustic attributes. Each of them is related to tonal and noisy signals.

2.1 Sharpness

Sharpness is a sensation that describes auditory perception related to the prevalence of high frequency components in a sound. Two of the parameters that influence sharpness are the spectral content and the centre frequency of narrow-band sounds. Sharpness is expressed in acums. Bandwidth is the other variable that strongly influences sharpness. Mathematical expressions for sharpness can be found in [2].

2.2 Roughness

Using a 100% amplitude-modulated 1-kHz tone and increasing the modulation frequency from low to high values, three different areas

of sensation are obtained. At low modulation frequencies fluctuation sensation is produced. At about 10 Hz, roughness starts to increase. Roughness is a complex effect that quantifies the subjective perception of rapid (15-300 Hz) amplitude modulation of a sound. Roughness can be expressed in aspers.

2.3 Fluctuation Strength

As mentioned above, at low modulation frequencies, fluctuation strength sensation is perceived when modulating a 1kHz tone with 100% amplitude modulation.

3. Psychoacoustic test using Similarity Rating methodology

In psychoacoustic applications, it is very common to obtain the mutual similarity of objects, or the inverse variable dissimilarity, δ. This model gives rise to the so called distance model, based on the concept of distance d. This distance d between points of the model configuration represents the obtained experimentally dissimilarity δ. It requires, therefore, the dissimilarity values as the input data for the analysis.

In this work, for the subjective measurement of similarities of acoustic stimuli, we use direct comparison of all pairs of stimuli. The listener's task is to express the degree of similarity between each pair in a definite scale, and where its end points are in a bipolar scale, namely "perfect similarity" and "minimal similarity".

Observations made for a pair of objects, i and j, results in a proximity value p_{ij}. The term proximity is used in a generic way to denote both similarity and dissimilarity.

3.1 Compression parameters

All the original tones and noises from [3] were compressed at a compression bit rate of 80 kbit/s. Lower bit rates were not taken

into account since for strong compression it may be difficult to isolate the mentioned psychoacoustic attributes. Higher bit rates were also discarded, because the perception of compression artifacts becomes poorer.

The compression technologies evaluated were MP3-Fraunhoffel, Ogg Vorbis, WMA and AAC-MP4.

3.2 Excerpts

Eight stimuli were used in the test:

1) Sharp broad band noise
2) Sharp tonal noise
3) Amplitude modulated white noise
4) Amplitude modulated 1000 Hz tone.

The modulation frequency of 3 and 4 was set to 70 Hz.

For the next two stimuli, corresponding to the sensation of Fluctuation Strength, the type of modulation is the same (amplitude modulation), but the modulation frequency is 4 Hz. 5) and 6) therefore correspond to 3) and 4) but at different modulation frequencies. Modulation index in both cases is set to m=100%
7) Compressed tone
8) Compressed noise.

3.3 Pair-wise comparison

N empirical objects (types of distortion) are investigated. Pairs of psychoacoustic affected tones noises/ compressed tones noises were formed, as shown in Table 1:

Table. 1. Pair-wise comparison. Listener's task is to evaluate the degree of similarity between the pair of acoustic stimuli

|	
Sharp-Compressed	Sharp-Rough
Compressed-Rough	Sharp-Fluctuation
Rough-Fluctuation	Compressed-Fluctuation
Compressed-Sharp	Rough-Compressed

Source: own elaboration

where S T-N, Sharp Tone/Noise., R T-N, Rough Tone/Noise., F T-N, Fluctuation Strength Tone/Noise., C T, Compression Tone/Noise

There are therefore eight evaluated objects.

Typical excerpts from each type of distortion were obtained from [3]. Therefore, an excerpt characterizing a sharp, rough and fluctuating tones/noises were obtained from this source. These excerpts were used for the evaluation, comparing them to the original tones/noises affected by compression. A horizontal scale for determining the degree of similarity is given to the listener. Listener's task is to express the degree of similarity between the given pairs of acoustic-stimuli within the frame of this scale, Fig 1.

Figure 1. Similarity scale of the experiment
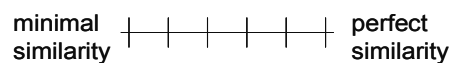
Source: own elaboration

The similarity rating methodology must be further evaluated by the MDS (Multidimensional Scaling) Method in order to find a geometric interpretation and description of the problem. In our case, the problem focuses in finding the geometric interpretation of the position of the compression artifact perception inside the psychoacoustic space mentioned.

4. Multidimensional Scaling

The objective of using the Multidimensional Scaling Technique is to discover the hidden structure of the measured values of a multidimensional variable or phenomenon, namely compression distortion in this case, and to express it as a geometric space

model, inside which each measured value is represented by a point localized in such a way that the geometric relations inside the point configuration correspond to the most important characteristics of the object.

In a MDS problem, we encounter the following aspects of the experimental problem:

1) A definite number of empirical objects n, namely four acoustic stimuli in the present study.

2) If an observation has been made for a pair of objects, i and j, a proximity value p_{ij} is given. The term proximity is used to denote both similarity and dissimilarity values.

3) A dissimilarity is a proximity that indicates how dissimilar two objects are, and is denoted by δ_{ij}.

4) X denotes (a) a point configuration (a set of n points in m-dimensional space) and (b) the $n \times m$ matrix of the coordinates of the n points relative to m Cartesians coordinate axes.

5) The Euclidean distance between any two points i and j in X is the length of a straight line connecting points i and j in X.

The term $f(p_{ij})$ denotes a mapping of p_{ij}, that is, the number assigned to p_{ij} according to rule f. This is sometimes written as

$f : p_{ij} \rightarrow f(p_{ij})$. We also say that $f(p_{ij})$ is a transformation of p_{ij}. Instead of $f(p_{ij})$ we often write $\hat{d}_{ij}$.

The error of the representation will be

$$e_{ij}^2 = (d_{ij} - \delta_{ij})^2 \qquad (1)$$

Summing the last equation over i and j yields the total error (of approximation) of an MDS representation, namely

$$\sigma_r(X) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (d_{ij} - \delta_{ij})^2 \qquad (2)$$

for all avalaible δ_{ij}, which is often written as

$$\sigma_r(X) = \sum_{i<j} (d_{ij} - \delta_{ij})^2 \qquad (3)$$

for all avalaible δ_{ij}.

Our goal, is to mathematically minimize $\sigma_r(x)$ over (X).

The basic non-metric distance model MDS uses Euclidean metrics. The distance d_{jk} between points j and k in a r-dimensional space is defined by,

$$d_{jk} = \sqrt{\sum_{t=1}^{r} (x_{jt} - x_{kt})^2} \qquad (4)$$

Where x_{jt} and x_{kt} are the points coordinates in dimension $t = 1, 2, ..., r$ The objective of the non-metric analysis MDS is to find an MDS solution in an Euclidean space of the least number of dimensions r, in which the mutual dimensions d_{jk} correspond to the dissimilarity values δ_{jk} in a monotonic relation.

The mapping of the measured values onto a geometric psychoacoustic space should fulfill the so-called "admissible transformations" in order to meet requirements such as "invariance", in the sense that every rotation of the geometric body should not change the overall configuration (i.e. distance between points should be conserved).

4.1 Minimizing the stress function

Stress Function as its metric, namely the squared error of the representation,

$$e^2 = \left| f - d(x) \right|^2 \qquad (5)$$

MDS models require to be evaluated due to an existence of noise in their models. Empirical proximities always contain noise due to measurement imprecision, unreliability, sampling effects, and so on.

Summing e_{ij}^2 over all pairs (i, j), raw stress is obtained

$$\sigma_r = \sigma_r(X) = \sum \left[f(p_{ij}) - d_{ij}(X) \right]^2 \qquad (6)$$

The raw stress is normalized, then,

$$\sigma_1 = \sqrt{\frac{\sum \left[f(p_{ij}) - d_{ij}(X) \right]^2}{\sum d_{ij}^2(X)}} \qquad (7)$$

Regression of the proximities onto the distances computed on X requires the use of linear regressions. The regression yields transformed proximities, $f\left(p_{ij}\right)$s, that are "approximated distances" or "d-hats", also referred to as disparities in the MDS-literature.

An observed proximity, after transformation by f, is thus conceived as

$$f(p_{ij}) = d_{ij}^{(e)} = \sum \left[x_{ia}^{(e)} - x_{ja}^{(e)} \right)^2 \right]^{\frac{1}{2}} \qquad (8)$$

Where

$$x_{ia}^{(e)} = x_{ia} + e_{ia} \qquad (9)$$

And e_{ia} is a value from the random distribution of point i.

4.2 Proximities

Proximities are collected by directly judging the (dis-) similarity of pairs of objects; they may also be derived from score or attribute vectors associated with each of these objects.

A "simple" distance of any two objects, i and j, in the m-dimensional attribute space can be given by the city-block distance,

$$d_{ij}^{(1)}(X) = \sum_{a=1}^{m} \left| x_{ia} - x_{ja} \right| \qquad (10)$$

Where i and j are two objects of interest, and x_{ia} and x_{ja} are the scores of these objects on attribute a. Euclidean distances are less attractive for deriving proximities because they involve some kind of weighting of the intra-attribute differences $x_{ia} - x_{ja}$.

A list of measures of proximities derived from attribute data can be found in [4]

5. Finding the coordinates of the model

The basic idea of classical scaling is to assume that the dissimilarities are distances, so that the coordinates that explain them can be found. The squared distances are computed from X by

$$\underline{D}^{(2)} = \underline{c}1' + \underline{1}\underline{c}' - 2\underline{X}\underline{X}' = \underline{c}1' + \underline{1}\underline{c}' - 2\underline{B} \qquad (11)$$

Where c is the vector with the diagonal elements of X'. Multiplying the left and the right sides by the centering matrix $\underline{J} = \underline{I} - \underline{n}^{-1}\underline{1}\underline{1}'$ and by the factor -1/2 yields

$$-\frac{1}{2}\underline{J}\underline{D}^{(2)}\underline{J} = -\frac{1}{2}\underline{J}(\underline{c}1' + \underline{1}\underline{c}' - 2\underline{X}\underline{X}')\underline{J} \qquad (12)$$

$$-\frac{1}{2}\underline{J}\underline{c}1'\underline{J} - \frac{1}{2}\underline{J}\underline{1}\underline{c}'\underline{J} + \frac{1}{2}\underline{J}(2\underline{B})\underline{J} \qquad (13)$$

$$-\frac{1}{2}\underline{J}\underline{c}0' - \frac{1}{2}\underline{0}\underline{c}'\underline{J} + \underline{J}\underline{B}\underline{J} = \underline{B} \qquad (14)$$

The first two terms are zero, because centering a vector of ones yields a vector of zeros $\left(\underline{1}'\underline{J} = \underline{0} \right)$. The centering around B can be removed because X is column centered, and hence so is B. To find the MDS coordinates from B, we factor B by eigen-decomposition,

$$\underline{Q}\underline{\Lambda}\underline{Q}' = (\underline{Q}\underline{\Lambda}^{1/2})(\underline{Q}\underline{\Lambda}^{1/2})' = \underline{X}\underline{X}' \qquad (15)$$

Summarizing, the steps for the calculation are as follows:

1. Compute the matrix of squared dissimilarities $\underline{\Delta}^{(2)}$.

2. Apply double centering to this matrix:

$$\underline{B}_\Delta = -\frac{1}{2}\underline{J}\underline{\Delta}^{(2)}\underline{J} \qquad (16)$$

3. Eigen-decomposition computation

$$\underline{B}_\Delta = \underline{Q}\underline{\Lambda}\underline{Q}' \qquad (17)$$

4. Let the matrix of the first m eigenvalues greater than zero be $\underline{\Lambda}_+$ and $\underline{Q}_+$ the first m columns of Q. Then, the coordinate matrix of classical scaling is given by

$$\underline{X} = \underline{Q}_+\underline{\Lambda}_+^{1/2} \qquad (18)$$

If $\underline{\Delta}$ happens to be a Euclidean distance matrix, then classical scaling finds the coordinates up to a rotation. Note that the solution $\underline{Q}_+\underline{\Lambda}_+^{1/2} = \underline{X}$ is a principal axes solution.

Classical scaling minimizes the loss function

$$L(\underline{X}) = \left\| -\frac{1}{2}\underline{J}[\underline{D}^{(2)}(\underline{X}) - \underline{\Delta}^{(2)}]\underline{J} \right\|^2 \qquad (19)$$

$$= \left\| \underline{X}\underline{X}' + \frac{1}{2}\underline{J}\underline{\Delta}^{(2)}\underline{J} \right\|^2 \qquad (20)$$

$$= \left\| \underline{X}\underline{X}' - \underline{B}_\Delta \right\|^2, \qquad (21)$$

sometimes called Strain.

6. Data collection [5]

The distance matrix D is collected from the dissimilarity values, rated by the listeners on the scale of Figure 1. Every listener was instructed to rate the degree of similarity between the two types of distortion. Sound excerpts were obtained from [3]. Parameters of the sound reproduction are in accordance to [6].

7. Results: MDS Calculation

Dissimilarity data from the experiment is organized in the dissimilarity data,

$$\underline{\Delta} = \begin{bmatrix} 0 & 4.05 & 8.25 & 5.57 \\ 4.05 & 0 & 2.54 & 2.69 \\ 8.25 & 2.54 & 0 & 2.11 \\ 5.57 & 2.69 & 2.11 & 0 \end{bmatrix}$$

so that

$$\underline{\Delta}^{(2)} = \begin{bmatrix} 0.00 & 16.40 & 68.06 & 31.02 \\ 16.40 & 0.00 & 6.45 & 7.24 \\ 68.06 & 6.45 & 0.00 & 4.45 \\ 31.02 & 7.24 & 4.45 & 0.00 \end{bmatrix}$$

The second step is to compute

$$\underline{B}_\Delta = -\frac{1}{2}\underline{J}\underline{\Delta}^{(2)}\underline{J}$$

$$\underline{J} = \begin{bmatrix} \dfrac{3}{4} & -\dfrac{1}{4} & -\dfrac{1}{4} & -\dfrac{1}{4} \\ -\dfrac{1}{4} & \dfrac{3}{4} & -\dfrac{1}{4} & -\dfrac{1}{4} \\ -\dfrac{1}{4} & -\dfrac{1}{4} & \dfrac{3}{4} & -\dfrac{1}{4} \\ -\dfrac{1}{4} & -\dfrac{1}{4} & -\dfrac{1}{4} & \dfrac{3}{4} \end{bmatrix}$$

Yielding the result for

$$\underline{B}_\Delta = \begin{bmatrix} 20.52 & 1.64 & -18.08 & -4.09 \\ 1.64 & -.83 & 2.05 & -2.87 \\ -18.08 & 2.05 & 11.39 & 4.63 \\ -4.09 & -2.87 & 4.63 & 2.33 \end{bmatrix}$$

In the third step, we compute the eigen-decomposition of $\underline{B}_\Delta$; that is, $\underline{B}_\Delta = \underline{Q}\underline{\Lambda}\underline{Q}'$ with

$$\underline{Q} = \begin{bmatrix} .77 & .04 & .50 & -.39 \\ .01 & -.61 & .50 & .61 \\ -.61 & -.19 & .50 & -.59 \\ -.18 & .76 & .50 & .37 \end{bmatrix}$$

And

$$\underline{\Lambda} = \begin{bmatrix} 35.71 & 0 & 0 & 0 \\ 0 & 3.27 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5.57 \end{bmatrix}$$

There are two positive eigenvalues, one zero eigenvalue due to the double centering and one negative eigenvalue. The task then leads to two dimensions in the Euclidean space. The configuration X is found by

$$\underline{X} = \underline{Q}_+ \underline{\Lambda}_+^{1/2}$$

$$= \begin{bmatrix} .77 & .04 \\ .01 & -.61 \\ -.61 & -.19 \\ -.18 & .76 \end{bmatrix} \begin{bmatrix} 5.98 & 0 \\ 0 & 1.81 \end{bmatrix} = \begin{bmatrix} 4.62 & 0.07 \\ 0.09 & -1.11 \\ -3.63 & -.34 \\ -1.08 & 1.38 \end{bmatrix}$$

Therefore, the MDS solution of the task yields a final result in the form of a 2-dimensional perceptual space with resulting X-coordinates (position of the four objects relative to this 2-principal axis) given by the last matrix. Therefore, we can infer that the compression phenomenon can be partially explained in terms of the three psychoacoustic attributes investigated in the experiment.

The 2-dimensional perceptual space of compressed tones, related to sharpness, roughness and fluctuation strength of tone, is depicted in Figure 2. Figure 3 shows the same description for a noise signal.

Figure 2. S1: MP3 compressed tone, S2: Ogg, S3: WMA, S4: AAC, S5: Sharp tone, S6: Fluctuation strength tone and S7: rough tone

2-D position of the stimuli
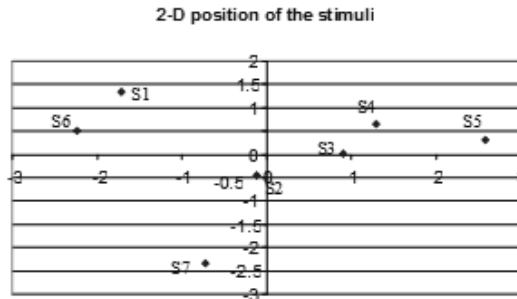
Source: own elaboration

An MP3 compressed tone is located perceptually close to the fluctuated strength tone, as depicted in Figure 3. The sharp tone is located closer to the AAC and WMA technologies. The Tone compressed using OGG Vorbis is positioned in the middle of the perceptual space. Therefore it can be stated that, for the tonal case, MP3 has a fluctuating character, and AAC and WMA a sharp character. OGG remains in a neutral point between the psychoacoustic attributes investigated.

Figure 3. S1: MP3 compressed noise, S2: Ogg, S3: WMA, S4: AAC, S5: Sharp noise, S6: Fluctuation strength noise and S7: rough noise. In both axes, perceptual distances for each dimension are shown

2-D Position of the stimuli
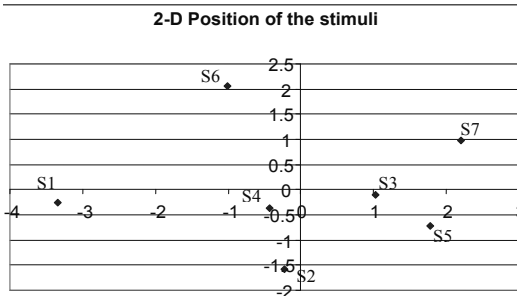
Source: own elaboration

WMA is perceptually close to the sharp noise, as depicted in Fig. 3. WMA is also perceptually close to the rough noise. OGG Vorbis is situated opposite to the fluctuation strength noise, but perceptually close to AAC and WMA compression.

8. Conclusions

In the present work, a subjective psychoacoustic evaluation of audio-coding technologies was performed using the method of "Similarity Ratings". The perceptual multidimensional structure of eight types of distortion, among which compression is included, was searched.

Results show that after the MDS evaluation, the four tonal objects have a dimensionality of two, thus perceptual space reduction was achieved by MDS. A similar reduction of dimensionality might be performed using other multivariate methods such as Principal Component Analysis, Correspondence Analysis, Discriminant Analysis, Cluster Analysis and others, but the author selected the MDS solution since it yields an elegant solution. Future work should deal also with other attributes such as noisiness, tonality and tonalness in order to enhance the psychoacoustic perceptual space proposed, [7].

References

[1] ITU-R 1116-1 Recommendation, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems", 1994-1997.

[2] F. Zwicker, "*Psychoacoustics: Facts and Models*", Springer, 1990.

[3] University of Salford Manchester, "Sound quality making products sound better". Available in: http://www.acoustics.salford.ac.uk/research/arc/cox/sound_quality/html.

[4] I. Org, P.J.F. Groenen, "*Modern Multidimensional Scaling, Theory and Applications*". New York: Springer, Second Edition, 2007.

[5] P. Hebak,. et al, "*Vícerozmêrné statistické metody 2*", Vícerozmêrné statistické metody, Informatorium, 2005.

[6] M. Herrera, "*Summary of the subjective evaluation of audio-coding testing at the CVUT during the period 2003-2005*". In XI International Symposium of audio and video. Krakov (Poland), 2005.

[7] A. Melka, "*Zaklady experimentalni psychoakustiky*". HAMU, 2005.