GENERATING MEANINGFUL SOUND: QUANTIFYING THE AFFECTIVE ATTRIBUTES OF SOUND EFFECTS FOR REAL-TIME SOUND SYNTHESIS IN AUDIO-VISUAL MEDIA

KAREN COLLINS[1]

[1] *Canadian Centre of Arts and Technology, University of Waterloo, Waterloo, Canada*
collinsk@uwaterloo.ca

Much research has been undertaken to discover what parameters in a musical composition carry emotional meaning. We now take for granted that harmonic content, instrumentation, tempo, timbre, pitch range, and dynamics (etc.) all play some role in music's affective abilities. However, there has been little research into similar aspects of affect when it comes to sound effects. Though many audio synthesis methods strive for greater realism, realism is not always the most believable sound in multi-media situations. This paper seeks to explore a methodology for research into the affective attributes of sound effects. An understanding of these affective elements can lead to more advanced sound synthesis methods for audio-visual media.

INTRODUCTION

The on-going drive towards greater realism in sound synthesis methods, while certainly invaluable, fails to take into account the fact that in sound for audio-visual media (such as film or games), the most realistic sound is not always the most effective or the most appropriate. Sounds can be metaphoric and meaningful on levels other than just as an attempt to approximate reality, contributing to the overall affective power of the audio, as sound designer Walter Murch [1] argues; "This metaphoric use of sound is one of the most flexible and productive means of opening up a conceptual gap into which the fertile imagination of the audience will reflexively rush, eager (even if unconsciously so) to complete circles that are only suggested, to answer questions that are only half-posed". In video games, even realistic sampled sounds are often treated with various digital signal processing (DSP) effects to make them 'more real than real'. Substitutions of unrealistic sounds or exaggerated sounds can become realistic for the audience, and have a greater impact than straight realism. In other words, sound in audio-visual media is as much an aesthetic choice as it is a reproduction of the imagined space. Sound synthesis methods, therefore, to be effective, need to create a balance between realism on the one hand and affect on the other.

1 SOUND *E*FFECTS AND *A*FFECTS

Much research has been undertaken in a variety of fields to discover the various parameters in a musical composition that can help to convey affective meaning to the listener. We now take for granted that harmonic content, instrumentation, tempo, articulation, timbre, pitch range, and dynamics all play some role in music's affective abilities. However, there has been significantly less research into the affective role of sound effects. An understanding of sound's affective properties can be used to set parameters in the real-time synthesis of sound effects. But which parameters of a sound effect can be altered to change the meaning of the sound, and is it possible to quantify these parameters? Moreover, what parameters of a sound effect can be changed in real-time *without* altering the meaning of the sound, to offer a player constant variation? For example, footstep sounds are ubiquitous and a now-expected stylistic trait of First Person Shooter games. However, with thousands of footstep sounds required for a game, those games that rely on unaltered sampled sound can expect to get short mileage out of those sounds. By adjusting parameters of the sound in real-time, however, we could theoretically show an increased weight in objects being carried, a change in mood, player health, and so on. However, we also need to know the extent to which a sound can be changed, so that when a sound's elements (such as DSP effects) are changed in real-time for the sake of variability, the sound is not altered to the extent that the player keys in on the change and spends time searching for some imagined associated narrative change.

Assuming that it is possible to in some way quantify the parameters that influence the reception of sound, how do we determine these factors or parameters? The study of sound perception is challenged by countless variables that affect study results (see below). The most significant variable in the context of video games is, of course, how the image will influence the reception of the sound (and vice versa). What methods can be used to determine meanings when sound is combined with imagery or narrative content? This paper seeks to lay some foundation for exploration by introducing some

methodologies for research into the affect of effects, exploring whether or not it is possible to quantify affective elements of a sound effect, what methods we can use to attempt to quantify these elements, and how we might determine which factors or parameters play a role in sound's affective properties. A brief introduction to the issue is presented first, followed by a series of prototype tests and results showing the reception of sound effects.

2 SOUND EFFECTS PERCEPTION

A variety of branches of research have been concerned with the study of sound perception. These include for instance ecological acoustics (largely concerned with the perception of the metrical properties of sound that enable us to determine size, distance, species recognition, shape and material composition), psychomechanics (including the perception, recognition and identification of vibrating objects in a three dimensional environment)[2], auditive kinetics (the study of the relationship between objective kinetic properties of sound and the perception of these properties)[3], sound semiotics (the study of sounds as a form of symbolic communication), and psychoacoustics. In other words, it is possible to study sound perception from a great variety of angles, and each of these has potential to contribute valuable information to the study of sound perception. Since we are focused on the affective properties of sound, it is the semiotic properties of sound research that are of particular interest here.

2.1 Factors influencing the perception of sound effects

Sound perception is difficult to quantify, since it is influenced by many factors. Some recent studies have been undertaken in music reception in particular that seek to understand how various elements obscure perceptual quality. Lewis and Schmidt [4], using the Myers-Briggs Type Indicator, concluded that personality type had an influence on musical reception. It has been proposed that pre-existing emotion or mood influences the following affective evaluation of an auditory event [5]. Gregory and Varney show significant differences between British and Indian listeners in their perception of music [6], while Iversen *et al.* show the influence that nationality has on the perception of rhythmic grouping [7]. Grouping itself can influence the perception of a sound, as has been studied since at least Bolton [8], who determined that loud sounds tend to mark the beginning of a group, and a lengthened sound or interval between sounds tends to mark the end of a group. Expertise can also affect reception. Neuhoff *et al.* [9] argue that the musical training of subjects influences conceptual relationships to pitch, for example. Sound's affective ability, then, lies in its cultural and personal associations as well as its bio-acoustic and physical associations. In other words, we can never arrive at a single definitive meaning or affective value of any sound, but rather through collecting large amounts of data we might arrive at a *series of assumptions* that can be brought into synthesis methods.

2.1.1 Sound and Image

There is perhaps little debate that sound effects in combination with a visual image (such as in a film or a game) can change the affective response to either the sound or the image. In many of cases of audio for visual media, sound reinforces what is occurring on-screen, and the audience is subtly affected by the sounds while they focus on the image and narrative. In television advertising, sound plays a critical role when substituting for other senses, such as taste or smell. Sound is closely allied with image in audiovisual media: so close that Moncrieff *et al.* [10] successfully wrote algorithms using ADSR information from sound effects as a search tool for suspense scenes in horror film.

The study of the influence of visual stimuli on the perception of sounds has been a relatively recent endeavour, with a majority of the work focused on the perception of environmental sounds, with particular attention paid to loudness or noise, such as for instance wind turbines [11], cars [12], and general urban soundscapes [13]. All of these studies concluded that image had a notable influence on the perception of loudness or other unpleasantness. Cox [14], in a large internet experiment, found that images affected the perception of 'horrible' sounds (dentist's drill, fingernails scraping, coughing, and so on).

The influence sound has on our perception of images is also of note here. Abe *et al.* [15] found that participants responded more positively to white noise when accompanied by an image of a waterfall. More relevant to video games, Mastoropoulou *et al.* [16] found that sound had a considerable response on the perception of motion smoothness in an animation (that is, frame rate).

This is not to suggest that sounds devoid of associated image lack connotative ability. Bisping [17] found an increase in the height of ADSR envelope was associated with unpleasant feelings, while fast attacks were associated with power. Nevertheless, most theorists agree that even when devoid of context, sounds cannot be separated from imagery because sounds have a tendency to evoke imagery in our minds, set in what composer Pierre Schaeffer referred to as the *écran sonore* or 'sound screen'. Elsewhere, I have shown that listeners hearing decontextualized sound effects often associate them with the physical actions that cause the sounds (such as striking or plucking)[18]. Dubois has

likewise shown that listeners respond to sounds by trying to associate them with their emitters (source) or a meaningful event [19].

3 METHODOLOGICAL APPROACHES TO SOUND PERCEPTION

Working on the assumption that sound effects do in fact have affective properties, how do we determine what factors or parameters play a role in this communication of affective meaning? There are a large variety of methods that have been used previously to test perceptual quality of both music and sound effects. Physiological approaches (EECs, EEGs, skin conductance response, respiration, heart rate, psychometric ratings for instance) such as those undertake by Paanksepp [20] provide us with basic physiological affective data, but cannot tell us what the sound is expressing in terms of other or cognitive meanings. Cognitive research on human perception and evaluation of sound have traditionally sought to establish the underlying perceptual dimension that people use when evaluating sounds [21]. These studies have focussed on physical, acoustical perception, psycho-acoustical perception and psychological studies. Interobjective approaches to musical meaning involving the comparison of a quantity of previously existing audio visual material have been undertaken by Tagg [22], who has shown that particular musical inflections occur in similar narrative contexts in film and television. Similar approaches to sound effects might include comparisons of asset lists for video games. Another possible approach would be to quantitatively or qualitatively compare the advertising of these sounds in sound effects collections [18], which are occasionally described by the creators in terms of their affect. Subjective approaches on the other hand are more commonly used, and rely on quantitative and qualitative data from groups of listeners.

3.1 Using Games for Subjective Sound Reception Studies

Since the study of the semiotic nature of sound is so subjective and depends to a large extent on a user's personal experiences, a *distributed classification* system (also known as collaborative tagging) is a logical step in a quantitative study of sound perception. Distributed classification is a method of collecting a large number of responses of multiple users, commonly used for meta-data. This approach can be seen in popular image- or content-tagging websites such as Flickr, Del.icio.us, and so on. Users tag media objects with text keywords in a free-association fashion. Tags can then be combined into non-hierarchical groups of associated terminology (a *folksonomy*), which can be accessed and searched by users. In order to engage the audience and increase the amount of tags collected, tagging games can be created. Von Ahn and Dabbish [23] created a way of encouraging users to accurately and effectively tag images through the *ESP Game*, in which players are paired with an unknown online partner, shown an image, and must guess the ways in which their partner will describe the image, thus tagging it with metadata that both users have agreed upon. Approaches that encourage players to tag musical data have also been developed [24][25].

Distributed classification is an affordable and efficient form of collecting data on multimedia objects such as audio content on the Internet. Moreover, the processes involved create a built-in feedback system to correct errors. Despite the many advantages of a distributed classification system, however, there are also some disadvantages. These include the ways in which users tag data in a free-association system, which can lead to ambiguities, imprecision, synonyms/homonyms, and so on. User-generated tagging can also lead to misinformation, or spamming (i.e. intentionally using inappropriate terminology). There are also problems with the lack of control over volume, and the fact that judgements on sound are made in a particular context in front of a computer [14]. Nevertheless, findings can be confirmed with other methods at a later time.

3.2 Prototype games

Two interactive online game prototypes were constructed for the purpose of this initial study into the perceptual qualities of sound effects, "The Psychic Psychiatrist" and "Roar-Shock". These games were built around the idea of collecting information on the communicative and affective role that five DSP pre-set effects had on sound effects. These effects were distortion, phaser, delay, reverb and flange from Logic Studio[26].

In the first prototype, The Psychic Psychiatrist, players are informed that they are a "psychic psychiatrist" and are to guess what their patient's responses to the questions "How does this make you feel?" and "What images does this conjure in your mind?" will be in regards to a sound object that they can both hear (randomly extracted from the game's database). They each type in keywords for fifteen seconds, whereupon the number of keywords matching existing words in a database set up by the researchers is added to their score before going to the next level (i.e. the next sound). If they fail to make a match, they lose that game and must begin again (with new randomly selected audio content).

Roar-Shock, the second prototype game, is in the guise of a personality type test. Players hear the same sound with the five DSP settings and see five images, and are to indicate which sounds are most appropriate for each image. For example, the same footstep sample is treated with the five DSP effects. Players choose which character is most likely to be making the footsteps. At the end of the test, they are told how many other players shared similar responses, and a suggested personality profile is created for them based on their responses (completely arbitrary and fun). Their IP is logged, and each IP's data is only stored once, although they may repeat the game.

3.2.1 Preliminary results

Our preliminary testing was undertaken with one hundred sound effects (twenty unique samples times five DSP effects) and twenty students who were participating in a sound course at the University of Waterloo. Results from The Psychic Psychiatrist showed a remarkable similarity in response to sound effects. While many responses unsurprisingly related sounds to their emitters (airplane samples were heard as airplanes, for instance) in response to the question "What images does this conjure in your mind?", the responses to "How does this make you feel" were also fairly consistent. For example, distortion on the airplane sound (a commercial aircraft sample) was "unsettling" and "unnerving" or had similar semantically related responses in 18 out of the 20 respondents. On the other hand, phasing on the same sound was "trippy", "spacey", or similar in all twenty responses.

Results from Roar-Shock showed that it was also possible to quantify the changes in broad terms across the spectrum of respondents and sounds. Reverb was often associated with larger and older imagery. Delay was associated with futuristic imagery and space, as was phasing and flanging, though to a lesser extent. These sounds were also associated with hard edged shapes, and metallic coloured shapes. Distortion was associated with the largest imagery, fuzziest images and 'dirtiest' images.

While the prototype test on twenty people and twenty sounds cannot show conclusive results on the actual responses or the impact that different DSP effects have on sound perception, what we can deduce is that it is, in fact, possible to collect large amounts of data on audio affect through the use of distributed classification games. What is now necessary is to refine the games and the study samples into ways to scientifically test the affect of sound effects.

4 CONCLUSIONS AND FURTHER STUDY

The prototype tests revealed that distributed classification games are a useful way to collect data on audio affect. The two different prototype games have provided us with a broad overview which must now be refined. The next stage in the research is to create more detailed games that will involve larger amounts of data, as well as varying degrees of effects. In this way, we might know *how much* and *what type* of reverb effect may elicit a change in response. We can therefore quantify at what point a sound is perceived to change the meaning or affect of an image. We will continue to collect data through the initial prototypes in the meantime over a large-scale internet study.

We believe that in addition to this quantitative data, it is also necessary to compare smaller amounts of qualitative data. Since we are particularly focused on the creation of affective audio for video games, we will be implementing the results of the data into a video game for further testing by hypothetical commutation. For this purpose, we will be using audio middleware company Audiokinetic's demonstration first person shooter game, *Forbidden Terror on Station Z.* In this way, we can incorporate a series of sound effects that have been treated with varying degrees of DSP effects in real time, and test user responses through think-aloud protocols and eye tracking software.

Using a variety of testing methods, therefore, we believe it is possible to quantify the affective qualities of sound effects, and to implement this data in the real-time sound synthesis in video games.

REFERENCES

[1] W. Murch, "Womb Tone" *The Transom Review* vol. 5, no. 1. http://transom.org/guests/review/200504.review.murch.html (2005).

[2] W. McAdams, "Recognition of sound sources and events" *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by S. McAdams & E. Bigand (Oxford University Press: Oxford), pp. 146-198 (1993).

[3] R. Guski, "Studies in auditive kinetics" *Contributions to Psychological Acoustics. Results of the 8th Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, M. Meis & C. Reckhardt (Oldenburg: BIS), pp. 383-401 (2000).

[4] B. E. Lewis & C.P. Schmidt, "Listeners' Response to Music as a Function of Personality Type" *Journal of Research in Music Education*, vol. 39, no. 4, pp. 311-321 (1991).

[5] J. Blauert & U. Jekosch, “Sound Quality Evaluation - A Multi-Layered Problem” *ACUSTICA - Acta Acustica*. vol. 83, no.5, pp. 747-753 (1997).

[6] A. H. Gregory & N. Varney, “Cross-Cultural Comparisons in the Affective Response to Music” *Psychology of Music* vol. 24 no. 47 (1996).

[7] J. R. Iversen, A. D. Patel & K. Ohgushi, “Perception of rhythmic grouping depends on auditory experience.” *Journal of the Acoustical Society of America*, vol. 124 no. 4, pp. 2263–2271 (2008).

[8] T. L. Bolton, "Rhythm" *American Journal of Psychology* vol. 6, 145-238 (1894).

[9] J.G. Neuhoff, R. Knight & J. Wayand, “Pitch change, sonification, and musical expertise: which way is up?” *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto (2002).

[10] S. Moncrieff, C. Dorai & S. Venkatesh, “Affect Computing in Film through Sound Energy Dynamics” *MM* ‘01, Ottawa (2001).

[11] E. Pedersen & K.P. Waye, “Perception and annoyance due to wind turbine noise – a dose–response relationship” *Journal of the Acoustical Society of American* vol. 16 no.6, pp. 3460–70 (2004).

1 [12] D. MENZEL, H. FASTL, R. GRAF, J. HELLBRÜCK, “INFLUENCE OF VEHICLE COLOR ON LOUDNESS JUDGMENTS” THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, VOL. 123 NO. 5, PP. 2477-9 (2008).

[13] S. Viollon, C. Lavandier & C. Drake, “Influence of visual setting on sound ratings in an urban environment” *Applied Acoustics* vol. 63, no. 5, pp. 493–511 (2002).

[14] T. J. Cox, “The effect of visual stimuli on the horribleness of awful sounds” *Applied Acoustics* vol. 69, pp. 691–703 (2008).

[15] K. Abe, K. Ozawa, Y. Suzuki & T. Sone, “The effects of visual information on the impression of environmental sounds” *Inter-noise* vol.99, pp. 1177–82 (1999).

[16] G. Mastoropoulou, K. Debattista, A. Chalmers & T. Troscianko, “The Influence of Sound Effects on the Perceived Smoothness of Rendered Animations” *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, pp. 9–15 (2005).

[17] R. Bisping, “Emotional effect of car interior sounds: Pleasantness and power and their relation to acoustic key features” *SAE thesis 951284*, pp.1203-1209 (1995).

[18] K. Collins, *The Future is Happening Already: Industrial Music, Dystopia and the Aesthetic of the Machine*, Ph.D. Diss, University of Liverpool (2002).

[19] D. Dubois, “Categories as acts of meaning: the case of categories in olfaction and audition” *Cognitive Science Quarterly* vol. 1, pp. 35–68 (2000).

[20] J. Paanksepp, “The emotional sources of “chills” induced by music” *Music Perception* vol. 13, pp. 171-207 (1995).

[21] H. Fastl, “The Psychoacoustics of Sound-Quality Evaluation” *ACUSTICA – Acta Acustica, v*ol. 83, no.5, pp. 754-764 (1997).

[22] P. Tagg, *Kojak: Fifty Seconds of Television Music*: *Towards the Analysis of Affect in Popular Music* Mass Media Music Scholars’ Press, Inc. (2000).

[23] L. von Ahn & L. Dabbish, “Labeling images with a computer game”. *ACM Conference on Human Factors in Computing Systems*, pp. 319–326, (2004).

[24] E. Law. L. von Ahn, R. Dannenberg & M. Crawford, “Tagatune: A Game for Music and Sound Annotation**”** *Proceedings of the International Conference on Music Information Retrieval* http://ismir2007.ismir.net/schedule.html (2007).

[25] D. Turnbull, R. Liu, L. Barrington & G. Lackriet, “A Game-Based Approach for Collecting Semantic Annotations of Music” *Proceedings of the International Conference on Music Information Retrieval* http://ismir2007.ismir.net/schedule.html (2007).

[26] Distortion was set to: drive 21.5 dB, tone 390 Hz, output -0.5 dB. Orange Phaser was set to: feedback 30%, floor 210 Hz, ceiling 1700 Hz, order 12, LFO 1 0.20 Hz, LFO 2 0.92 Hz, Phase +10, LFO Mix 50%/50%, Env Follow 36%, Output Mix +50%. Delay Designer was set to: Total Delay Time 1746 ms, Taps 15, High Pass Cutoff Increases from 0 Hz to 16,800 Hz over the 15 taps, Dry -0.5 dB, Wet -11.0 dB, Pan Center, Reso Decreases from 81% to 14% over the 15 taps, Grid 1/32, Swing 50%. PlatinumVerb was set to: Predelay 0 ms, Room Shape 5 (Pentagon), Stereo Base 1.5 m, Room size 17 m, Initial Delay 12 ms, Spread 116%, Crossover 340 Hz, Low Ratio 66%, Low Freq. Level -5.5 dB, High Cut 6700 Hz, Density 79% , Diffusion 100%, Reverb time 1.65 sec, ER/Reverb

Balance 49%/51%, Dry 50%, Wet 50%, ER Scale 100% and Flange was set to: Feedback 49%, Rate 0.366 Hz, Intensity 13.5%, Mix 80%.