Rediscovering the auditory temporal window and 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 establish a fixed time duration of the temporal window [1] though the subjective responses are obtained in function of some hall’s objective acoustic parameters, using fixed time limits without enough known scientific evidence.

2. Those ones starting from recent studies on the audition physiology, submit all analysis to two types of brain processes: temporal and spatial. This makes that the subjective parameters to describe do not depend just on the room acoustics, but on the sound signal processing capabilities of the brain too. Undoubtedly this conducts definitively to subjective scales, which clearly connect the physical 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 passing 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 transmission method and a general representation of sound information distributed on time. Considering this is that the union (by frequency 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 stimulus, it would be directly related with the pitch temporal resolution model and other sound stimulus 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 stimulus and the firing discharge probability of the neurons. This correlation can be produced by receptors that follow some aspect of the stimulus waveform somehow that it imprints its time structure on their firings. Virtually at all sensorial modality of the human being exists some aspect with echolocation 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 stimulus. There are psychophysical and neurophysiological evidences that suggest that this temporal information contains important musical qualities, as pitch, timbre and rhythm [8].

1.4 Running Autocorrelation Function (rACF):

$\phi_r(r) = \phi_x(r; t, T) = \frac{1}{2T} \int_{-T}^{T} r'(s) \cdot r'(s + r) ds$ (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 $\tau_e$ values in a histogram display type, as can be seen in figure 1, for Mozart’s “Divertimento” music motif.

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

$\tau_e$: The effective duration of the normalized ACF envelope. This value represents a repetitive property or “reverberation” contained within the signal. $\tau_e$ 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 $\tau_{emin}$ distributed into the whole analyzed sound file.

$\tau_1$ y $\phi_1$: The first maximum, time and amplitude location, respectively, of the ACF.

It is possible to mention that $\tau_e$ 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:

$$2T = 30 \cdot \tau_{emin}.$$ (2)

$\tau_{emin}$ Represents the information density (in function of the time unit) of the sound sources. A sound source with a small $\tau_{emin}$ value contains a denser information at that instant [14] of its histogram. This means more information quantity 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 $\approx 0.1s \sim 5s$.

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 $\tau_{emin}$ values which sometimes are substantially different from those of the real 2T’s.

The study of the $\tau_e$ histogram, permits identify the various instants of local $\tau_{emin}$ moments in which the auditory 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 $\tau_{emin}$ 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 \( \tau_{\text{emin}} \) 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 \( \tau_{\text{emin}} \) exist in a whole music motif? Which is the level the brain considers a \( \tau \) 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 \( \tau_{\text{emin}} \)? From [6] is concluded that:

a) The parameters defined in terms of brain specialization are orthogonal between them and strongly depend on \( \tau \). b) If audition adapts its 2T value in a volitive way (to the \( \tau \) 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 subjective 2T window duration in function of \( \tau_{\text{emin}} \) 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 \( \tau_{\text{emin}} \), 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 “Fantasia”. This conditions impose diverse \( \tau \) 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 levels 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 listening training or not understanding 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 decreasing) 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 \( \tau_{\text{emin}} \) 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 \( W_{\text{Audition}} \) 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 \( \tau \) of a sound file. These results confirm the existence of several local \( \tau_{\text{emin}} \) within a music motif, which bring the sound “cues”; this motivates the use of a statistic descriptor: the percentile. By
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, $W_{\text{Audition}} \ [\text{ms}]$, 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.](image)

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

<table>
<thead>
<tr>
<th>Music Motif</th>
<th>$W_{\text{Audition}}$ [ms]</th>
<th>$\tau_{\text{emin}}$ [ms]</th>
<th>95th Percentile [ms]</th>
<th>99th Percentile [ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anechoic Female Vocal</td>
<td>61</td>
<td>3.680</td>
<td>4.882</td>
<td>5.820</td>
</tr>
<tr>
<td>Jazz</td>
<td>50</td>
<td>2.350</td>
<td>3.686</td>
<td>4.972</td>
</tr>
<tr>
<td>Mozart’s “Divertimento”</td>
<td>300</td>
<td>2.970</td>
<td>3.247</td>
<td>4.045</td>
</tr>
<tr>
<td>Gibbons Organ</td>
<td>1400</td>
<td>8.520</td>
<td>11.059</td>
<td>14.438</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Subject Result’s</th>
<th>Detection Threshold</th>
<th>Delay [ms]</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 1</th>
<th>Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOZART’S DIVERTIMENTO</td>
<td>61</td>
<td>20</td>
<td>-12.6</td>
<td>-11.7</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>-13.4</td>
<td>-12.6</td>
<td>52</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>-14.6</td>
<td>-13.9</td>
<td>10</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>-16.2</td>
<td>-15.7</td>
<td>8</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>300</td>
<td>-18.8</td>
<td>-18.9</td>
<td>44</td>
<td>32</td>
</tr>
</tbody>
</table>

On figure 4 can be found the $W_{\text{Audition}}$ in function of $\tau_{\text{Percentile95%}}$, a linear and an exponential fitting.

$$W_{\text{Audition}} = f(\tau_{\text{Percentile95%}})$$

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

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

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 $\tau_{\text{emin}}$ 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


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


