

A need for universal definitions of audio terminologies and improved knowledge transfer to the audio consumer

E. R. Toulson

e.r.toulson@anglia.ac.uk

Lecturer in Audio and Music Technology, Anglia Ruskin University, Cambridge.

Presented at The Art of Record Production Conference, Edinburgh, 8 – 10 September 2006.

Abstract

It is widely regarded that many music producers, engineers and other audio industry professionals are generally employed for their ears, and their ability to articulate what they hear into dialogue and scientific terms. This scenario has generated an element of expertise associated with the language of the music industry professional – but has this in turn created a barrier which impedes the development of audio devices for consumer use?

Modern consumer listening trends have moved towards the use of more functional and convenient electronic devices, often at the expense of audio quality. For example, portable MP3 reproduction devices have enabled consumers to listen to hundreds of audio tracks in any particular order, as well as in random, undefined order. This trend plays fully against the role of the mastering process of record production, where a mastering engineer will meticulously adjust and align audio tracks so that they play on a record in an audibly accurate and artistically pleasing manner.

It could be envisaged that future advances in consumer audio technology will be in high-level control devices that allow consumers to integrate studio production processes into their listening chain. Unfortunately, for these equipment to be commercially viable, the consumer will be required to have an improved understanding of the language and terminology associated with such audio processing. This highlights a knock-on issue in that much audio terminology is metaphoric and only loosely defined scientifically.

In conclusion, to allow the development of advanced devices to enhance audio quality within current consumer listening trends, we should first attempt to develop more universal definitions of audio terminologies. Once this has been done, we can expect consumers to become better educated towards audio processing, thus allowing the development of advanced high-level consumer devices to become more commercially viable.

1 Introduction

It is a widely regarded fact that audio engineers, producers, A&R men etc, are generally employed for their ears, and their ability to articulate what they hear into dialogue and scientific terms. A simple example here is whether someone can distinguish perfect pitch, or perfect time. Pitch and timing are easily quantifiable by scientific means, but other descriptive terms relating to clarity, timbre, intelligibility and audio quality are not generally defined mathematically.

There is an element of expertise associated with the language of the music industry professional – but does this generate a barrier which hampers the development of audio devices for consumer use? Consumer listening trends are rapidly changing towards more convenient media such as MP3 compressed data, which has a severe effect on the reproduced audio quality.

It is anticipated that consumers will look towards advance processing methods within their chosen listening trend to reduce the loss of audio quality as much as possible. Only, for this to be achievable, consumers need to become better educated and aware of audio production methods and the associated technologies and terminologies.

2 Modern listening trends – “quantity above quality”?

Modern digital broadcast methods have allowed much improved systems for home consumer use, such as High Definition broadcast for television. Paradoxically, these advances have been used within audio to reduce audio quality with the aim of increasing the functionality of download and reproduction systems – as argued by Owen (2006).

The compressed data structure of MP3 compressed audio allows consumers immediate access to music through the internet; this is a trend which consumers are currently embracing more and more. Similarly, listening trends are moving towards a more 'quantity rather than quality' approach by users preferring to listen to low quality MP3 audio on a portable device allowing more data or audio to be carried within. The access to MP3s allows users to listen to any track at any time, and often in a random unspecified order.

Effectively, the introduction of MP3 compressed audio for listening goes completely at odds with the formal mastering process applied to a commercial record release. During the mastering stage, the song order and relative output levels are meticulously chosen to give the greatest quality of output – "the last creative step in the process of producing a record album" (Katz, 2002). The reduced audio quality of MP3 versus 16-bit uncompressed audio is further at odds with the whole ethos of quality record production and the development of advanced technology to improve audio systems. Modern digital technology allows us to produce audio at up to 196kHz and 64-bits (approx 25000 kbps) if we like, but what's the point in making this effort if the track is going to be digitally compressed down to MP3 (usually 128 kbps) for download and listening?

One issue under discussion here is where the audio production industry should accept these listening trends or whether they should act to deter users from using them. Either way, the current situation appears to be unsustainable; either listeners are educated in the importance of quality, or production and distribution methods are updated to fit in with consumer listening trends. The latter is not so unheard of, particularly with engineers finalising tracks specifically for radio, or vinyl or whatever the final reproduction outlet might be.

It could be hoped that the 'quantity' fascination of current listening trends will reach saturation point, in that consumers will not need to store more than, say, 50,000 tracks at once on a particular device! Given continually evolving data storage technologies, it could be hoped that listeners will eventually look to improve the audio 'quality' of their stored data given that they have already achieved the desired 'quantity'. If this is the case, and data storage capacities continue to increase, then eventually consumers will be able to have the best of both worlds and store a large quantity of audio with appreciable quality resolution.

Along with the modern listening trends and the availability of advanced production software and hardware, consumers are more and more becoming home 'producers' and 'mastering engineers', manipulating their audio to their own preferences. Consumers therefore require greater education and appreciation for quality and better knowledge of best practices for reproducing audio. If consumers continue to embrace low-fidelity reproduction without consideration for audio quality, then it could become the case that listeners' musical preferences could be more based on the reproduction system used as apposed to the quality and integrity of the music being listened to.

3 Advanced audio processing with high-level control

Historically consumers haven't had access to, or even been trusted with, advanced audio processing systems. We all know that television sets could benefit from some type of audio dynamic range processor to remove the blaring adverts' unsavoury loudness or to normalise output levels between channels. While modern TVs might include audio processing circuitry, few (if any) actually allow a consumer to configure these levels to their desired taste; it is assumed that a home 'consumer' doesn't understand the meaning and value of dynamic 'compression', 'expansion' or 'limiting' sufficiently to benefit. But shouldn't we actually transfer to the consumers a little more knowledge so that we can develop a practical business case for these types of devices and functionality? Television adverts are generally heavily compressed to create impact – but for many viewers, it usually ends up just being too loud and disruptive! We could easily develop systems to limit or expand these signals, but it would be up to the user to manage and configure the system to their taste. This applies similarly to multi-disc CD players when used in random or shuffle mode. Different CDs have different output levels, but if a listener can't be bothered to physically change the disc between songs, then they're also unlikely to be wiling to alter the volume each time.

A further example is the development of equalisation systems for home hi-fi. Early home hi-fi EQ systems allow graphical adjustment of the frequency range, over maybe five or seven frequency bands. But it was quickly decided that a consumer doesn't significantly understand the meaning and application of equalisation enough to benefit. In some cases this was replaced with a simple 'loudness' button which just switched in a single EQ pattern (usually based on the Fletcher-Munson loudness characteristic). More recently home hi-fi systems allow us to select between EQ shapes, usually categorised by music genre or a particular venue type. However, how often do you think a listener

would select the 'classical' EQ setting to listen to heavy rock music, and, even, how often would you use the 'classical' setting for listening to classical music – is it not classical enough for you? This reflects a definite dumming down of society, as it is assumed that consumers of audio don't sufficiently understand electronic processing devices and their associated terminology to benefit from advanced functionality.

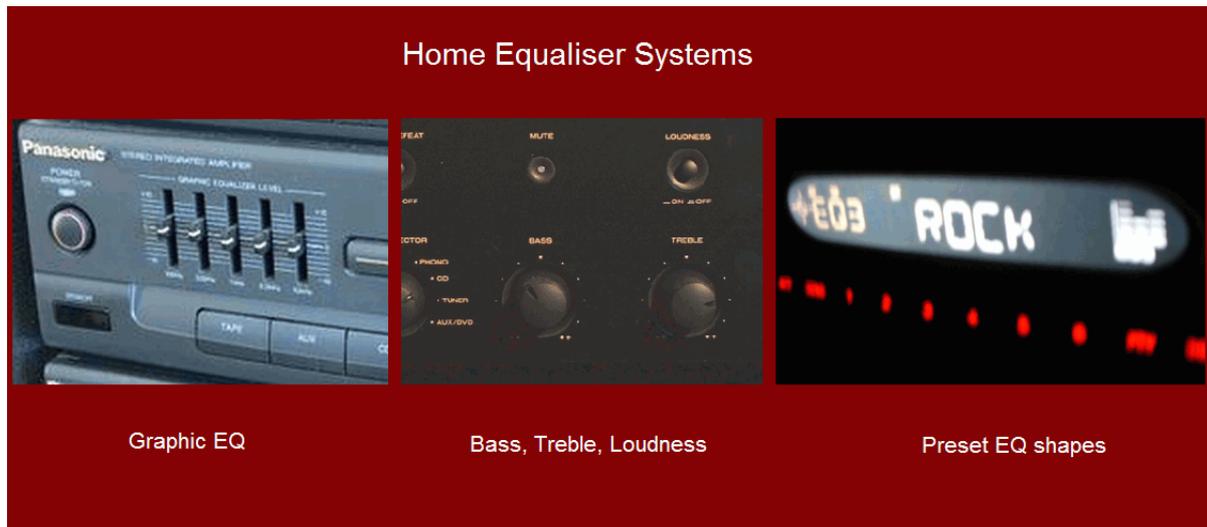


Figure 1. Equaliser systems of home audio devices.

It is suggested that a more appreciable system would be to allow consumers to manipulate their home hi-fi audio quality by increasing the 'warmth' or reducing 'sibilance' in a track, or attenuating problem 'standing waves' for a given entertainment system in a particular room – if only the consumer understood what these terms meant! Furthermore, 'warmth' and 'brightness', for example, can be manipulated by many means other than simple EQ adjustment; manipulation of musical harmonic content is also a contributing factor for enhancing such audio properties (White, 2003, p77-92).

Development of intelligent audio processing devices with high-level control is becoming more possible with advances in electronic hardware and software design. Such devices which could become commercially valuable might include, for example:

- Intelligent closed-loop EQ systems for correction of room acoustics
- Intelligent output level control for CD duke box and random MP3 playback
- Intelligent random song selection to analyse and select songs of a particular genre
- Intelligent automated DJ (beat and pitch) mixing
- Advanced correction/optimisation for automotive vehicle cabins

Systems such as these, however, require a little interaction from the user and at least an appreciation for the improvement of reproduced audio. Listeners therefore need educating towards the values of quality audio reproduction before these types of processing systems can become commercially viable. This opinion, in turn, raises a further problem: what exactly is 'warmth', what is 'sibilance', what makes something sound 'boxy' or 'present'? Perhaps a more unified scientific understanding of these terms and parameters is required before we can educate the listener and hence take advantage of such high-level audio processing systems.

4 Moving towards universal audio terminologies

Given advances in audio technology, consumer listening trends and the availability of advanced processing devices for amateur audio enthusiasts, it can be emphasised that knowledge transfer between the audio production industry and the audio consumer should be improved. This transfer of knowledge and education of listeners can only be accomplished with a universal language of audio terminologies and descriptors. In turn, an improved appreciation of audio quality will generate the commercial viability for development of high-level intelligent audio devices for home use.

In order to develop a universal language of audio terminologies, the issue of defining scientific characteristics for common subjective descriptors should be considered.

4.1 Categorising audio terminologies

Rumsey (2005) identifies the need for clear definition of subjective attributes of audio. Rumsey also indicates that descriptors of audio quality can be divided into three different categories; technical qualities, spatial qualities and timbral qualities.

Technical quality adjectives such as 'compressed', 'noisy' or 'distorted' are generally quite easy to justify with reference to scientific parameters. Simple waveform analysis can be used to determine how much compression, noise or distortion is present in an audio track. Similarly, signal processing algorithms can be used to quantify harmonic content and whether a sound is at a desired pitch or not. Other technical observations of quality that can be quantified scientifically include loudness, timing and waveform envelope shapes.

Spatial qualities usually refer to the positioning of instruments within a final stereo mix. Obviously panning can be used to position instruments left and right, but further techniques can be used with reverb and other time based modulation to give a sense of depth and room size. This gives rise to descriptors such as 'wide', 'deep' and 'up-front' which can all be quantified with scientific terms relating to stereo position and reverb characteristics if necessary.

Terminologies referring to timbral qualities often utilise imagery and metaphors to relate musical sounds to emotive feelings and visualisations. As such, timbral adjectives are rarely correlated with specific scientific parameters - though recent investigations show that this practice would certainly be beneficial (Rumsey, 2005), (Atsushi & Martens, 2005) and is indeed possible (Disley et al, 2006), (Johnson & Gounaropoulos, 2006).

4.2 Describing timbre

The timbre of musical sound is defined by the American National Standards Institute as

“that attribute of auditory sensation in terms of which a listener can judge two sounds simultaneously presented and having the same loudness and pitch as being dissimilar”

(ANSI, 1960).

Pitch and loudness are obviously quantifiable by scientific terms, and indeed so can many other parameters relating to the quality and make-up of musical sound. So, essentially, timbre relates to all the other descriptors of musical sound that are not quantified by pitch or loudness. In general, timbral adjectives revolve around metaphoric imagery related to emotions and visualisations such as 'smooth', 'sweet', 'warm' or 'bright'. But, each of these metaphoric descriptors can be broken down into a combination of scientific parameters. Katz (2002, p43) defines a unique chart relating subjective timbral terms with their relative energy presence and deficiency across the audio spectrum. Subjective timbral adjectives are therefore used to describe many specific properties and characteristics of sound in a single term, but as this has propagated through time, the actual scientific characteristics being described have been left behind. Often such metaphoric adjectives are used because it allows us to avoid the need to specifically quantify the technical attributes being referenced. This highlights an interesting paradox in that something intrinsically metaphoric is being used to define something definite and scientific.

As an example, the timbral description of 'warmth' is one of the most widely used timbral descriptions, aimed usually at describing qualities of high-end analogue recording and reproduction systems. (Use of the metaphoric term 'warmth' probably originates from the implementation of thermionic valve devices which glow and give off heat in vintage analogue amplifiers). The term warmth is referred to so frequently that it has been possible over time to identify specific scientific properties which correlate with those properties of a warm sound, for example

- Energy in the lower mid-range of the audio spectrum (200 – 500 Hz)
- Depletion in the upper mid-range of the audio spectrum (2.5 – 6 kHz)
- Subtle non-linear compression as in analogue audio recording devices (analogue tape)
- Subtle harmonic distortion as in analogue (thermionic valve) audio amplification devices

(Hood, 1997,), (Hood, 1999, p4-12), (Katz, 2002, p43).

It should similarly be possible to define technical characteristics relating to other timbral adjectives.

4.3 Visualising musical sounds

Metaphoric timbral adjectives are often associated with the human senses (other than hearing). For example 'warmth' relates to touch and 'sweet' relates to taste. In many cases, adjective imagery is used to relate how something sounds to a comparable visual representation. As in Katz's relationship between subjective terms and spectrum make-up, it is generally agreed that 'bright' sounds have greater energy in the high-midrange and treble regions (2-20 kHz) and/or depletion in the bass and lower midrange (60 – 2000 Hz). A particularly bright percussive instrument is the cymbal, but many timbral descriptions used for instruments are often contradictory and confusing, for example

"Byzance Medium Thin Crash - Significant, washy and fairly dark sound with a full frequency spectrum. Voluminous attack with moderate sustain"

(<http://www.meincymbals.com>, 2006)

Here a cymbal which is obviously 'bright' by nature is described as 'fairly dark' as well as being both 'thin' and 'with a full frequency spectrum'. It must be appreciated that these types of descriptions are often used as marketing tools, and indeed descriptions such as these do give a consumer the opportunity to make a choice based on a subjective definition, but in many cases these types of descriptions only add confusion and hence become meaningless - as well as being scientifically incorrect. Defining scientific parameters for universal timbre adjectives would allow the confusion and contradiction to be avoided in such instances.

Another common visualisation is to use colour to describe the timbre of a particular instrument. For example, the sound of a trumpet might be described as 'scarlet', or the cello as 'rich brown' (Howard & Angus, 2006, p 217), (Scholes, 1970). The use of colour as a descriptor for music might seem a little abstract at first, particularly as people can easily visualise colours in a very subjective way; some people like the colour brown, others do not. Does that mean if you don't like the colour brown then you probably won't like the sound of the cello too?! It is possible, however, to dissect this metaphor to relate to the scientific properties being referred to. In the example of using 'scarlet' as an adjective for musical sound, we can break this down into descriptive terms for scarlet which describes a warm, smooth and bold colour, and these descriptive terms equally describe the sound of a trumpet.

The use of metaphors and adjectives to describe musical sounds therefore generally defines a method for grouping terms and scientific properties in order to describe a sound in a more concise and colloquial manner. This enables musical sounds to be discussed without the need to individually specify each and every scientific component of the sound. Again with reference to the 'scarlet trumpet' example, we can see that the term 'warm' defines the spectral makeup and harmonic content of the sound, 'smooth' defines certain properties associated with the attack, sustain and envelope shape of the instrument and 'bold' can relate to the dynamic range and perceived loudness of the sound – see Figure 2.

The use of timbral adjectives can therefore be a very powerful tool in describing and referencing music and audio. Unfortunately, the subjective nature of the current use of these terminologies allows for regular contradiction and confusion.

5 Conclusions

Modern consumer listening trends on the whole appear to revolve around the desire to retain audio in a media format which allows convenient and functional access to a large database of music. The use of MP3 compressed audio facilities this consumer demand, though generally at the expense of the reproduced audio quality.

The use of MP3 compressed audio plays against the production of high quality and artistically mastered music records. There is also a danger that listeners' musical preferences could be more based on the reproduction method employed as apposed to the quality and integrity of the music being listened to. There is therefore a current need to educate listeners in the values of quality audio which in turn can provide a business case for more intelligent home audio processing hardware and software systems.

Educating the audio listening public is not a simple task, particularly given that many audio terminologies are subjective and metaphoric, and rarely defined scientifically. Before knowledge can be transferred to consumers, the industry professions must first develop a uniform language of terminologies.

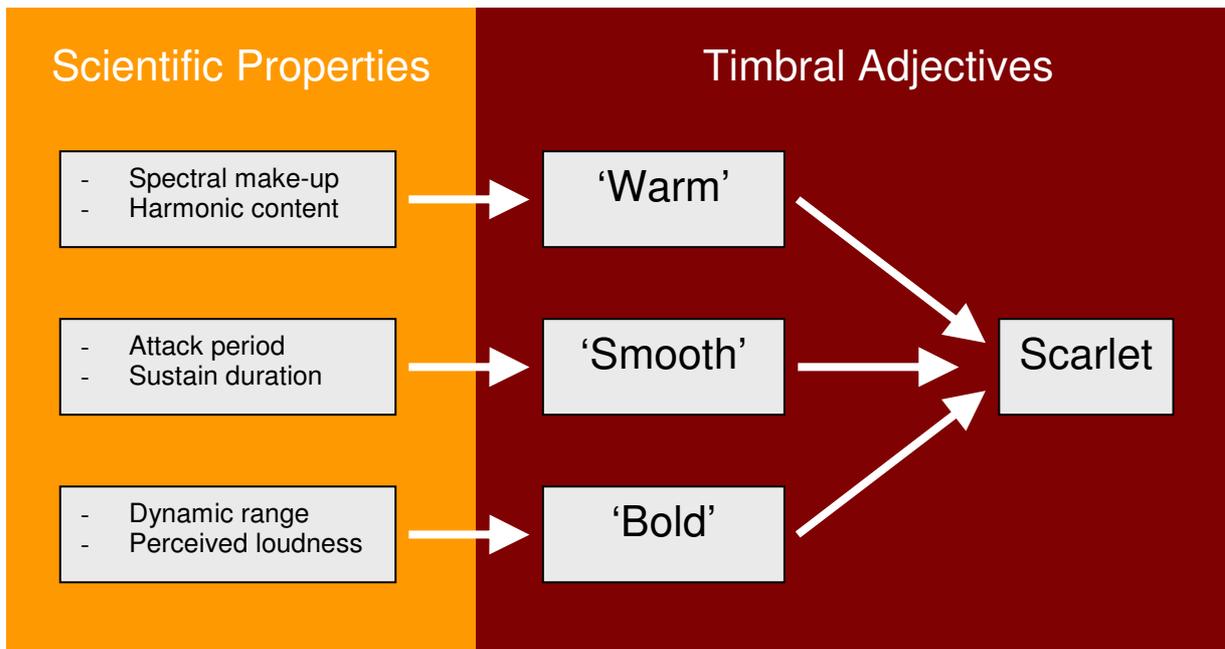


Figure 2. Example relationship between scientific properties and timbral adjectives of musical sound.

It has been shown here and previously that it is possible to define particular scientific qualities of audio which can be referenced using simple metaphoric terms or timbral adjectives. Groups of adjectives can further be grouped to describe complex sounds as higher-level adjectives or metaphoric visualisation, such as colour. More detailed psychoacoustic analysis correlating timbral adjectives with specific scientific properties is required to achieve a universal understanding of music and audio descriptors.

In conclusion, to allow the development of advanced devices to enhance audio quality within current consumer listening trends, we should first attempt to develop more universal definitions of audio terminologies. Correlating scientific parameters with descriptive terms should allow improved education and knowledge transfer to the consumer and hence provide a business case for the development of advanced high-level consumer processing products.

6 References and Sources

- Atsushi, M. & Martens, W. L. (2005).** Timbre of Nonlinear Distortion Effects: Perceptual Attributes Beyond Sharpness? Proceedings of the Conference of Interdisciplinary Musicology, Montreal, Canada.
- Disley, A. C., Howard, D. M. & Hunt, A. D. (2006).** Spectral correlation of timbral adjectives used by musicians, *The Journal of the Acoustical Society of America*, 119(5), p3333.
- Hood, J. L. (1997).** Valve and Transistor Amplifiers, Oxford, Newnes Publishing.
- Hood, J. L. (1999).** Audio Electronics, Oxford, Newnes Publishing.
- Howard, D. M. & Angus, J. (2006).** Acoustics and Psychoacoustics, 3rd Edition, Oxford, Focal Press.
- Howard, D. M. & Tyrrell, A. M. (1997)** Psychoacoustically informed spectrography and timbre, *Organised Sound* (1997), 2: 65-76 Cambridge University Press.
- Johnson, C. G. & Gounaropoulos, A. (2006).** Timbre interfaces using adjectives and adverbs, Proceedings of the 2006 International Conference on New Interfaces for Musical Expression, Paris, France.

- Katz, B. (2002).** Mastering Audio, Focal Press.
- Owen, O. (Editor) (2006).** Tech Angst, Future Music, Future Publishing Ltd, p13.
- Rumsey, F. and McCormick, T. (2006).** Sound and Recording, 5th Edition, Oxford, Focal Press.
- Rumsey, F. (2005).** Psychoacoustics of Sound Quality, Proceedings of the Art of Record Production Conference, London, 2005.
- Scholes, P. A. (1970)** The Oxford Companion to Music, London, Oxford University Press.
- White, P. (2002).** Creative Recording – part one, effects and processors, 2nd Edition, London, Sanctuary.

World Wide Web References

- <http://www.meinlcymbals.com>. (2006)** http://www.meinlcymbals.com/cymbal_series/cymbals_byzance_regular.html, accessed September 2006.