Audio Engineering Society

Convention Paper 8655

Presented at the 132nd Convention
2012 April 26–29 Budapest, Hungary

Amplitude Manipulation For Perceived Movement In Depth

Sonia Wilkie,[1] Tony Stockman,[1] and Joshua D. Reiss[1]

[1] *Centre for Digital Music, Queen Mary University of London, London, E1 4NS, UK*

Correspondence should be addressed to Sonia Wilkie (sonia.wilkie@eecs.qmul.ac.uk)

ABSTRACT

The presentation of objects moving in depth towards the viewer (looming) is a technique used in film (particularly those in 3D) to assist in drawing the viewer into the created world. The sounds that accompany these looming objects can affect the extent to which a viewer can perceptually immerse within the multidimensional world and interact with moving objects. However the extent to which sound parameters should be manipulated remains unclear. For example, amplitude, spectral components, reverb and spatialisation can all be altered, but the degree of their alteration and the resulting perception generated, need greater investigation. This paper presents the results from an investigation into one of the sound parameters used as an audio cue in looming scenes by the film industry, namely amplitude, reporting the degree and slope of its manipulation.

1. INTRODUCTION

One of the features of 3D presentation that entices viewers to attend a 3D screening, as opposed to a 2D screening, is the opportunity to see objects appear to leap out of the screen towards the viewer.

This presentation of objects moving through a multidimensional space assists in drawing the viewer into the created world and making it appear more immersive, not only by presenting the third dimension of depth and bringing particular objects closer to the viewer, but also by transforming the experience from a passive one of motionless watching and listening, to an active one in which viewers often physically move to avoid objects as an instinctive reaction to their perceived increasing proximity. Whilst the image representation of the object's movement in depth is often the focus of amazement, the generation of a rich perception of the event is dependent on the simultaneous presentation and integration of both sound and image, and the degree to which the sound accurately represents the object's movement.

2. BACKGROUND

A great deal of research is currently being undertaken on the presentation of 3D images, and the multi-speaker spatial arrays for 3D sound transmission [1]. However research also needs to be undertaken to develop post-production tools and sound manipulation techniques to maximise the viewer's experience.

Scientific studies that have analysed different sound parameter's effect on the perception of approaching virtual objects have been extremely limited with respect to the parameters investigated ([2], [3], [4], [5], [6]). These studies often use simple sound signals (sine or triangle waves at 400 to 1000 Hz), and limit the sound parameter manipulation variables to amplitude increase, frequency change, and interaural temporal differences.

An increase in the amplitude has been the core (sometimes only) variable in previous auditory looming studies, with the level being increased by 10 to 30 dB ([2], [4], [5]). However, the manipulation of this sound parameter was not the focus of these studies, and limiting the variables used in these studies compromises the ecological validity of the results, the parameters manipulated, and the real world application. As such, we have chosen this audio cue for a detailed analysis, for this first study.

3. FEATURE ANALYSIS

Analysis of the sound manipulation techniques that sound designers and post-production audio engineers use as cues to depict an approaching object, sheds light on current practice in the film industry, by which engineers seek to maximise the impact and level of immersion of viewers.

This will provide a broader range of sound parameter variables for closer investigation in scientific studies, that will provide insight on how sound affects viewer's perception of approaching objects. It will also provide the foundation for developing specifications for techniques on generating sound moving in depth, based on the desired perceptual and immersive experience.

Comparison between films of the amplitude level, degree of manipulation, and the slope of change (linear or non-linear), reveals similarities between the techniques used, and the degree of manipulation.

We propose that these levels, degree of manipulation, and the slope of change used, can form the basis for specifications on the manipulation of sound (for an object's movement in depth) according to the desired viewer experience.

4. STUDY

4.1. Aim

The aim of this study is to determine

- If a change in amplitude is used as a cue to suggest an approaching object
  - how much it changes;
  - if the amount of change is dependent on the sample duration.
- If there is a common amplitude (dB) level that sound designers use at the start, and the peak of the sound.
- If the slope of the amplitude increase is manipulated as a cue to suggest an approaching object;
  - if so, how much it is increased;
  - if the increase is relative to the sample duration.

4.2. Hypothesis

It was hypothesised that sound designers manipulate amplitude as one of a number of sound parameters to generate the percept of an approaching object. More specifically, it was hypothesised that the

- Amplitude would change by increasing
  - to the point of impact or passing;
  - at a level similar to the auditory looming studies;
  - relative to the time duration, i.e. that shorter durations would allow less time, therefore would have less amplitude increase, than longer durations.
- Amplitude levels
  - would be similar to those used in the previous auditory looming experiments;
  - would be different per duration band (specifically that short durations would commence at a higher level, and peak at a lower level, than longer durations).

- Amplitude would increase on a linear slope (representative of the sound designers physical movement of sliding a fader on a mixing desk; or in audio editing software, the linear crossfade modifying the change in amplitude level).

4.3. Method

4.3.1. Stimuli

Looming scenes were selected from 27 films scenes and are listed in Table 1. Each sound source was a stereo track between 1 and 3000 ms in duration.

4.3.2. Apparatus and Procedure

The audio signal (.wav) of each film looming scene sample was analysed (using the MIRtoolbox v.1.3.4 [10] for MATLAB) to measure the signal level, over time. For this initial study, the analyses conducted were on amplitude change (over time); the minimum and maximum amplitude levels; and the slope of the amplitude change.

4.4. Results

4.4.1. Amplitude Change

As expected, all of the samples increased the amplitude as the proximity of the object became closer. The minimum and maximum amplitude (dB) level for each looming sample is listed in Table 1, as well as the total increase (dB), and the duration of the change.

The average total increase (in dB) across all of the looming scene durations (1 - 3000 ms) was calculated at 62.68 dB, standard deviation = 15.49; minimum = 32.47 dB (Final Destination); and maximum = 89.7 dB (I Am Legend).

The amplitude increase per sample duration is plotted in Figure 1 (the looming duration was measured from the start of the sample to the amplitude peak). The longest looming duration was 2995 ms (The Day After Tomorrow) and the shortest duration was 313 ms (Sin City). The spread of the data shows there were more looming scene samples under 1500 ms (74.08 % of total samples) than over, and the number of samples under 1500 ms were spread almost almost equally within that time span, with 5 samples for the $\leq$ 500 ms band, 8 samples for the 501 to 1000 ms band, and 7 samples for the 1001 to 1500 ms band. There is also a general trend for samples to cluster around the 60 dB level, and neither increase or decrease in amplitude level as the sample duration increases. This is supported by a linear regression analysis performed on the data, which suggested a small increase in the amplitude of 1.99 dB for each 1000 ms increase in the sample duration, however the correlation is weak, with a coefficient of determination $R^2 = 0.01$.

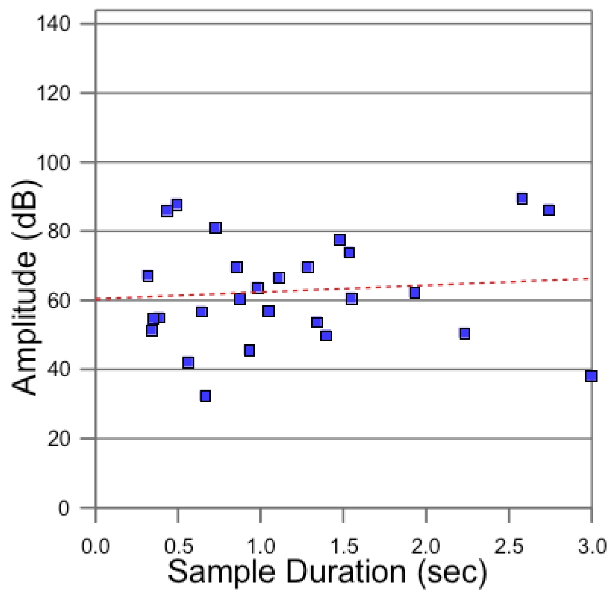

Fig. 1: Amplitude change per sample duration. Linear regression equation of line: $y = 1.99x + 60.34$; $R^2 = 0.01$.

Amount of Amplitude Increase per dB Band

The amount of amplitude increase was subcategorised into band levels (at 10 dB increments) and is listed in Table 2. The results show that the increase in amplitude of between 60 to 70 dB, and 50 to 60 dB occurred most often (8 and 7 samples respectively, totalling 55.57% of all samples), and that increases using bands $\leq$30 dB, and 90.01 to 100 dB occurred least often (with zero samples).

Amplitude Change Discussion

Comparison of all of the looming scenes analysed indicate that they all increased the amplitude, as is consistent with the physics of an approaching object, and its application as a variable in the previ-

#	Title, Year	Chapter, Time (min : sec)	Decibel Level Min	Max	Total Incr.	Looming Duration (sec)	Linear Slope m value	dB Incr. per 100ms
1	The Matrix (1999)	Ch. 1, 1:22 - 1:25	-116	-49.31	66.69	1.103	27	2.7
2	Return of the Jedi (1983)	Ch. 3, 0:20 - 0:24	-130.9	-61.08	69.82	1.277	25	2.5
3	Revenge of the Sith (2005)	Ch. 31, 3:08 - 3:09	-132.6	-68.78	63.82	0.975	34	3.4
4	X-men (2006)	Ch. 15, 0:35 - 0:36	-120.2	-34.04	86.16	0.430	162.18	16.2
5	The Day After Tomorrow (2004)	Ch. 12, 2:29 - 2:33	- 88.67	-50.43	38.24	2.995	2.8	0.28
6	King Arthur (2004)	Ch. 7, 10:46 -10:48	-117.4	-60.46	56.94	1.045	26	2.6
7	Sherlock Holmes (2009)	Ch. 22, 4:36 - 4:38	-125.3	-55.61	69.69	0.847	49	4.9
8	Van Helsing (2004)	Ch. 17, 1:52 - 1:54	-106.2	-52.35	53.85	1.335	17	1.7
9	I Am Legend (2007)	Ch. 17, 0:00 - 0:03	-143.9	-54.2	89.7	2.577	28	2.8
10	Troy (2007)	Ch. 27, 2:22 - 2:24	-105.3	-54.67	50.63	2.229	4.1	0.41
11	Beowulf (2007)	Ch. 2, 4:03 - 4:05	-115.4	-37.67	77.73	1.474	17	1.7
12	The Bourne Identity (2002)	Ch. 12, 2:10 - 2:12	-106.9	-50.06	56.84	0.639	27	2.7
13	Charlie & the Chocolate Factory (2005)	Ch. 15, 1:24 - 1:26	-114.6	-59.5	55.1	0.383	88	8.8
14	Mr and Mrs Smith (2005)	Ch. 20, 0:40 - 0:44	-122	-61.51	60.49	0.871	19	1.9
15	Sin City (2005)	Ch. 18, 1:06 - 1:07	- 87.9	-20.6	67.3	0.313	181.57	18.16
16	28 Days Later (2002)	Ch. 11, 0:01 - 0:04	-135.7	-61.71	73.99	1.532	21	2.1
17	Gattaca (1997)	Ch. 21, 2:39 - 2:40	-117.2	-35.92	81.28	0.720	75	7.5
18	Alice in Wonderland (2010)	Ch. 15, 0:19 - 0:20	- 95.73	-40.77	54.96	0.348	68	6.8
19	Avatar (2009)	Ch. 22, 1:42 - 1:45	- 77.79	-26.24	51.55	0.337	124.84	12.48
20	Clash of the Titans (2010)	Ch. 13, 4:11 - 4:13	-115.6	-65.69	49.91	1.381	19	1.9
21	Despicable Me (2010)	Ch. 18, 2:23 - 2:24	-112.2	-69.87	42.33	0.557	30	3.0
22	Kill Bill vol2 (2004)	Ch. 6, 0:03 - 0:06	-120.5	-58.27	62.23	1.927	9.2	0.92
23	Mission Impossible 3 (2006)	Ch. 4, 1:06 - 1:08	- 95.86	-50.26	45.6	0.929	28	2.8
24	Yogi Bear (2010)	Ch. 1, 1:25 - 1:27	-115.8	-55.05	60.75	1.544	18	1.8
25	Final Destination (2009)	Ch. 15, 0:06 - 0:07	- 77.67	-45.2	32.47	0.662	43	4.3
26	Salt (2010)	Ch. 9, 3:13 - 3:14	-131.4	-43.45	87.95	0.488	116.75	11.68
27	Saving private ryan (1998)	Ch. 19, 3:17 - 3:21	-143.9	-57.63	86.27	2.740	26	2.6

Table 1: List of the film scenes that were used in the study, with the DVD chapter and time. Other columns detail the minimum and maximum amplitude levels measured for each film; the total amplitude increase; the duration of the looming scene; the slopes linear equation m value; and the linear slope increase per 100 ms ($m \times 0.1$).

dB Band	Number of samples	% of Total Samples
≤ 30 dB	0	0.00 %
30.01 - 40 dB	2	7.41 %
40.01 - 50 dB	3	11.11%
50.01 - 60 dB	7	25.93 %
60.01 - 70 dB	8	29.63 %
70.01 - 80 dB	2	7.41 %
80.01 - 90 dB	5	18.52 %
90.01 - 100 dB	0	0.00 %

Table 2: The number of looming samples increase in amplitude, per decibel band, across all of the sample durations.

ous auditory looming studies. However, the amount of amplitude increase used in the previous auditory looming experiments ranged from a 10 dB increase ([2], [6]) to a 30 dB increase ([4], [5]), whereas the amount of amplitude increase in the film samples, was significantly greater with an average increase of 62.68 dB.

When categorised into amplitude increase band levels, the increase in the amplitude was most often in the 60 to 70 dB, and 50 to 60 dB bands, as opposed to the looming experiments (at 10 to 30 dB). Further, none of the film samples used this low band level, and only two of the film samples (Final Destination at 32.47 dB, and The Day After Tomorrow at 38.24 dB) were even close to the 30 dB maximum level used in the experiments. And these two particular film samples were comprised of broadband noise (depicting images of snow and air) and were not focused sounds, such as the simple triangle and sine tones as used in the experiments.

It was also hypothesised that the amount of amplitude increase may be dependent on the duration of the looming scene, with longer scenes allowing a greater total increase of amplitude, and that sound designers would exploit this opportunity. This hypothesis was rejected however, with no general change (either increasing or decreasing) in the amplitude over the sample duration, that was further supported by the linear regression analysis which indicated there was no correlation between the amount of amplitude increase and the duration of the sample.

4.4.2. Amplitude Levels

The minimum and maximum amplitude level per film were taken from Table 1 and averaged. The average minimum level was -113.8 dB (standard deviation = 17.87; minimum = -143.9 dB (I am legend); maximum = -77.67 dB (Final Destination)); and the average maximum level was -51.12 dB (standard deviation = -12.35; minimum = -69.87 dB (Despicable Me); maximum = -20.6 dB (Sin City)).

The minimum and maximum levels in Table 1 were further categorised according to dB bands (at increments of 10 dB), with the number of samples per band tallied (see Table 3).

The results indicate that the minimum level that sounds commenced most often was the -110 to -120 dB band (with 8 samples, 29.63% of the total samples), and the maximum level that sounds most often peaked was the -50 to -60 dB band (with 11 samples, 40.74% of the total samples).

dB Band	Samples Min dB		Samples Max dB	
	#	%	#	%
0 to -10	0	-	0	-
-10 to -20	0	-	0	-
-20 to -30	0	-	2	7.41 %
-30 to -40	0	-	3	11.11 %
-40 to -50	0	-	4	14.81 %
-50 to -60	0	-	11	40.74 %
-60 to -70	0	-	7	25.93 %
-70 to -80	2	7.41 %	0	-
-80 to -90	2	7.41 %	0	-
-90 to -100	2	7.41 %	0	-
-100 to -110	3	11.11 %	0	-
-110 to -120	8	29.63 %	0	-
-120 to -130	4	14.81 %	0	-
-130 to -144	6	22.22 %	0	-

Table 3: The number of looming samples increase in amplitude, per decibel band (at 10 dB increments), across all of the durations (1 - 3000 ms), and the percentage of total samples.

Amplitude Levels Discussion

The amplitude level that the looming commences (minimum level) and peaks (maximum level) may act as a cue to the physical properties of the moving object, such as the distance of the object, proximity, or speed at which it is travelling.

Previous experiments on auditory looming presented the audio signal at more moderated levels, using minimum (starting) levels, and maximum (peak) levels of 55 to 75 dB [8], 65 to 75 dB [2], 40 to 70 dB and 60 to 90 dB [4] [5], and 77 to 87 dB [6]. However the results from the film samples were much more extreme in the minimum and maximum levels used, with the average minimum level at -113.8 dB, with sounds commencing at a level between -110 and -120 dB occurring most often, and peaking at an average maximum level of -51.12 dB, with sounds peaking at a level between -50 and -60 dB occurring most often.

4.4.3. Amplitude Slope

Linear, quadratic and cubic equations were applied to the amplitude level to determine the line of best fit. The measurement of the slope line for each film sample was made between the amplitude start, and the amplitude peak.

Observation of the line of best fit suggests that the amplitude increase has a linear or near-linear relationship, and not a higher order relationship, so further analyses of the polynomial equations was not conducted.

Linear Equation Results

For each film sample, the linear equation's first coefficient (m value) was multiplied by 0.1 ($m \times 0.1$) to determine the amount of amplitude increase (dB) per 100 ms (and is listed in Table 1, columns 'Linear slope m value' and 'Linear slope dB increase per 100ms').

The average total increase in amplitude per 100 ms (across all of the looming scene durations (1 - 3000 ms)) was 4.76 dB (standard deviation = 4.75 dB, minimum = 0.28 dB (The Day After Tomorrow), maximum = 18.16 dB (Sin City)).

Linear Slope Regression Analysis

Regression analysis was also performed on the data, of the amplitude change (dB) over 100 ms, per sample duration, and is plotted in Figure 2.

Looking at the data, the spread appeared to be more non-linear than linear, so both linear and non-linear regression analyses were performed on the data.

The linear regression equation of line was calculated to be $y = -4.2x + 10$, and the co-efficient of determination $R^2 = 0.43$.

The nonlinear (logarithmic) regression equation line $y = -3.52x + 4.25$ had a better fit, with the co-efficient of determination $R^2 = 0.51$ suggesting a stronger correlation.

Overall, the lines of regression show a decreasing trend in the amount of 'amplitude (dB) increase over 100 ms', with shorter duration samples having a greater increase in amplitude (dB) over 100 ms, than the longer duration samples.

Linear Amplitude Increase per dB Bands

Subcategorisation of the 'linear slope dB increase per 100 ms' results from Table 1, into amplitude increase bands (at 1 dB increments) is listed in Table 4. The number of samples were tallied and the results indicated that an increase of between 2 to 3 dB, and 1 to 2 dB, per 100 ms occurred most often (7 and 5 samples respectively, totalling 44.45 % of all samples).

Linear Amplitude Increase per Sample Duration Bands

Further categorisation was again made of the 'linear slope dB increase per 100 ms' results from Table 1, into sample duration bands (at 500 ms increments) and is plotted in Figure 3. The results show that the ≤ 500 ms band had the greatest average increase

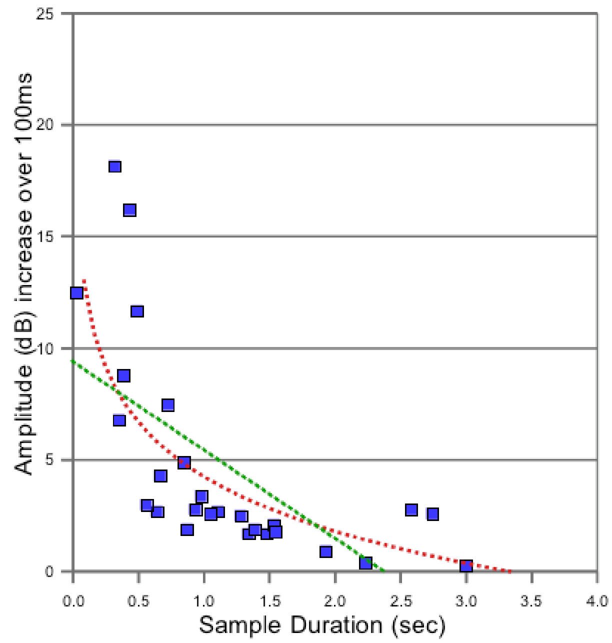

Fig. 2: Amplitude change (dB) over 100 ms, per sample duration. Linear regression equation of line: $y = -4.2x + 10$; $R^2 = 0.43$; Nonlinear (log) regression equation of line: $y = -3.52x + 4.25$; $R^2 = 0.51$.

dB Band	Number of samples	% of Total Samples
≤ 1 dB	3	11.11 %
1.01 - 2 dB	5	18.52 %
2.01 - 3 dB	7	25.93 %
3.01 - 4 dB	3	11.11 %
4.01 - 5 dB	2	7.41 %
5.01 - 10 dB	3	11.11 %
10.01 - 15 dB	2	7.41 %
15.01 - 20 dB	2	7.41 %

Table 4: The number of looming samples (linear) increase in amplitude over 100 ms, per dB band, across all of the sample duration's (1 - 3000 ms).

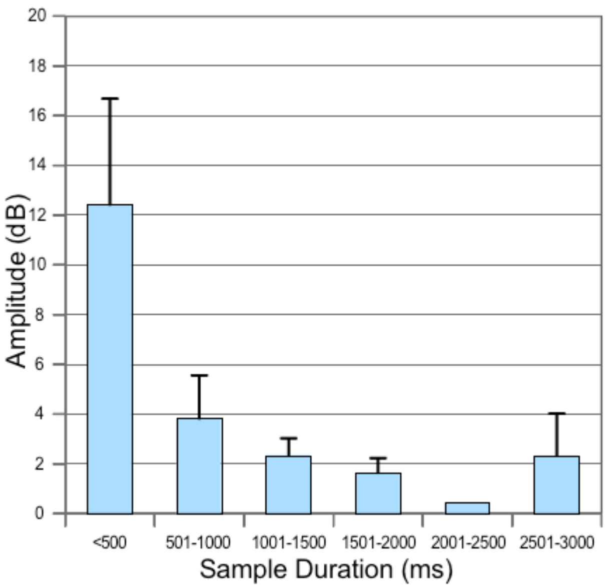

Fig. 3: The average (linear) increase of amplitude (in dB) over 100 ms, per duration band. The samples were categorised into duration bands (at 500 ms increments). The average increase in amplitude (dB) over 100 ms, was made for each duration time band, and the standard deviation is indicated by the error bars.

in amplitude (mean = 12.40, standard deviation = 4.29, minimum = 6.8, maximum = 18.16).

Amplitude Slope Discussion

The slope of the amplitude increase may bias perception of an approaching object, and when people expect the object to reach them. A nonlinear slope may suggest that the approaching object is accelerating, therefore people may illicit a faster response to contact time than for a linear increase, which may suggest a constant (and slower) speed of approach.

It is unknown whether the previous studies increased the amplitude on a linear or nonlinear slope. However this analysis of the film samples suggests that sound designers tended to apply a linear slope, which may be a reflection of the DAW impact on sound design.

5. CONCLUSION

This initial study on the manipulation of amplitude in depicting an approaching object has shown a number of similarities among the techniques used by sound designers, and how they contrast to those used by psychologists in the psychoacoustic auditory looming experiments. In general, the ranges and levels of manipulation used by the sound designers were much greater than those used by the psychologists, which may reflect the intention of the sound designers to maximise viewers experience, as compared to the psychologists who may have used a conservative amount to simply elicit a response.

Further research will investigate peoples responses to the stimuli, to determine if the greater levels of manipulation bias people's perception of the approaching object, and also into other sound parameters that are used as cues to generate the perception of an approaching object.

6. REFERENCES

[1] Morrell, M. J., Baume, C., and Reiss, J. D. (2012). "VAMBU Sound: A mixed technique 4-D reproduction system with a heightened frontal localisation area", presented at the AES 25th Conference, York, UK.

[2] Rosenblum, L, Carello, C, & Pastore, R. (1987). "Relative effectiveness of three stimulus variables for locating a moving sound source", Perception, 16, 175186.

[3] Rosenblum, L, Wuestefeld, A, & Saldana, H. (1993). "Auditory looming perception: Influences on anticipatory judgements", Perception, vol 22, 1467-1482.

[4] Neuhoff, J.G. (2001). "An adaptive bias in the perception of looming auditory motion", Ecological Psychology, 13 (2), 87-110.

[5] Neuhoff, J.G., & Heckel, T. (2004). "Sex differences in perceiving auditory looming produced by acoustic intensity change", In Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia.

[6] Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). "Selective integration of auditory-visual looming cues by humans", Neuropsychologia, 47, 1045-1052.

[7] Ghazanfar, A.A., Neuhoff, J.G., & Logothetis, N.K. (2002). "Auditory looming perception in rhesus monkeys", In Proceedings of the National Academy of Sciences, USA 99, 15755-15757.

[8] Maier, J, Neuhoff, J, Logothetis, N, & Ghazanfar. A. (2004). "Multisensory integration of looming signals by rhesus monkeys", Neuron, 43 (2), 177-181.

[9] Neuhoff, John G. (2004). "Ecological psychoacoustics: introduction and history", Ecological Psychoacoustics. Elsevier Academic Press, California.

[10] Olivier L., and Toiviainen, P. (2007). "A Matlab Toolbox for Musical Feature Extraction From Audio", International Conference on Digital Audio Effects, Bordeaux.