Rediscovering the auditory temporal window consequences on the acoustics science

Alejandro Bidondo

Universidad Nacional de Tres de Febrero (UNTREF), Buenos Aires, Argentina.

Dipl. Sound Engineering career Coordinator.

abidondo@untref.edu.ar www.untref.edu.ar

(Received 30 November 2009; accepted 7 December 2009)

On this research the variability of the integration window duration of the auditory system in function of the perceived acoustic signal is verified. The obtained results validate the perception model based on the specialization of the brain's hemispheres and the existence of the autocorrelation and cross-correlation functions achieved at the brainstem. It is essayed a formulae of the duration of the temporal window in function of the effective duration of the autocorrelation's (ACF's) envelope of the sound stimuli.

Key words: Autocorrelation, ACF, Psychoacoustics, Auditory system.

1. INTRODUCTION

1.1. Actual auditory window models analysis:

The research paths on the auditory window integration duration known by now can be divided as follows:

1. Those which stablish a fixed time duration of the temporal window [1] though the subjetive responses are obtained in function of some hall's objetive acoustic parameters, using fixed time limits without enough known scientific evidence.
2. Those ones starting from recent studies on the audition phisiology, submit all analysis to two types of brain processes: temporal and spatial. This makes that the subjetive parameters to describe do not depend just on the room acoustics, but on the sound signal procesing capabilityes of the brain too. Undoubtely this conducts definitively to subjetive scales, which clearly connect the phisical phenomena with the hearing human sensations.

It was intended to demonstrate which research path has more solid sustenance and from there in the future study diverse themes in terms of the proved scientific line, LEV (listener envelopment) among them.

1.2 . The audition phenomena: from the cochlea to the brain

At [2] and [3], after several years of research about the brain's response to several perceived sounds, by means of SVR's (*Slow Vertex Responses*) and ABR's (*Auditory Brainstem Responses*), electro-neuro-phisiological analysis methods that permit register the electro-chemical signals passig through the brainstem, raise an audition model based on the sound information processing specialization of each hemisphere.

The power spectral density of the signal coming out from the cochlea can be mapped at certain neuronal position as a temporal activity. This neuronal activities own enough information to process its autocorrelation. In fact, the firing ratios of the auditory nerve of a cat reveals an autocorrelation pattern of the incoming information instead of a frequency analysis of it [4].

The firing interval's distributions at a set of neurons constitute a coding method, a transmision method and a general representation of sound information distributed on time. Considering this is that the union (by frequecy bands) of this firing distributions resembles a running autocorrelation function envelope [5].

If the temporal forward masking is kept constant, it would be related with a mechanical resolution limitation of the basilar membrane, maintaining the spatial theory and spectral patterns being received by the brain. Otherwise, if the duration of the integration window varies in function of certain characteristics of the sound stimuly, it would be directly related with the pitch temporal resolution model and other sound stimuly characteristics.

1.3 Time coding of auditory forms:

Temporal codes are those composed by neural pulses in which relative firing times transport information. This means which neurons fire more in function of the presented tonotopy at the basilar membrane. The temporal codification of the sensory information is possible when there exists some correlation between the waveform of the stimuly and the firing discharge probability of the neurons. This correlation can be produced by receptors that follow some aspect of the stimuly waveform somehow that it imprints its time structure on their firings. Virtually at all sensorial modality of the human being exists some aspect wich perception may correspond to a time coding information [7], [8] and [17]. Particularly at the auditory system is specially evident the existence of a time structure produced by any sound sitmuly. There are psicophisical and neurophisical evidences that suggest that this temporal information contains important musical qualities, as pitch, timbre and rythm [8].

1.4 Running Autocorrelation Function (*r*ACF):

$$\phi_p(\tau) = \phi_p(\tau;t,T) = \frac{\frac{1}{2T}\int_{t-T}^{t+T} p'(s)\cdot p'(s+\tau)ds}{\sqrt{\left[\Phi_p(0;t,T)\cdot\Phi_p(0;t+\tau,T\right]}} \qquad (1)$$

This type of autocorrelation reflects the self similarity of the signal in function of time is calculated for a limited integration interval named 2T, which is time shifted for the whole sound file under analysis. This calculation permits, among other things, obtain the ACF and the τ_e values in a histogram display type, as can be seen in figure 1, for Mozart's "Divertimento" music motif.

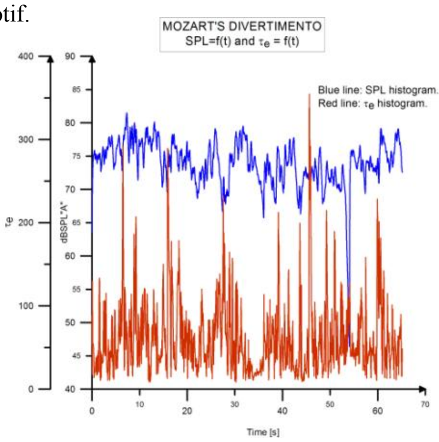

Figure 1. *SPL=f(t)* and τ_e*=f(t)*, for Mozart's "Divertimento".

The ACF has several microscopic parameter within it [11]:

$\Phi_p(0)$: Is the energy contained in the signal itself. It is possible to know the signal's sound level by measuring $\Phi_p(0)$, applying the appropriate time window, 2T, corresponding to the duration of the auditory time window.

τ_e: The effective duration of the normalized ACF envelope. This value represents a repetitive property or "reverberation" contained within the signal. τ_e Is the fundamental time unit of every sound field in a concert hall. It's minimum value reflects the most active part of the signal, existing several local τ_{emin} distributed into the whole analyzed sound file.

τ_1 y $\phi 1$: The first maximum, time and amplitude location, respectively, of the ACF.

The fine structure of the ACF (time instants and their respective positive amplitudes), after τ_1.

It is possible to mention that τ_e vary in function of the selected integration interval, 2T, to make the ACF analysis over the sound signal.

The appropriate time integration window goes from 30ms to 1000ms depending on the sound source [18]; to be precise:

$$[2T] = 30\cdot\tau_{e_{minimo}} \qquad (2)$$

τ_{emin} Represents the information density (in function of the time unit) of the sound sources. A sound source with a small τ_{emin} value contains a denser information at that instant [14] of its histogram. This means more information quantity[a] than the mean value in the time unit, more information in less time, or both.

From a psychological point of view, the integration time interval would be the "psychological present" [19], which is 2T ≅ 0.1s ~ 5s.

The higher the integration time for ACF analysis works as a low pass filter of the sound information with lower cut off frequency. This would smooth the parameter's histogram, resulting in different τ_{emin}, values wich sometimes are substantially different from those of the real 2T's.

The study of the τ_e histogram, permits identify the various instants of local τ_{emin}, moments in which the audition attention is maximal, instants that would strongly determine the auditory window duration [9] and [10].

At previous studies, [12], [13] and [14], it has been computed just only one τ_{emin} because of a short duration of the analyzed sound files, between 2 and 20 sec and/or the sound sources had been mainly simple ones (not complex neither of high dynamic range and several musical expressions), for example: vocals, telephone rings, working HVAC, a very short piece of a string quintet, etc. This prevented to observe several τ_{emin} within a complex sound source (in the present case, the duration of each sound file was of 1 minute), which would result in extremely low values, several instants of very high sensitivity to 2T duration, that would suggest a constant (running) adaptivity of the duration of the auditory time window.

How many τ_{emin} exist in a whole music motif? Which is the level the brain considers a τ_e as a local minima, so to dominate the duration of the auditory time window during a certain time period? Would it depend on the hearing desire? Does the dynamic audio signal processing modify the local τ_{emin}?

From [6] is concluded that:

a) The parameters defined in terms of brain specialization are orthogonal between them and strongly depend on τ_e. b) If audition adapts its 2T value in a volitive way (to the τ_e wanted to be heard), the cocktail party effect could be re-defined.

Established the previous and trying to sustain the development of the time adaptive firing neural nets theory based on temporal coded signals [7], it only remains to verify in the most direct way the presented model.

2. A SIMPLE EXPERIMENT

One of the "*a priori*" conclusions of the time coding theory is the adaptivity of the auditory time window to the minimal effective duration of the normalized ACF envelope. By the negation, if somehow could be tested the non variability of the subjetive 2T window duration in function of τ_{emin} from diverse audio programs, the hypothesis would be negated automatically. So it was come to stand out the auditory temporal window with 4 music motifs.

2.1 Experiment:

They were reviewed the responses of 8 subjects (effectively), with normal hearing, with smedium trained audition, capables to understand auditory instructions.

They were put under 4 mono music motifs (1 minute each), all with different $\tau_{e\,min}$, with only one reflection. The reflection was located to the center initially; then it was located with a panning of 45% on a stereophonic system (40.5º from the center).

The music motifs used were: an anechoic Jazz theme unsinged, a string part of Mozart's "Divertimento", an anechoic pop female vocal and a piece of an organ concert of Orlando Gibbons named "Fantasía". This conditions impose diverse τ_e histograms. The listening was done by means of Audio Technica ATH D40 headphones. The sound level was constant between essays. The LeqA used was 75dBA. This sound leves was registered inside the headphones cavity, with a sound level meter Svantek SV959. The audio interface used was M-Audio 410. A sound field composed by the direct sound and just one reflection was reproduced, this one with delay times of 20ms, 50ms, 100ms, 200ms and 500ms. 4 Tests were done on each subject for each delay, so to obtain the mean of the 4 as results. The variances of each subject and delay were analized. The results with high variance were discarded because they showed low llistening training or not undestanding the instructions or phisical exhaustion. The direct signal was reproduced at a relative level of 0dB on the multitrack software, and kept constant all over the experiment. The delayed signal was reproduced with increasing (or decresing) steps of 3dB or 0.1dB. The subjects were asked to stablish the level, relative to the direct sound, at wich they just notice the reflection. They were asked also to keep constant the detection threshold all over the experiment. The order of the tests was random (music motif and reflection to evaluate). The mean final relative level of detection of the delayed signal was the reslt on each test.

For the calculus of the τ_{emin} on the used music motifs they were not discarded those values result of of breathing intervals or almost-silences, because at that moments the subjects could perceive the sound "*cues*" to judge the valuations asked on every test. One was due to instruct the subjetcs on the diverse limits of perception of a single reflection by means of Olive and Toole findings on [16].

3 RESULTS

At table 1 are shown the $W_{Audition}$ results corresponding to lateralized reflections. At table 2 is shown one subject's detection of the same delay at different instants. This means that the brain does not perceive the only one minimum τ_e of a sound file. These results confirm the existence of several local τ_{emin} within a music motif, which bring the sound "*cues*"; this motivates the use of a statistic descriptor: the percentile. By

[a] The term "information quantity" makes reference to its valuation, where a higher value is assigned to a message that has less probabilities to be received or happen [15].

iteration was found the most relevant percentile for the founded results (figure 4).

On figure 2 are displayed the results for the reflection placed 40.5°, because real world is binaural, not monoaural.

The duration of the auditory time window, $W_{Audition}$ [ms], is defined as the time interval between the perception of the direct signal and the one at which the reflection perception is -10dB from the one found at the 20ms delayed signal. This can be found at figure 3.

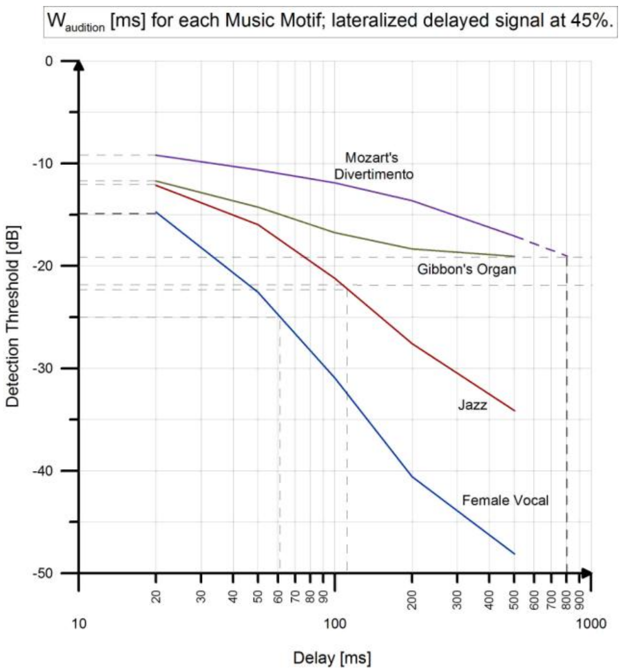

Figure 2. Detection limits results for lateralized delay reflection.

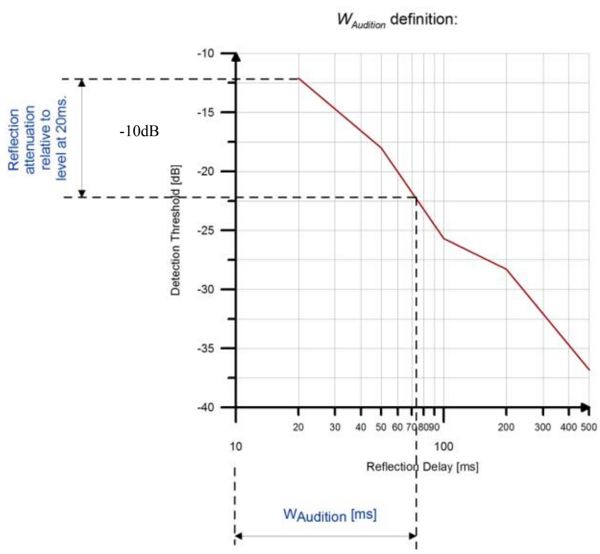

Figure 3. $W_{Audition}$ Definition.

Table 1. $W_{Audition}$ results. Lateralized reflection.

Music Motif	Panning = 45% WAudition [ms] = 2T	τemin [ms]	τe Percentile 99% [ms]	τe Percentile 95% [ms]
Anechoic Female Vocal	61	3.680	4.682	5.820
Jazz	107	5.050	6.086	8.272
Mozart's "Divertimento"	800	7.570	9.247	13.045
Gibbons Organ	1400	8.520	11.059	14.438
75dBA				

Table 2. Detection instants of the cues for each delay.

Subject Result's				
Music Motif	MOZART'S DIVERTIMENTO			
	Detection Threshold		Detection instant [s]	
Delay [ms]	Test 1	Test 2	Test 1	Test 2
20	-12.6	-11.7	12	35
50	-13.4	-12.8	62	20
100	-14.6	-13.9	10	48
200	-16.2	-15.7	8	55
500	-18.8	-18.9	44	32

On figure 4 can be found the $W_{Audition}$ in function of τ_ePercentile95%, a linear and an exponential fitting.

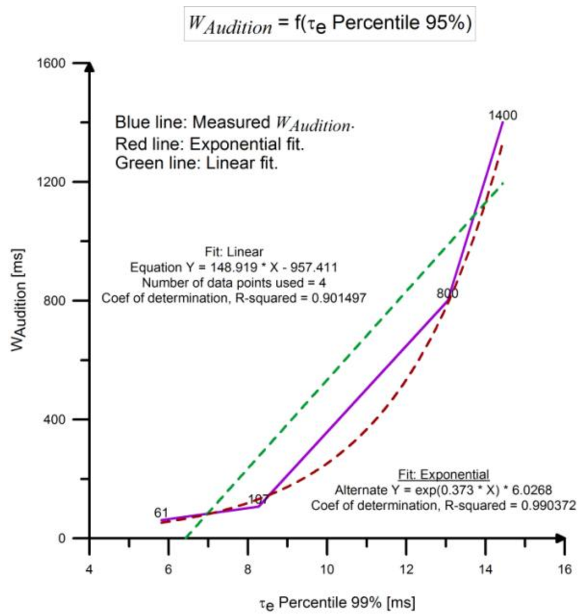

Figure 4. $W_{Audition}$ in function of τ_ePercentile95%.

The best fitting curve in figure 4 corresponds to equation 3:

$$W_{Audition}[ms] = 6 \cdot e^{(0.373 \cdot \tau_e Percentile95\%)} \quad (3)$$

4 CONCLUSIONS

The duration of the auditory time window varies in function of certain parameters of the sound stimuli.

A predictive equation is presented.

Brain detects sound data at several τ_{emin}, during a music motif or sound file.

From the previous evidences, more acoustic descriptors should be developed based on brain processes, ACF & CCF, so to be directly related to subjective responses to sound stimulus.

REFERENCES

[1] Rossing, Thomas. (2007). "Handbook of acoustics", Ch. 10, page 360. Springer.

[2] Tohyama, M, Suzuki, H & Ando, Y. (1995). "The Nature and Tecnology of Acoustic Space". Academic Press. Capítulo 5.

[3] Ando, Yoichi. (1994). "Proceedings of the international conference on Acoustic Quality of Concert Halls. State of the art and research trends". Fundación Ramón Areces. Capítulo. 3, pags 63 – 71. Madrid, october 24 – 25.

[4] Secker-Walker, H. E.; Searle, C. L. (1990). Journal of the Acoustical Society of America 88, 1427–1436. Time domain analysis of auditory-nerve-fiber firing rates.

[5] Cariani, P. A.; Delgutte, B. (1996). Journal of Neurophysiology 76, 1698–1716. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience.

[6] Ando, Yoichi. (2002). "Correlation factors describing primary spatial sensations of sound fields". Journal of Sound and Vibration. 258(3), 405–417, doi:10.1006/jsvi.5264.

[7] Cariani, Peter; Tramo, Mark. (1997). "Neural representation of pitch through temporal autocorrelation" Proceedings, Audio Engineering Society Meeting (AES), New York, September, 1997, Preprint #4583 (L-3).

[8] Cariani, Peter. (2001). "Temporal codes, timing nets, and music perception". Journal of New Music Perception, vol. 30, issue 2. Special issue on Rhythm Perception, Periodicity, and Timing Nets.

[9] Ando, Yoichi. (1998). "Architectural Acoustics. Blending sound sources, sound fields, and listeners". Ch. 3, page 8 - 23. Modern Acoustics and Signal Processing, AIP Press.

[10] Ando, Yoichi. (2008). Theory of Auditory Temporal and Spatial Primary Sensations", Journal of Temporal Design in Architecture and the Environment, Japan.

[11] Ando, Yoichi. (1998). "Architectural Acoustics. Blending sound sources, sound fields, and listeners". Modern Acoustics and Signal Processing, AIP Press.

[12] Kato, Kosuke; Noson, Dennis; Ando, Yoichi. (2002). "Blending Vocal Music With Sound Fields by Variations in Lyrics and Music Tempo". S.E.A.

[13] Mouri, K., Akiyama, K., Ando, Y. (2001). "Preliminar study on recommended time duration of source signals to be analyzed, in relation to its effective duration of the autocorrelation function". Journal of Sound and Vibration, 241, pages 87 – 95.

[14] Kato, Kosuke; Fujii, kenji; Kawai, Keiji; Ando, Yoichi; Yano, Takashi. (2004). "Blending vocal music with the sound field –the effective duration of autocorrelation function of Western professional singing voices with different vowels and pitches". Proceedings of the International Symposium on Musical Acoustics, March 31st to April 3rd, (ISMA2004), Nara, Japan.

[15] http://es.encarta.msn.com/encyclopedia_761577650/Teor%C3%ADa_de_la_informaci%C3%B3n.html

[16] Olive, S. E.; Toole , F. E. (1988). "The Detection of Reflections in Typical Rooms," presented at the 85th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 36, p. 1029, preprint 2719.

[17] Cariani, Peter. (1995). "As if time really mattered: Temporal strategies for neural coding of sensory information". Communication and Cognition-Artificial Intelligence (CC-AI), March, 1995, Vol 12, nos. 1-2, pp. 157-219. Special Issue on Self-Reference in Biological and Cognitive Systems, L. Rocha (ed).

[18] Ando, Yoichi. (2009). Recomendaciones propias del Dr. Ando asentadas en varias comunicaciones personales con él intercambiadas durante el año 2009.

[19] Fraisse, P. (1967). Psychologie du temps. París. PUF.