ACOUSTICS 2012

Noise/Signal: two different listening experiences for deterministic and non deterministic sound stimulus

R. C. Borges

Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 - Pampulha, 31270-901 Belo Horizonte, Brazil
rborges@cpdee.ufmg.br

The duality between signal and noise is widely observed in the scientific context of the 20th century. Noise is retrospectively associated to nuisance, annoyance, and was even subjectively defined as a non-signal. Definition, anyway, takes noise away from its original meaning, turns it into signal and keeps this duality existing through time. This work treats the subject as a matter of perception, more specifically, as a matter of two different listening experiences for deterministic and non deterministic sound stimulus. People with trained ears were asked to freely choose adjectives for pairs of sounds took from a group of: three different probability distributed white noises, a pink noise, a Brownian noise, a square, a square with aliasing effect, a sawtooth, and a pure sine wave. Twenty seven acoustic descriptors were extracted from the samples, from spectral kurtosis to dissonance level. The results were submitted to factorial analysis for finding the best descriptors when separating both groups of sounds, and which physical parameters are correlated to perceptive ones. The results points 'diffuse/concentrated' as the most successful adjectives for separating signal from noise, the salience of a fundamental frequency as a determinant for the distinction, and some correlation between subjective and objective data.

1 Introduction

Noise seems to be a key term when trying to understand the scientific context of the 20th century, once assumed that generic signals are only coherent when presented under some desorganized background. Signals are ordered phenomenons, noises are not.

Communication channels aims to transmit messages, but noise is what defines its boundaries. Noises probably came first then signals, and also because of this, limits the way humans perceive the world.

As a sound element, noise is widely complex. Its meaning is subjective and its perception varies historically, socially and culturally. Noise is intuitively associated to denial, and retrospectively defined as non-information, non-music, nuisance, annoyance. In other words, noise is defined not as a signal.

According to Moles there is no difference in terms of absolute structure between perturbations and signals: "signals and noises have the same nature, and the only logical and appropriate difference that can be established between them have to be based in the conception of intention from the emitter part: noise is a signal which has no intention of being transmitted." [1]

From a objective point of view, deterministic processes are the ones that can have their future behavior determined by a finite number of measurements; non deterministic processes, on the contrary, are the ones that can not be determined [2]. Signals, then, may be completly random, but will be here assumed as deterministic sound elements. Noises are usually unpredictable, and are here represented as non deterministic phenomenons.

Papers dealing with noise as a sound element were already published [3] [4], and some were even about its relation with music [5], but none of these focused on noise as playing a role in human sound perception.

This work tries to discuss the way people listen to the differences between these two classes of sounds, deterministic and non deterministic, and which are the physical features of their waves that are correlated to this distinction.

The paper is structured as follows. In section 2 the sound samples are presented and shortly described. Both methods for feature extraction are presented in section 3; listening tests were used for extracting perceptive parameters, and a computational tool was applied for extacting the physical ones. The results were submited to principal components analysis in both cases and are presented in section 4, as well as the correlation between them. The last section discuss the results and propose some future work.

2 Presentation

This work makes use of: uniform white noise, gaussian white noise, laplacian white noise, pink noise, brownian noise, sine wave, sawtooth wave, square wave and a square wave with aliasing effect samples.

White noise is perfect noise. It has all frequency contents ocurring with a flat power spectral density, which turns its existence only possible in theory: an infinite energy would be necessary to represent all these frequencies in the real world [6]. White noise should also be mentioned as having no memory, each of its hiss is completly uncorrelated and independent of the previous one. Doesn't matter how far they are from each other, they don't contribute with any knowledge about the past of the signal.

White noises can have any probability distribution and this reflects in the incidence of the frequency contents. If this is uniformally distributed, all values have equal probability of ocurring; if its gaussian distributed, this should present major part of its content around the mean, following a gaussian curve; and if it has laplacian distribution, the same should occur but according to a laplacian curve.

Pink and Brownian noises have their spectral energy density decreasing in a inverse proportion to the frequency bin number, the first decreases as $1/f$, the second $1/f^2$ [7]. When compared to the sound of the white noise, these two 'colored' noises presents enhanced low frequency spectrum content, and may give the listener the sensation of a lower sound.

The sine wave, as well known, has no superior harmonics; the sawtooth have all of them, with magnitude inversely proportional to the number of the harmonic; and the square wave have only the odd partials [8]. Aliasing effect is a recurrent effect in the signal process field, which is closely related to the frequency used for sampling the signal. It introduces perceptual content in the digitalized signal that were not present in the original one.

Non deterministic samples were sinthesized departing from their probabilty distribution curves with MATLAB software, and the deterministic ones were synthesized with Audacity software, with a 1 KHz fundamental frequency. All the sounds have 2 seconds of duration.

3 Feature Extraction

3.1 Listening Tests

Subjective listening tests are normally realized with the aim of obtaining information about how people listen to an specific group of sounds [9]. The process of presenting adjectives to the listeners for them to associate to each sound have great potential of influencing the results, and this is why, in this experiment, people were free to choose whatever description they wanted according to a method called *Repertoty Grid Method* [10].

The sounds were pesented to a group of 10 people from Brasil, Colombia, Chile, Spain and Russia, between 22 and 35 years old, 6 female and 4 male, all involved with some activity related to music. They were informed with the fewest possible information about the tests, and were asked to describe the sounds with adjectives in the most natural way they could. The only restriction imposed was that the elected term had to be related to the timbre of the sound.

In a first part of the process pairs of sounds were elected, each one necessarily belonging to a different group (deterministic and non deterministic), and the listener were asked to describe the difference between them by choosing an adjective as well as its opposite. Once the description had already been made, each of the 9 samples were rated in a scale from 1 to 5, varying from the adjective to its opposite. This was repeated 5 times for each volunteer resulting in 50 pairs of adjective for the 9 sound samples.

The tests were applied at the Music and Technology Group, and at the University of Barcelona, Spain, and took approximately 30 minutes each.

The results were all translated to english and were combined in such a way that when identical descriptons pairs (adjective/opposite) were choosen, the rates were summed. The matrix with the data was normalized for the following procedures.

3.2 Acoustic Parameters

The physical features of the samples were extracted by using a tool for retrieving acoustical parameters of music files called *Essentia* [11], and only the ones that make reference to the timbre of the sounds were maintained for analysis. They were all calculated with a sampling frequency of 44,1 KHz, taking 2048 samples in each iteration, and sucessives frames should coincide 1024 samples for better resolution in transitions. The 27 features are listed below:

1. Barkbands Kurtosis: Bark bands are slices of the human hearing spectrum calculated trought a psychoacoustical scale called Bark scale. Kurtosis is a probabilistic measure which provides information about the behavior of the variable around its mean value. Negative kurtosis indicates flat bark bands, positive kurtosis indicates peakier ones, and null values indicates normal distributions. This are generally used for timbre description.

2. Barkbands Skewness: Skewness measures the assimetry of a distribution around its mean value. A negative skewness indicates bark bands with energy concentrated in the high frequencies, a positive skewness indicates bark bands with energy concentrated in low frequencies, and a null value indicates symmetric distributions. For silence and constant sounds the skewness has null values. This are generally used for timbre description.

3. Barkbands Spread: Spread is defined as the variance of a distribution around its mean value. Is the same as the second central moment and used for timbre classification.

4. Dissonance: Perceptual descriptor used for measuring the roughness of the sound, based in the fact that two sinusoidal spectral components share a dissonance curve which, values are dependent on their frequency and amplitude relations. The total dissonance value id derived by summing up the values for all components (i.e. spectral peaks) in a sampling window. This is generally used for sound segmentation.

5. Hfc: The high frequency content is a simple measurement, taken from the singal spectrum (usually a STFT spectrum) which can be used to characterize the amount of high frequency content present in a signal. In contrast to perceptive measures, it is not based in any evidency about its relevance to human hearing. Despite that, it can be usefull for some applications, like sound event detection for example.

6. Pitch: Is represented as the fundamental frequency of the analysed sound. This is calculated through the frequency spectrum for monophonic signals.

7. Pitch Instantaneous Confidence: A measure of the confidence of the pitch calculated for the signal. Provides the evidence about how much a pitch, calculated in a single sampling window, is affecting the whole spectrum. If its value is near to one, there is only one pitch value for the whole mixture, a zero value indicates multiple pitches indistinguishables.

8. Pitch Salience: The pitch salience is given by the relation between the highest peak and the zero difference peak of the autocorrelation function. Sounds without defined pitch have a mean value of relevance near to zero, while harmonic sounds presents values near to one. Unvarying pitch sounds have a small pitch salience variance while sounds with varying pitch have a high pitch salience variance.

9. Silence Rate 60dB: This is the rate of frames where the level is above a given threshold, here -60dB. Returns one whenever the instant power of the input frame is below the given threshold, zero otherwise.

10. Spectral Centroid: The spectral centroid is a measure used in digital signal processing to characterize an audio spectrum. It indicates where the ”center of mass” of the spectrum is.

11. Spectral Complexity: Timbral complexity is a measure of the complexity of the spectrum of the audio file. Typically, in a piece of audio several elements are present. This increases the complexity of the spectrum of the audio and therefore, it represents a useful audio feature for characterizing a piece of audio.

12. Spectral Crest: The crest is the ratio between the maximun value and the arithmetic mean of the spectrum. It is a measure of the noisiness of the spectrum.

13. Spectral Decrease: Extracts the decrease of an array of Reals (which is defined as the linear regression coefficient). The range parameter is used to normalize the result. For a spectral centroid, the range should be

equal to Nyquist and for an audio centroid the range should be equal to (audiosize - 1) / samplerate.

14. Spectral Energy: The spectrum energy at a given frame.
15. Spectral Energyband High: The Energy Band Ratio of a spectrum is the ratio of the spectrum energy from start cut off frequency to stop cut off frequency to the total spectrum energy. For the Energy Band Ration High, start Cut off Frequency = 4000Hz and stop Cut off Frequency = 20000Hz.
16. Spectral Energyband Low: The Energy Band Ratio of a spectrum is the ratio of the spectrum energy from start cut off frequency to stop cut off frequency to the total spectrum energy. For the Energy Band Ration Low, start Cut off Frequency = 20Hz and stop Cut off Frequency = 150Hz.
17. Spectral Energyband Middle High: The Energy Band Ratio of a spectrum is the ratio of the spectrum energy from start cut off frequency to stop cut off frequency to the total spectrum energy. For the Energy Band Ration Middle High, start Cut off Frequency = 800Hz and stop Cut off Frequency = 4000Hz.
18. Spectral Energyband Middle Low: The Energy Band Ratio of a spectrum is the ratio of the spectrum energy from start cut off frequency to stop cut off frequency to the total spectrum energy. For the Energy Band Ration Middle Low, start Cut off Frequency = 150Hz and stop Cut off Frequency = 800Hz.
19. Spectral Flatness dB: This is a kind of dB value of the Bark bands. It characterizes the shape of the spectral envelope. For tonal signals, flatness dB is close to one, for noisy signals it is close to zero.
20. Spectral Flux: Spectral Flux is a measure of how quickly the power spectrum of a signal is changing, calculated by comparing the power spectrum for one frame against the power spectrum from the previous frame. The spectral flux can be used to determine the timbre of an audio signal, or in sound event detection, among other things.
21. Spectral Kurtosis: The kurtosis gives a measure of the flatness of a distribution around its mean value. A negative kurtosis indicates a flatter signal spectrum. A positive kurtosis indicates a peakier signal spectrum. A kurtosis equal to zero indicates a spectrum with normal distribution.
22. Spectral RMS: The root mean square spectrum energy. This is used for measuring the loudness of the sound frame.
23. Spectral Rolloff: Computes the roll-off frequency of a spectrum. The roll-off frequency is defined as the frequency under which some percentage (cutoff), of the total energy of the spectrum is contained, 85% in this case. The roll-off frequency can be used to distinguish between harmonic (below roll-off) and noisy sounds (above roll-off).
24. Spectral Skewness: The skewness is a measure of the asymmetry of a distribution around its mean value. A negative skewness indicates a signal spectrum with more energy in the high frequencies. A positive skewness indicates a signal spectrum with more energy in the low frequencies. A skewness equal to zero indicates a symmetric spectrum. For silence or constants signal, skewness is zero.
25. Spectral Spread: The spread is defined as the variance of a distribution around its mean value. It is equal to the 2nd order central moment. Used for timbral characterization.
26. Spectral Strongpeak: The Strong Peak is defined as the ratio between the spectrum maximum magnitude and the bandwidth of the maximum peak in the spectrum above a threshold (half its amplitude). It reveals whether the spectrum presents a very pronounced maximum peak. The thinner and the higher the maximum of the spectrum is, the higher the value this parameter takes.
27. Zero Crossing Rate: The Zero Crossing Rate is the number of sign changes between consecutive signal values divided by the total number of values. A measure of the noisiness of the signal: noisy signals tend to have a high value.

4 Results

Seven pairs of adjectives were repeated during the listening experiments: 'dark/bright','deep/superficial', 'dirty/clean', 'full/empty', 'harsh/high', 'rounded/spiky', 'together/lonely'. The first referring to noises and the second to the deterministic signals. These are comun references people make when differentiating these two groups of sound.

The five pairs of adjectives that separate more accurately the two different classes of sound according to the method of analysis of variance (ANOVA) are presented in table 1.

Table 1: Pairs of adjectives choosen by ANOVA

Pairs of Adjectives	**p-valor**
Diffuse/Concentrated	4.597 e-05
Comfortable/Annoying	5.554 e-05
Salt/Sweet	6.839 e-05
Dark/Shiny	7.525 e-05
Neutral/Disturbing	8.565 e-05

91% of the total variance of the data of the listening tests are explained by the 3 first principal components, and 86% by the 2 first ones, according to Principal Component Analysis (PCA). The figure 1 shows these two principal components.

The first component (horizontal axis) mainly separates the two groups of sound, one from another, and the major loads are related, in order, to the adjective pairs: 'noisy/clean' (1.756e-01), 'neutral/disturbing' (1.749e-01), 'salt/sweet' (1.747e-01), 'harsh /high' (1.740e-01), and 'natural/artificial' (1.730e-01).

The second component (vertical axis) separates pink noise, gaussian white noise, uniform white noise, sawtooth wave and square wave with aliasing, from square wave without aliasing, brownian noise, laplacian white noise and sinewave. The highest loads that contribute for this separation are: 'contaminated/pure' (4.093e-01), 'atonal/tonal' (2.405e-01), 'harsh/sweet' (2.026e-01), 'dirty/clean' (1.865e-01), and 'bitter/sweet' (1.431e-01).

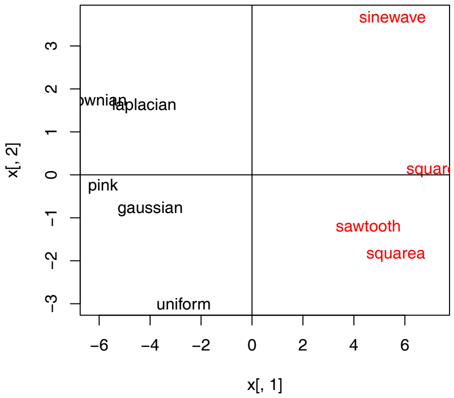

Figure 1: Two first principal components explaining 86% of the variance of the data obtained from the listening tests.

The five physical features that also separate more accurately the two different classes of sound according to ANOVA are presented in table 2.

Table 2: Acoustic features choosen by ANOVA

Acoustic Features	p-valor
Pitch Instantaneous Confidence	4.552 e-03
Spectral Spread	7.849 e-03
Spectral Decrease	3.655 e-02
Spectral Energyband Middle High	4.469 e-02
Spectral Energyband High	5.369 e-02

88% of the variance of the data obtained from the acoustical descriptors is explained in the 3 first principal components, and 75% is represented in the 2 first ones. The figure 2 shows these two principal components.

The first principal component of the objective measurements also separates mainly deterministic from non deterministic sounds, and the acoustical descriptors that contributes with highest loads are: 'spectral flatness dB' (8.630e-02), 'pitch instantaneous confidence' (8.533e-02), 'spectral complexity' (7.622e-02), 'zero crossing rate' (7.541e-02), 'spectral skewness' (7.394e-02).

The second component observed in figure 2 separates brownian noise, pink noise and sine wave, from white noises, sawtooth wave and square wave with aliasing effect. The descriptors with more influence were: 'spectral energyband low' (1.088e-01), 'spectral energyband middle low' (1.051e-01), 'high frequency content' (8.916e-02), 'spectral rolloff' (8.718e-02), and 'spectral decrease' (8.700e-02).

The 3 first principal components of the perceptive data were compared to each of the physical features and the 3 highest correlation degrees are shown in table 3

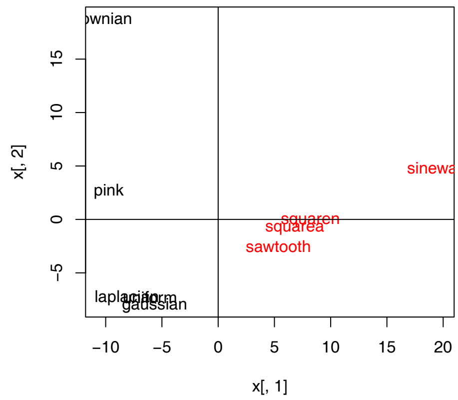

Figure 2: Two first principal components explaining 75% of the variance of the data obtained from acoustical analysis.

Table 3: Correlation between the first three perceptive components and acoustic descriptors.

	Descriptor	Correlation Index (mod)
PC1	Pitch Instant. Confidence	0.90
	Spectral Spread	0.77
	Spectral Energyband Middle Low	0.75
PC2	Spectral Flatness dB	0.67
	Spectral RMS	0.67
	Barkband Skewness	0.65
PC3	HFC	0.69
	Spectral Decrease	0.66
	Spectral Energyband Low	0.62

5 Conclusion

People do differentiate random sounds from coherent ones by listening to them, and they do this also using terms related to vision, touch and taste sensations, like 'dark/bright', 'rough/smooth' and 'salt/sweet'.

Noise is historically associated to annoyiance and nuisance, but the results here presented shows that people also elect positive terms for describing it. Adjectives like 'relaxing', 'pleasant', 'confortable' and 'natural' were attributed to noises when compared to synthesized deterministic sounds, which were described like 'irritanting' and 'strident'.

The principal component of the perceptive data associates mainly noises with terms like 'noisy' and 'harsh', but also associate the same ones to 'natural' or 'neutral'. The second component indicates pink noise, gaussian white noise and uniform white noise as beeing as 'atonal', 'harsh' and 'dirty', as sawtooth and square wave with aliasing effect. Brownian noise, laplacian with noise, square wave and sine wave are 'pure', 'clean', 'sweet' and 'tonal' according to these results (figure 1).

The first principal component referred to the objective data separates noisy sounds from harmonic ones through: 'spectral flatness', 'pitch instantaneous confidence', 'spectral complexity' and 'zero crossing rate'. These are acoustic descriptors developed for distinguishing complex spectrums from organized ones, which had their purpose validated here.

The second component for this type of data associates pink noise to sine wave and to brownian noise. This results that way specially because of spectrum energy location descriptors, two referred to low frequencys and one to the high spectrum content. Pink and brownian noises presents spectral power density decreasing when frequency increases, and sine wave don't have superior partials. White noises have flat spectrum energy, square and sawtooth waves are well known as having enhanced high frequency content. This may explain these results.

Grill et al. [10] achieved some correlation, for example, between 'high/low' adjectives and 'spectral centroid' descriptor (-0.69), between 'smooth/coarse' and 'spectral skewness' (-0.63), and between 'edgy/flowing' and 'spectral flatness' (-0.61) when using textural musical sounds as hearing stimulus. These are results that can not be compared with the ones presented here once the elected adjectives are different in each experiment, and were not directly compared to acoustic features.

'Pitch instantaneous confidence' is highly correlated to the most significant perceptual component, and was also the descriptor that presented the better p-value for separating coherent samples from random ones according to the objective analysis. This validates the instantaneous confidence of pitch as closely related to the perception of sound randomness, and as an efficient acoustic descriptor for distinguishing between synthesized deterministic and non deterministic sounds.

'Spectral flatness' and 'spectral root mean square' are weakly correlated to the second principal perceptive component that separate 'atonal', 'dirty' and 'bitter' sounds from 'tonal', 'clean' and 'sweet' ones.

Noise is a contemporary term used by engineers, physicists, economists, biologists and musicians, which can be discussed from many different points of view. This works tries some simple analysis in the subject by treating noise from the perceptual perspective. Future experiments should increase the number of participants as well as the number of sounds samples. The concepts here presented should also be expanded for enriching the discussion.

Acknowledgments

This work was partially funded by CAPES. Many thanks to Xavier Serra, Gil Sevalho and Renato Souza for the colaboration.

References

[1] A. Moles, *Teoría de la información y percepción estética*, 142-143, Ediciones Jucar, Madrid (1976)

[2] D. Middleton, IEEE Communications Society, *An introduction to statistical communication theory*, McGraw-Hill New York (1960)

[3] B. Uzkent, B.D. Barkana, J. Yang "Automatic environmental noise source classification model using fuzzy logic", *Expert Systems with Applications*, Elsevier, **38**, 8751-8755 (2011)

[4] V.T.K. Peltonen, A.J. Eronen, M.P. Parviainen, A.P. Klapuri, "Recognition of everyday auditory scenes: potentials, latencies and cues", *110th Audio Engineer Society Convention*, Amsterdam, Netherlands (2001)

[5] R.F. Vossa, J. Clarke, "1/f noise in music: Music from 1/f noise", *J. Acoust. Soc. Am.* **63**(1), 258 (1978)

[6] B. Kosko, *Noise*, Penguin Books (2006)

[7] R.F. Voss "1/f (flicker) noise: A brief review", *Proceedings of the 33rd Annual Symposium on Frequency Control*, 40-46 (1979)

[8] S. Tempelaars, *Signal processing, speech and music (studies on new music research)*, Routledge, 1st Edition (1996)

[9] G. Darke "Assessment of timbre using verbal attributes", *Proceedings of the Conference on Interdisciplinary Musicology*, Montreal, Canada (2005)

[10] T. Grill, A. Flexer, S. Cunningham "Identification of perceptual qualities in textural sounds using the repertory grid method", *Proceedings of the 6th Audio Mostly Conference*, Coimbra, Portugal (2011)

[11] Pompeu Fabra University, Music Technology Group, "Essentia", http://mtg.upf.edu/technologies/essentia (2012)

[12] R.O. Duda, P.E. Hart, D.G. Stork *Pattern classification*, John Wiley & Sons (2001)