Asociación Argentina
de Mecánica Computacional

Mecánica Computacional Vol XXVIII, págs. 9-22 (artículo completo)
Cristian García Bauza, Pablo Lotito, Lisandro Parente, Marcelo Vénere (Eds.)
Tandil, Argentina, 3-6 Noviembre 2009

FLUCTUATION STRENGTH OF MIXED FLUCTUATING SOUND SOURCES

Ernesto Accolti[a] **and Federico Miyara**[b]

Laboratorio de Acústica y Electroacústica, Facultad de Electrónica, Universidad Nacional de Rosario. Riobamba 245 bis, Rosario, Argentina, http://www.fceia.unr.edu.ar/acustica

[a]eaccolti@fceia.unr.edu.ar

[b]fmiyara@fceia.unr.edu.ar

Keywords: Fluctuation Strength, Loudness Fluctuation, Sound Quality, Psychoacoustics.

Abstract. A preliminary study was carried out with broadband noises and a few subjects. The result was that fluctuation strength models for sounds with the same modulation frequency, as those available in the literature, are not extensible for combined signals modulated with different modulation frequencies.

An exploratory study of fluctuation strength of combined fluctuating tone signals with different modulation frequencies was conducted by magnitude estimation experiments. For that purpose a graphical and interactive software interface was developed. The interface manages the test sounds and 10 anchor sounds that the user can listen until he or she is convinced of the magnitude estimation. A computational procedure was implemented to choose the modulation index of the anchor sounds that exhibits equally spaced fluctuation strength. Although no correlation was found, some possible-related objective-parameters are described. Computational models for accounting those possible-related objective-parameters are implemented.

1 INTRODUCTION

Fluctuation Strength (FS) is the psychoacoustic parameter that describes how strongly or weakly sounds fluctuate. This parameter is sometimes called subjective or, more accurately, perceptual. There are models for frequency- or amplitude-modulated (FM or AM) sinusoidal tones, as for amplitude-modulated broadband noises. These models depend on the loudness of sound, particularly on the frequency and depth of the loudness fluctuation. Literature models assume that the modulating frequency is the same for all sound sources present in the signal.

When multiple sources are present in a single sound field, differences of modulation patterns among sources may aid segregation of the auditory images that represent the individual sources. These considerations apply to both AM and FM (Sheft 2008). Experimental data from the literature support that human neural mechanisms extract AM information. When this segregation is possible FS could be, to some extent, separated too. But there are other cases when it is not possible, because of the individual perception differences, because the carrier frequency bands are very similar or just because the same source produces sound as combination of modulations (Moore et al., 2002).

In this paper the research focus on non-separated FS of mixed AM sounds with different modulation frequencies as a complement of a routine for the controlled combination of sounds to be used in experiments in noise perception (see Accolti et al., 2008).

2 MODELS

For single noises and tones, FS sensation has been proven to be proportional to the temporal masking depth (ΔL), e.g. the difference between the maximum and minimum level of the masking pattern. The masking pattern is defined as the necessary sound pressure level of a frequency-independent test tone to be just audible in the presence of a masking tone. For broadband noises ΔL is largely independent of frequency because the nonlinearity of the upper slope in masking patterns overlaps with the adjacent frequency bands but for AM tones some frequency dependence occurs. Furthermore maximum and minimum masking patterns are those after postmasking occurs. Postmasking can be defined as the temporal-dependent sound pressure level of a very short burst tone that occurs after a masking tone or noise ends necessary to be just audible. So, before computing the minimum and maximum masking levels, a two dimensional (time and frequency) response must be estimated. This response is computed in terms of the loudness (Chalupper et al., 2002) that could be seen as a nonlinear serial block with masking pattern. Experimental data show that FS shows a band-pass characteristic as a function of modulation frequency, with a maximum around 4 Hz (Zwicker et al., 2007). Zwicker proposed equation (1) for pure tones

$$FS = \frac{0{,}008 \displaystyle\int_{0Bark}^{24Bark} \frac{\Delta L}{dBBark}\, dz}{\left(\frac{f_{\mathrm{mod}}}{4Hz}\right) + \left(\frac{4Hz}{f_{\mathrm{mod}}}\right)} vacil \tag{1}$$

where f_{mod} is the modulation frequency of an AM or FM tone.

A block diagram of the algorithm for FS calculation is outlined in Figure 1 (Chalupper 2000). The input data is the time signal relative to the sound pressure filtered to account for free field to the inner ear transmission. For broadband noises, in line to simplify the model, this input data can be supplied in 24 frequency bands corresponding to the critical bands related to the basilar membrane (Zwicker et al., 2007).

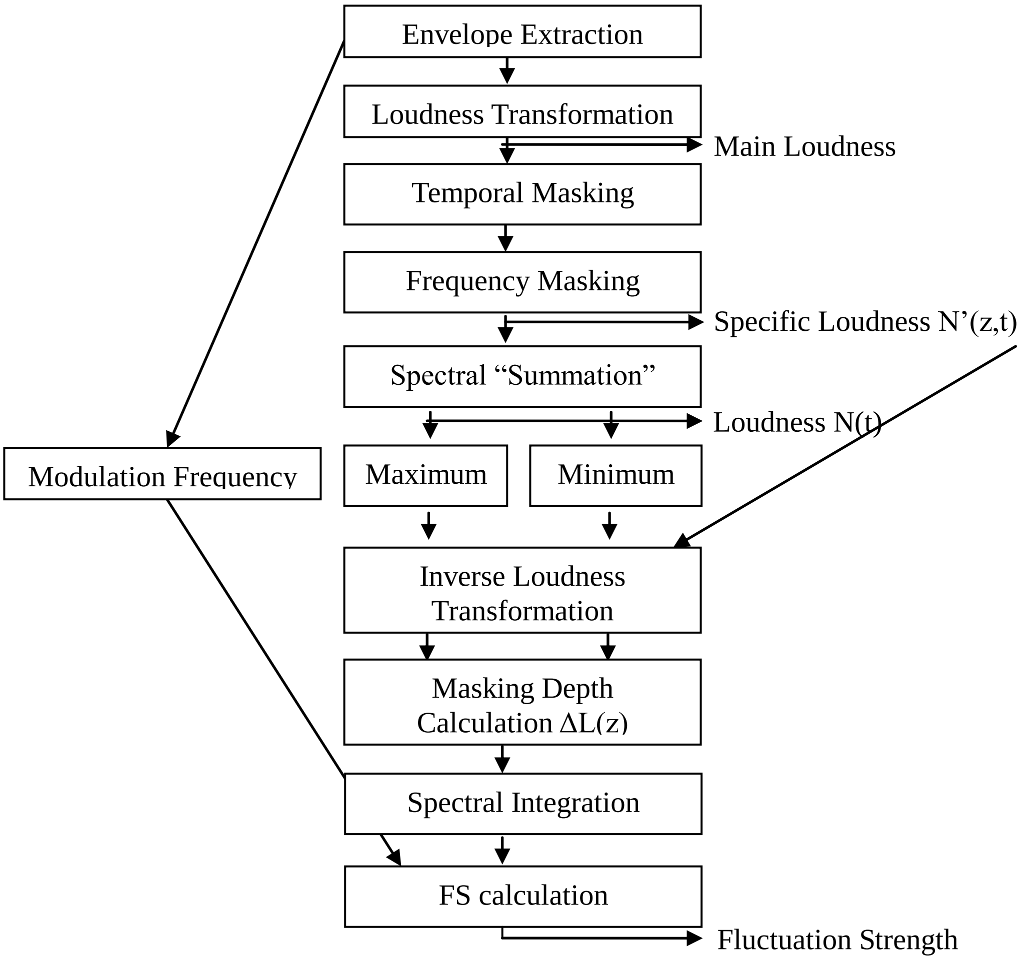

Figure 1: Fluctuation Strength calculation blocks diagram

2.1 Fluctuation Strength algorithm for simple sources

The envelope extraction is done by a first order low pass filter with time constant $\tau = 4$ ms applied to the squared signal followed by an 88 factor downsampling step. The antialiasing filtering involved in the downsampling process removes any remaining high frequency ripple, while retaining the relevant features of the envelope. Note that for broadband noises the envelope extraction is a 24-channel parallel operator. The modulation frequency (f_{mod}) is calculated as the reciprocal of the time period of two adjacent maximum.

The nonlinear loudness transformation is calculated by equation (2) in accordance with Zwicker et al., 2007.

$$N'(z) = N'_0 \left(\frac{E_{TQ}(z)}{s(z)E_0(z)} \right)^{0,23} \left[\left(1 - s(z) + s(z) \frac{Es(z)}{E_{TQ}s(z)} \right)^{0,23} - 1 \right] \tag{2}$$

where $N'(z)$ is the critical-band rate (z) dependent specific loudness and N'_0 is the reference specific loudness. The hearing threshold (E_{TQ}), the excitation (E) and the reference excitation (E_0) are single values for sinusoids and are frequency (or critical-band rate) dependent for broadband or multiple sinusoidal tone noises. And the ratio (s) between the sound intensity of a just-audible test tone and the sound intensity of the internal noise appearing within the critical band at the test tone's frequency is a frequency dependent value either for single tones or broadband noises.

The main loudness function is available after the loudness transformation block. The temporal masking block accounts for postmasking only because premasking time is similar to the response time of the filter used in the envelope extraction block. The postmasking is modeled as a nonlinear filter as proposed by Chalupper et al., 2002 in accordance with Zwicker 1984 results.

The masking pattern is achieved with the frequency masking block model by Zwicker et al., 2007, using the reported tuning curves (see Figure 2)

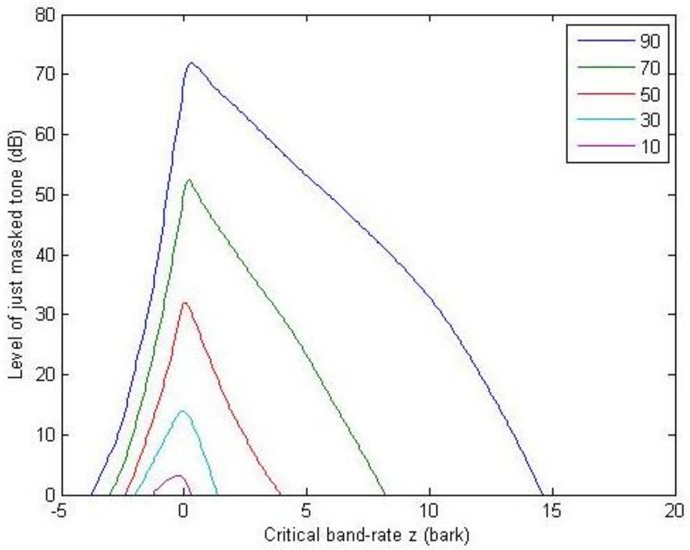

Figure 2: Masking patterns for test pure-tones masked by pure tones. The parameter is the masker level.

The output of this block is the specific loudness time pattern, a time and frequency dependent, two dimensional, matrix for broadband noise as for single or multiple pure tones. Figure 3 shows an example of a two-tone complex.

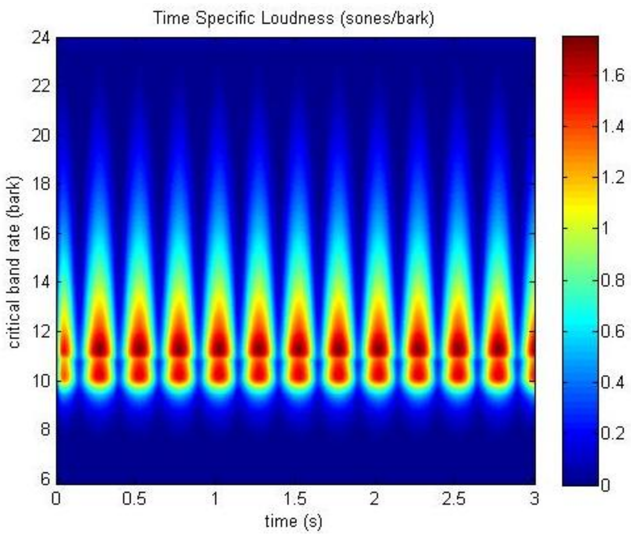

Figure 3: Specific loudness for two pure tones of 63 and 65 dB sound pressure level, 10 and 11 bark respectively and both sine-modulated at 4 Hz with 40 dB modulation depth.

A spectral perceptive summation is done by keeping the maximum values for each z from each z channel, achieving time-dependent loudness shown in Figure 4.

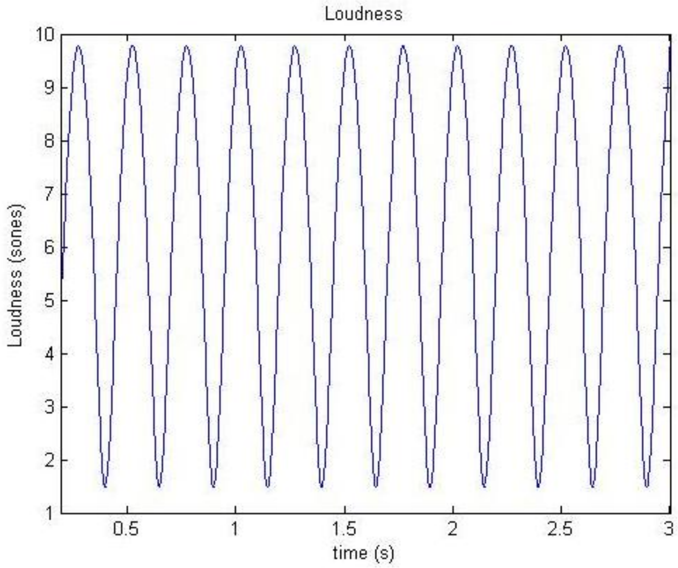

Figure 4: Loudness for two pure tones of Figure 3.

The instants for which the maximum and minimum loudness occur are used as candidates for which the maximum and minimum masking levels also occur. First the frequency-dependent specific loudness at such instants is sliced out from the matrix depicted in Figure 3. Then the inverse loudness transformation block calculates the inverse of equation (2) to get the maximum and minimum masking levels depicted in Figure 5.

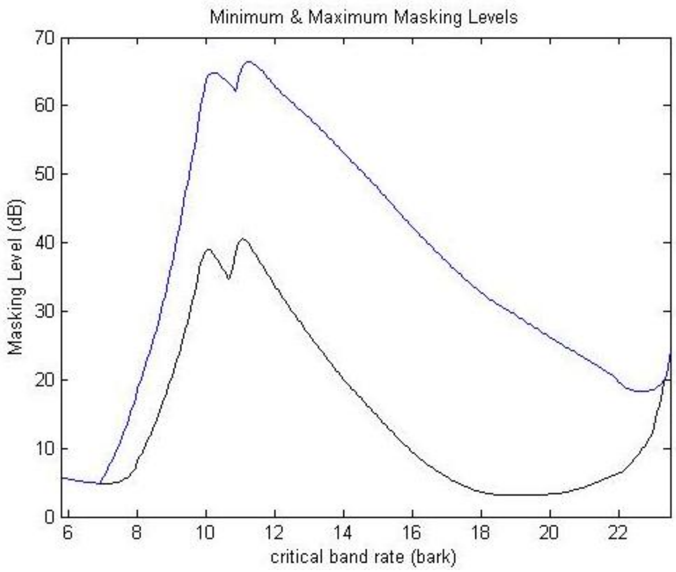

Figure 5: Maximum (blue) and minimum (black) masking levels for pure tones of Figure 3. The masking depth $\Delta L(z)$ is the distance between these two curves.

Finally the last block implements equation (1), where the inputs are the spectral integration of masking depth, achieved in the spectral integration block, and the modulation frequency.

2.2 Fluctuation Strength of mixed sources

Although a complete model is not yet available, results demonstrate some facts that can be useful for further studies as for a better understanding the phenomena.

A mixed signal of two (or more) AM signals with different modulation frequencies ($f_{mod,i}$; i = 1,2) exhibits a compound modulation frequency (f_{modc}). Throughout time f_{modc} varies around the greatest of $f_{mod,i}$. The distribution of f_{modc} (in a logarithmic transformation) is illustrated in Figure 6 for two AM sine signals with modulation frequencies $f_{mod,1}$ = 4 Hz and as function of $f_{mod,2}$, respectively. Note that when $f_{mod,2} \approx 4$ Hz = $f_{mod,1}$, f_{modc} is approached by a constant value (the same for $f_{mod,2} \approx 2$ Hz). For $f_{mod,2} < f_{mod,1}$. f_{modc} varies around $f_{mod,1}$, and for higher values of $f_{mod,2}$, f_{modc} varies around $f_{mod,2}$. This fact adds a confusion variable for the experimentation since it has been demonstrated that the time variance of psychoacoustical parameters such as loudness [Kuwano et al., 2003], roughness and FS [Hastings et al., 2001] tends to show a different global value than the average value of the instantaneous ones. The variance of f_{modc} is calculated to take into account that inherent variable (see Figure 7).

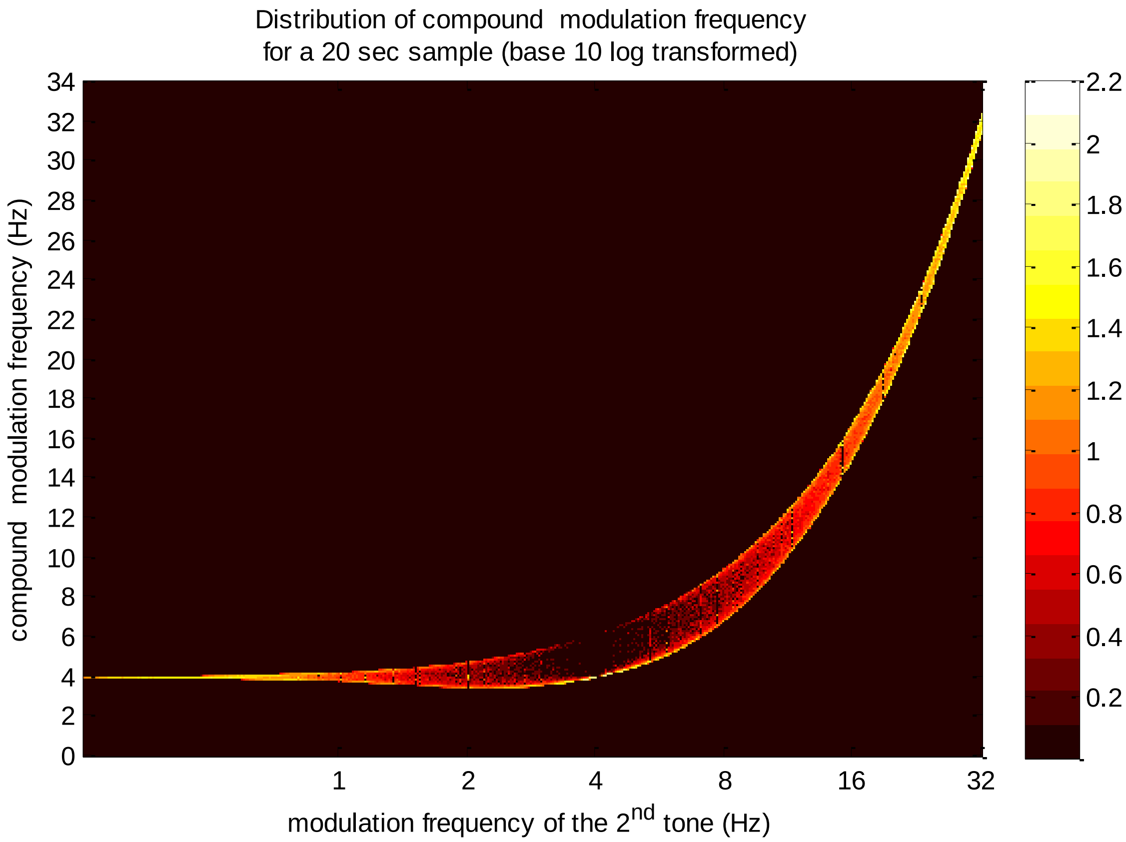

Figure 6: Distribution of compound modulation frequency for a 20 seconds sample of two AM sine tones with $f_{mod,1}$ = 4 Hz as function of $f_{mod,2}$.

Figure 7 shows the standard deviation relative to the compound modulation frequency of 2033 samples of two AM sine tones, the modulation frequency of the first tone was set at 4 Hz and the second one at 2033 values linearly spaced from 0,25 Hz to 32 Hz corresponding to each of the 2033 samples.

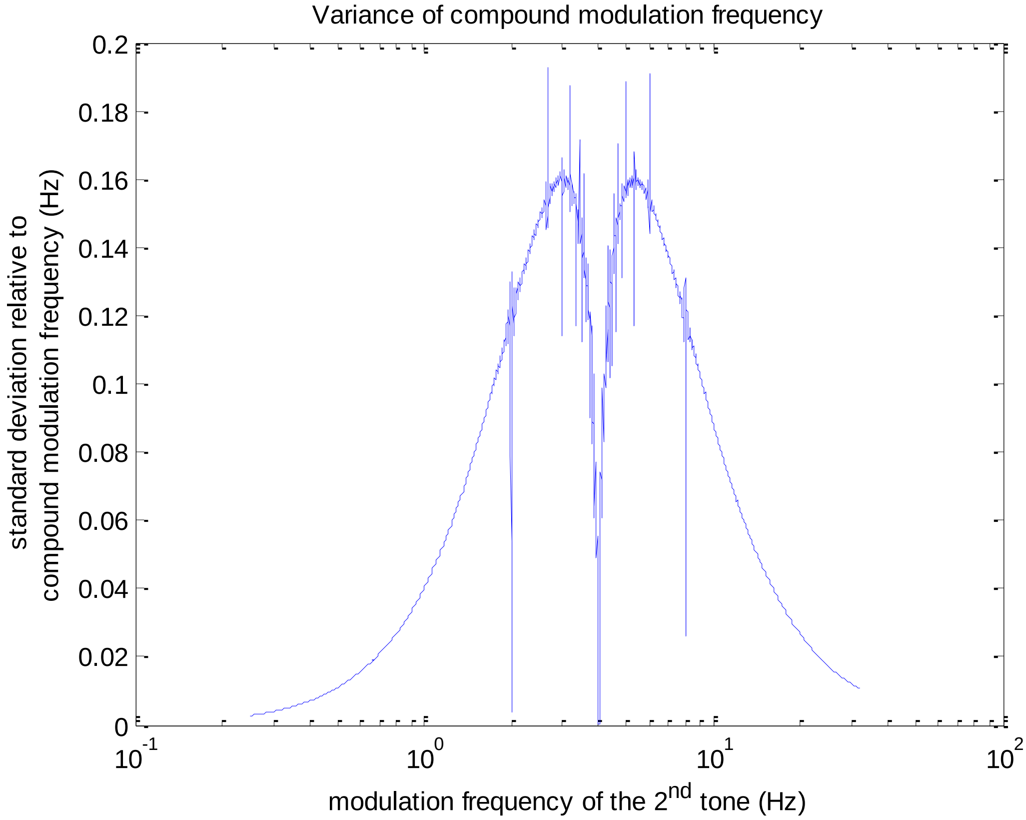

Figure 7: Standard deviation relative to compound modulation frequency for a 20 s sample of two AM sine tones with $f_{mod,1}$ = 4 Hz as function of $f_{mod,2}$.

3 SETUP

The setup used in the experiment is described in this section. Every signal was presented using the virtual auralization setup described above. The broadband experiment was conducted without any graphical interface and all the other experiments were carried out using the experiment platform described above

3.1 Virtual auralization

The auralization was done using an external soundcard (M-Audio Mobile Pre) in series with a self designed impedance adaptor loaded with extended range (up to 16 kHz) audiometric earphones (Sennheiser HDA 200).

A practical procedure was implemented to compensate the effects of the soundcard response as well as the pressure- to free- field and the voltage- to pressure- field effects. This procedure will be described in Accolti et al., 2009.

3.2 Experiment platform for a magnitude estimation test

A computational platform was developed for a magnitude estimation test (see Figure 8). The buttons Ref1 to Ref 10 play sounds with known and equidistant FS as anchor. The button Play Test plays the sound to be tested. The subject has to introduce his estimation in the text box and press Next button to pass to the next test sound, listening at the tested sound at less one time and the every anchor wanted at any times wanted. The test sounds appear in random order for each subject.

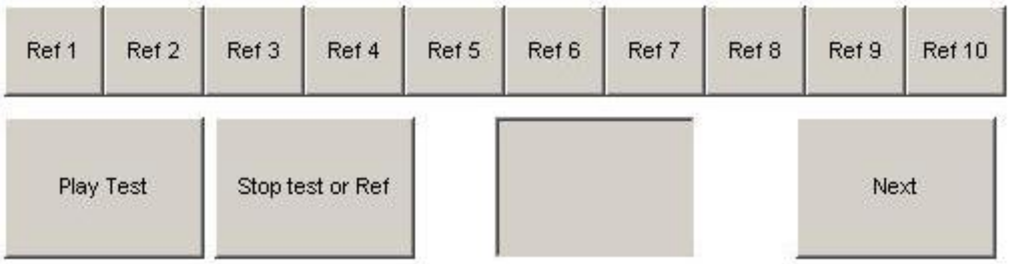

Figure 8: Test Platform.

This platform is also used for training purposes using just the anchors buttons.

3.3 Reference sounds (anchors)

Before describing the anchor, the fluctuation strength model was implemented for a set of 1 kHz sine tones amplitude modulated at 4 Hz with modulation index (m) from 0,1 to 1 in 0,01 steps getting an FS vs. *m* plot. A 10-element FS vector from the minimum to the maximum value of FS was generated and the modulation index *m* was estimated using a spline interpolation of the inverse of FS. The inverse of FS (m) is illustrated in Figure 9 (Note that FS values have been chosen equidistant).

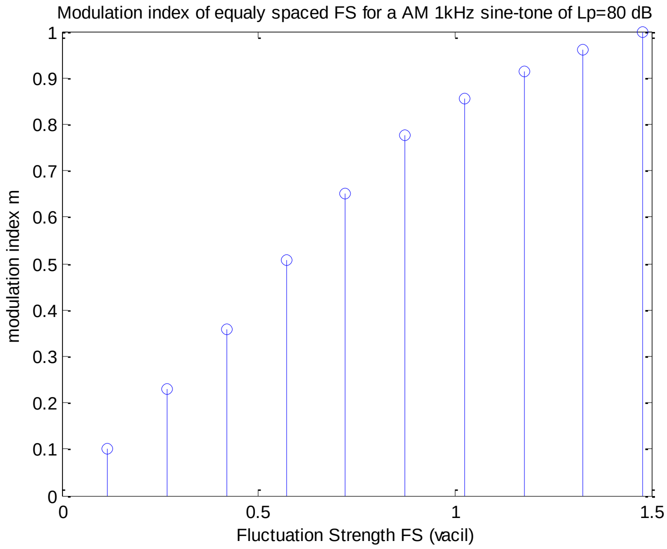

Figure 9: Modulation index for anchors.

Finally the anchors are generated using these values of *m* for AM sine tones keeping the other parameters equal to those used before, i.e., $f = 1$ kHz and $f_{mod} = 4$ Hz.

4 EXPERIMENTATION

4.1 Broadband noise signals at different modulation frequencies

Some preliminary perception tests were conducted to find out the influence on FS when noises with different fluctuation patterns are mixed.

A test for the comparison of stimulus in pairs has been conducted with three subjects using three sounds. The subjects are frequently involved in psychoacoustical tests so they can be considered trained subjects. All three sounds were generated in mathematical software:

a) White noise of sound pressure level L_p = 60 dB modulated by a 5 Hz sine wave with 40 dB modulation depth
b) White noise of L_p = 60 dB modulated by a 6 Hz sine with 40 dB modulation depth
c) The mix of sounds *a* and *b*

The sounds were presented in pairs. Both possible sequential orders were used for each pair. Between the six pairs of sounds a 5 s silence was used to separate them. Between each pair there was a 2 s silence. The subjects were asked to select the most strongly fluctuating sound of each pair.

4.2 Pure tone signals at different modulation frequencies

Training

A preliminary test with 10 subjects was conducted, concluding that the results were strongly variable from one subject to another. The reason of that variability may be that the very meaning of FS was not clear for subjects. Particularly, the well known problem that FS is frequently confused with roughness [Zwicker et al., 2007] was confirmed by spontaneous information provided by some of the subjects. To prevent it, a second experiment was conducted but after a training procedure.

The procedure is similar to the experiment itself but the FS value for each test is shown after the subject reports his judgment for that test. The test sounds are AM broadband noises with known FS.

The training procedure ends when subjects consider that FS concept is clear.

Test sounds

In order to better understand the behavior of FS for signals with mixed modulated signals, a magnitude estimation test was conducted using the experiment platform for the combination of two pure tones modulated at different modulation frequencies as the test signals.

The first tone was set at a carrier frequency of f_1 = 1456 Hz (11 bark), a sound pressure level L_{p1} = 70 dB, a modulation frequency $f_{mod,1}$ = 4 Hz and a modulation index m_1 = 0,98 (40 dB modulation depth).

The other tone was first set at a carrier frequency of f_2 = 1666 Hz (11,9 bark), and a modulation index m_2 = 0,98 (40 dB modulation depth). The sound pressure level (L_{p2} = 70,6 dB) was chosen to be on the same equal loudness level contour than the first tone. For that purpose the loudness level that corresponds to L_{p1} and f_1 was first computed as in Espinoza et al., 2006. Then the value of L_{p2} was derived by spline interpolation from the loudness level data generated using the ISO226 method. The modulation frequency $f_{mod,2}$ was varied exponentially to get the 12 test signals shown in Table 1.

0,25	0,39	0,60	0,9	1,5	2,3	3,5	5,5	8,5	13,2	21	32

Table 1: Modulation frequency for the second combining tone (Hz)

The frequency difference between signal carriers (210 Hz) has been selected to be greater than the modulation frequencies reported to cause high roughness sensation [Zwicker et al., 2007] in order to avoid that sensation by avoiding the intrinsic modulation arising from the beating of two simultaneous tones at different frequencies.

A second set of 12 test signals was generated using the same parameters as in the first case but with f_2 = 4554 Hz (18 bark) and L_{p2} = 67,4 dB.

The spectral integration of the difference in the masking pattern ($\int \Delta L(z) dz$) of the test signals has been estimated following the algorithm described in Figure 1, up to the spectral summation block. Figure 10 and Figure 11 show $\int \Delta L(z)\mathrm{d}z$ for first and second set of signals, respectively.

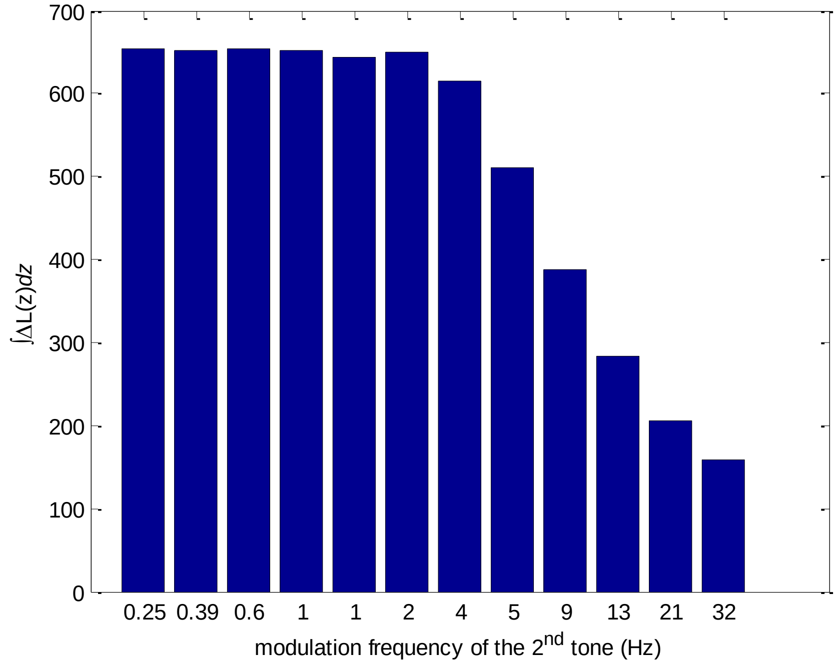

Figure 10: Spectral integration of the difference in the masking pattern of two AM sine tones with $f_{mod,1}$ = 4 Hz as function of $f_{mod,2}$. (carriers in the same critical band).

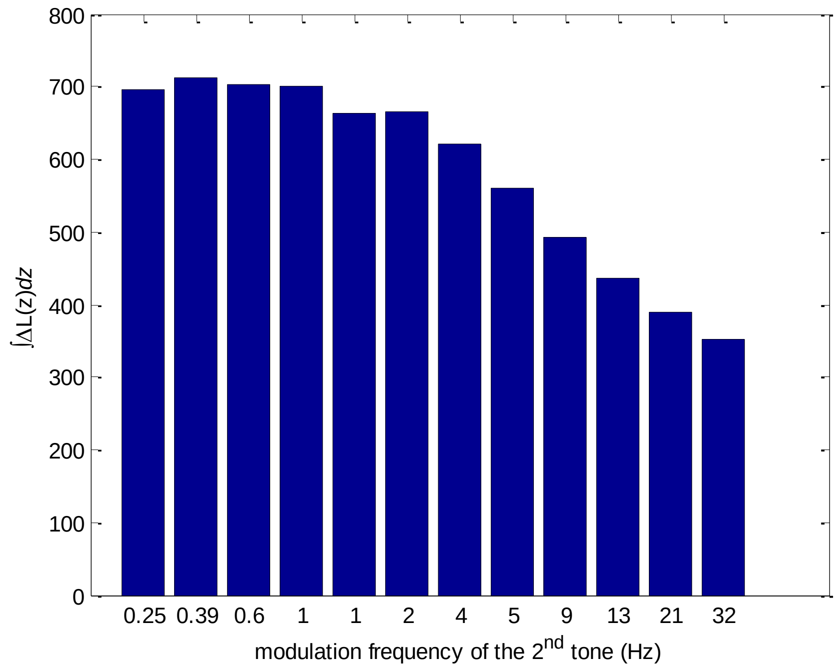

Figure 11: Spectral integration of the difference in the masking pattern of two AM sine tones with $f_{mod,1}$ = 4 Hz as function of $f_{mod,2}$. (carriers in different critical bands).

The standard deviation relative to the compound modulation frequency for each test signal is calculated as in procedure yielding to Figure 7.

5 RESULTS

5.1 Broadband signals at different modulation frequencies

Sound *c* was reported as the most fluctuating one, then *b* and finally *a*, which was never chosen as the most strongly fluctuating one.

The FS of sound *a* and *b* was computed numerically by the models given in 2.1, as expected it is in accordance with the empirical data exposed in the literature. Even though these models have not been developed for different modulation frequencies, an attempt was made to use them to approximate the FS of the sound *c*. From these procedure the estimation of the most fluctuating sound is *b*, the next is *a* and finally *c*, in contradiction with aforementioned experimental results.

5.2 Combination of 2 pure tone signals at different modulation frequencies

The training procedure results are illustrated in Figure 12 in terms of the relative error for each subject (subjects are coded as AS, FE, JN, EX and EO) as function of the sequential order of the test. The absolute maximum error is assigned a score of 100. Although the error tends to decrease as the number of training tests grows, the final error values are still large with a few exceptions. The standard deviation of the error percentage for each subject is reported in Table 2.

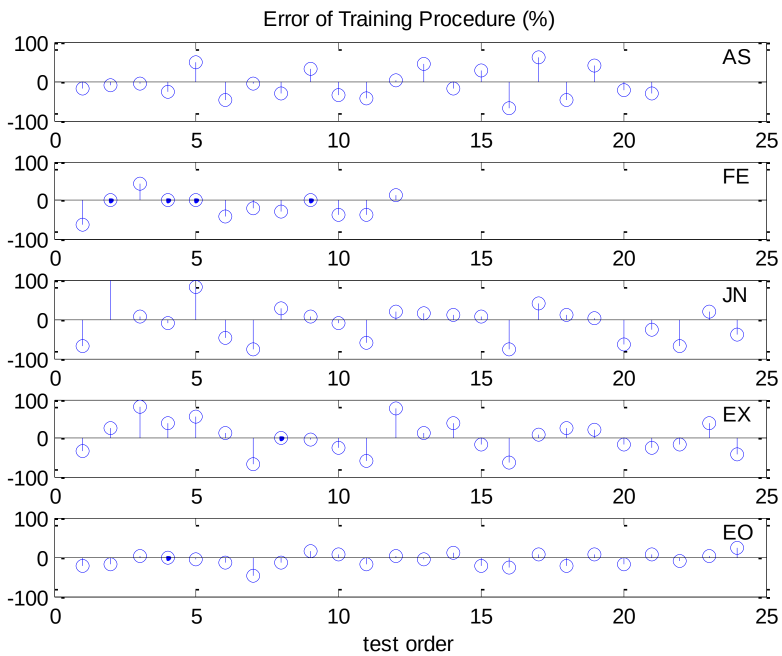

Figure 12: Results of training procedure in sequential order. (Each subject decided to end the training procedure when the FS concept was clear.)

Subject	AS	FE	JN	EX	EO
Standard Deviation of error (%)	40	22	44	39	17

Table 2: Standard Deviation of error in the training procedure for each subject.

Results of the experiment using the first set of 12 signals (both carriers in the same critical band) are shown in a box and whiskers plot (see Figure 13). Boxes represent the interquartile range, red lines the medians (50th percentile), black circles the average, whiskers the extreme values reported within 1,5 times the interquartile range and red plus signs the outliers, i.e. values exceeding the extreme values.

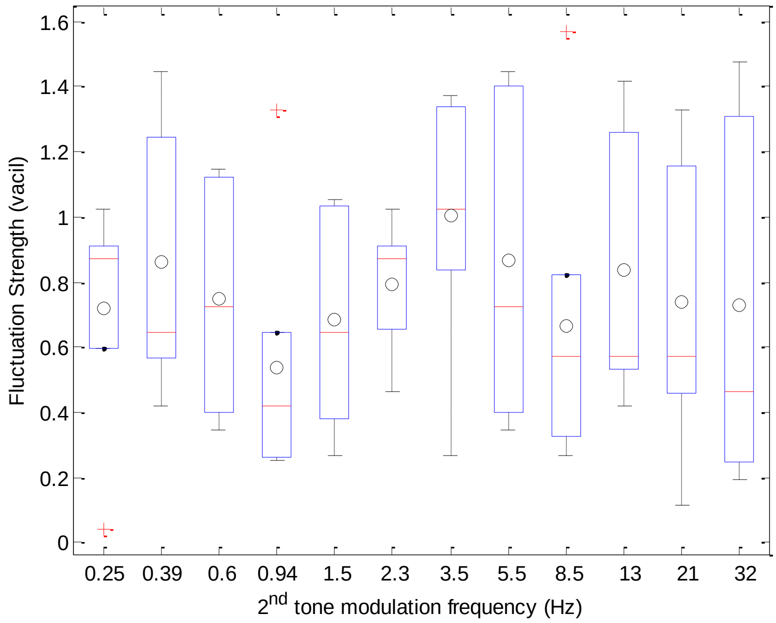

Figure 13: FS of two AM tones with carriers in the same critical band.

Results using the second set of 12 signals (each carrier in a different critical band) are shown in a box and whiskers plot (see Figure 14)

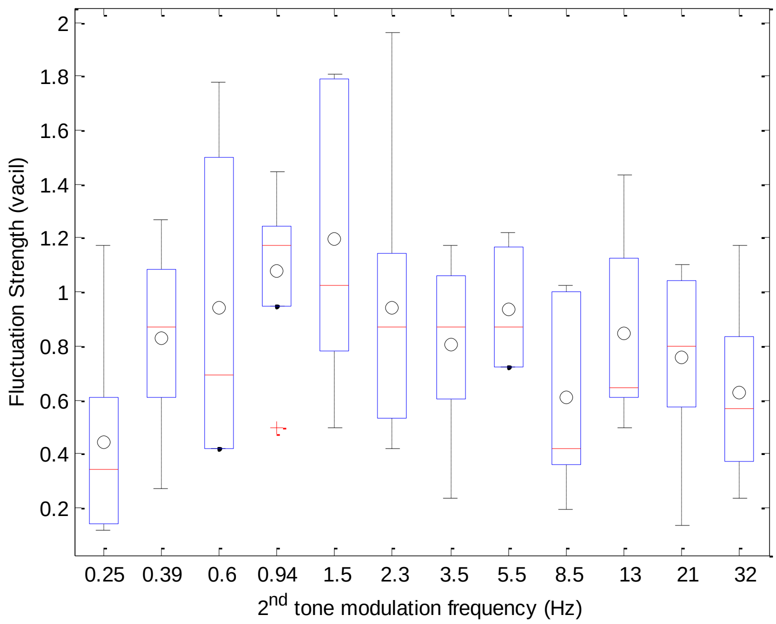

Figure 14: FS of two AM tones with carriers in different critical bands.

For both sets of test signals, the residuals are independent, normally distributed and homocedastic.

In an exploratory study, the correlation of both set of results (p = 0,10) with $\int \Delta L(z)\mathrm{d}z$, f_{modc} and $f_{\mathrm{mod,2}}$ has been studied. The same correlation analysis was carried out for the subjects treating them as blocks.

Although the data is not correlated even with $f_{\mathrm{mod,2}}$ or f_{modc}, an attempt was made to fit them to a polynomial curve. The best fits occurs with constant models for the first set of data and with quadratic models for the second set of data but are not reported because the R^2 statistics was poor (as expected since the correlation is not statistically significant).

6 CONCLUSIONS

Algorithms and procedures for magnitude estimation experiments as well as for training subjects for the proper identification of psychoacoustic parameters have been developed. Furthermore some analysis procedures and algorithms were proposed.

The fluctuation strength of combined sound sources was studied concluding that the estimation models must be extended to take into account this parameter, particularly when the modulation frequency is different for each of the sources.

Some interesting characteristics of AM tone signals with different modulation frequencies were studied, particularly the distribution of the compound modulation frequency and the spectral integration of the global masking pattern difference (as a function of time). Although the experimental data does not show yet significant correlation with those characteristics, it is conjectured that the FS is related to them. The lack of correlation observed is attributed to the spread of the experimental data which, in turn, may be due to an insufficient training.

New research directions point to extend the training procedure for more than a single session in order to avoid fatigue-related effects. Besides, a parallel study of the influence of the instantaneous FS on the global FS should be conducted.

NOTE: This work is part of a research project PICT Nº 38109 financed by the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT)

REFERENCES

Accolti, E., and Miyara, F., Combinación Digital Controlada de Ruidos. *VI Congreso Iberoamericano de Acústica* - FIA 2008, Bs As A090, 2008.

Accolti, E., Miyara, F., and Mignini, E., Protocolo de calibración de auriculares audiométricos para su uso en investigación psicoaústica. *Primeras Jornadas Regionales de Acústica* - AdAA 2009, Rosario, 2009. (In Review)

Chalupper, J., Modellierung der Lautstärkeschwankung für Normal- und Schwerhörige. Tagungsband, DAGA 2000, DEGA, 2000.

Chalupper, J. and Fastl, H., Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica*, 88(3):378-386, 2002.

Espinoza, V., Venegas, R., and Floody, S., Modelo de Sonoridad Utilizando Redes Neuronales. *V Congreso Iberoamericano de Acústica* FIA2006 Santiago de Chile, 2006.

Hastings, A., Davies P. and Takata, H., Effects of Modulation on the Quality of Diesel Engine Noise. *Proceedings of the 17th International Congress of Acoustics* ICA 2001, Rome, 2001.

ISO226. Acoustics - Normal equal-loudness-level curves. 2003.

Kuwano, S., Namba, S., Kato, T., and Hellbrück, J., Memory of the loudness of sounds in relation to overall impression. *Acoust. Sci. & Tech.* 24, 4, 2003.

Moore, B., and Gockel, H., Factors Influencing Sequential Stream Segregation. *Acta Acustica united with Acustica*, 88(3):320-333, 2002.

Sheft, S., Envelope Processing and Sound-Source Perception. *Auditory Perception of Sound Sources*, Springer, 2008.

Zwicker, E., Dependence of post-masking on masker duration and its relation to temporal effects in loudness. J. Acoust. Soc. Am. 75:219–223, 1984.

Zwicker, E., and Fastl, H., *Psychoacoustics: Facts and Models*, Springer, 2007.