



Deliverable D5.1

Hierarchical ontology of timbral semantic descriptors

Grant agreement nr	688382
Project full title	Audio Commons: An Ecosystem for Creative Reuse of Audio Content
Project acronym	AudioCommons
Project duration	36 Months (February 2016 - January 2019)
Work package	WP5
Due date	31 August 2016 (M7)
Submission date	31 August 2016 (M7)
Report availability	Public (X), Confidential ()
Task leader	Surrey
Authors	Andy Pearce, Tim Brookes, Russell Mason
Document status	Final





Table of contents

Executive Summary	3
Background	4
1 Introduction	5
1.1 Main objectives and goals	5
1.2 Methodology	5
1.3 Terminology	5
2 Attribute identification	7
2.1 Literature attributes	7
2.2 Attribute reduction	8
2.2.1 Automated redundancy reduction	9
2.2.2 Removal of non-timbral terms	9
2.2.3 Group discussion	10
2.3 Hierarchical grouping	10
2.4 Inclusion of D2.1	10
3 Search term analysis	12
3.1 Search term frequency	12
3.2 Manual filtering	13
3.3 Hierarchical frequency summing	14
4 Conclusion	17
5 References	18
APPENDIX 1: Dictionary of timbral terms and hierarchical structure	20





Executive Summary

This deliverable identifies the timbral attributes that have potential to add value to online audio as automatically-generated tags.

The task was accomplished in two sections. Firstly, a dictionary of timbral attributes and terms used to perceptually describe audio was compiled from the relevant academic literature, and this dictionary was then structured into a hierarchy. Secondly, the frequency-of-use of each dictionary term, in online searches for audio content, was established in order to give an indication of each term's potential value; every search within a one month time-frame on Freesound was considered.

This method identified that the term *deep* was the most frequently searched for, followed by *dark*, *soft*, and *electronic*. To allow for the fact that a timbral attribute might be labelled using multiple terms, the frequency-of-use for each term was summed into the hierarchical ordering. This showed that the *depth* attribute (as labelled with the term *deep*) was only the third most searched for attribute, with the *hardness* attribute (labelled by *hard*, *pillowy*, and *soft*) being the most frequently searched for.

The findings presented in this document can guide future research in WP5, which aims to develop automatic tagging tools for timbral attributes, and can also feed into the Audio Commons ontology.





Background

This deliverable is part of the “semantic annotation of non-musical sound properties” work package (WP5). This work package aims to enhance the usefulness of existing content and to facilitate more creative uses by: (i) developing better tools for manually annotating sound effects and soundscapes; and (ii) developing a system to automatically add timbral metadata, such that content can be searched by perceptual sound quality (e.g., piercing, crunchy, rich, etc.).

Towards this aim, this deliverable identifies the relevant timbral attribute terms that are used to perceptually describe an audio signal (based on the existing literature), structures these terms into a suitable *hierarchy* (developed by the authors), and establishes the frequency-of-use for each term (based on end-user searches made in *Freesound*) to give an indication of the potential value of the attribute as an automatically-generated tag.

This document also maps the output of deliverable D2.1: “Requirements Report and Use Cases” onto the timbral hierarchy introduced above, from the survey data regarding what users expected of systems developed within AudioCommons. The survey contained questions pertaining to the perceptual attributes that users may want to be tagged/searchable.





1 Introduction

1.1 Main objectives and goals

In order to improve on a user's experience with Creative Commons licensed audio, users may want to be able to narrow down searches by specifying the timbral attributes of the desired sound. The goal of this deliverable is to identify the timbral attributes that are currently being searched for on *Freesound*, a website which hosts Creative Commons licensed audio samples. Identifying which attributes are being searched for, and their frequency-of-use, will indicate their potential usefulness as tags, and thus give an indication of which attributes should be studied further.

1.2 Methodology

The *Freesound* search history is too large to manually inspect and parse all entries (in order to decide which term within the search, if any, is a timbral attribute). Current natural language processing algorithms are not well equipped to deal with this type of data and are unable to determine which terms would be timbral attributes due to the lack of a suitable lexicon. Therefore, in order to automatically extract the relevant timbral attributes from the search term data, a dictionary was first developed that contains a wide range of the timbral terms that people typically use to describe audio.

The development of this dictionary, discussed in [Section 2](#), was accomplished by a literature review of research into timbral description of audio across many different fields, collating all of the timbral attributes and descriptors used. Redundancy was removed with natural language processing algorithms, and a manual approach was then employed for sifting and structuring the terms, creating a dictionary of timbral attributes that are arranged in a hierarchy.

Natural language processing algorithms were then used to compare the *Freesound* search history data against this dictionary, as discussed in [Section 3](#). The relevant timbral searches were identified and the frequency-of-use calculated to give an estimate of the relative usefulness of each term within the dictionary. Hierarchical summation of the frequencies was also used to give an indication of the frequency-of-use for each attribute.

1.3 Terminology

AudioCommons: reference to the EC H2020 funded project AudioCommons, with grant agreement nr 688382.

Audio Commons Initiative: reference to the AudioCommons project core ideas beyond the lifetime and specific scope of the funded project. The term "Audio Commons Initiative" is used to imply i) our will to continue supporting the Audio Commons Ecosystem and its ideas after the lifetime of the funded project, and ii) our will to engage new stakeholders which are not officially part of the project consortium.

Audio Commons: generic reference to the Audio Commons core ideas, without distinguishing between the concept of the initiative and the actual funded project.

Audio Commons Ecosystem (ACE): set of interconnected tools, technologies, content, users and other actors involved in publishing and consuming Audio Commons content.





Audio Commons content (ACC): audio content released under Creative Commons licenses and enhanced with meaningful contextual information (e.g., annotations, license information) that enables its publication in the ACE.

Content creator: individual users, industries or other actors that create audio content and publish in the ACE through content providers.

Content provider: services that expose content created by content creators to the ACE.

Content user: individual users, industries or other actors that use the content exposed by content providers and created by content creators in their creative workflows.

Tool developer: individual users, industries or other actors that develop tools for consuming (and also potentially publishing) Audio Commons content.

Embeddable tools: tools for consuming Audio Commons content that can be embedded in existing production workflows of creative industries.



2 Attribute identification

Any sound has a number of acoustical attributes that make up its acoustical quality (e.g. frequency, sound pressure level, harmonic structure). When that sound is perceived by a listener, its perceptual quality comprises a number of perceptual attributes (e.g. pitch, loudness, richness). Timbre is a broad term traditionally covering those perceptual attributes that are not pitch or loudness. It is now generally agreed that timbre should also exclude attributes relating specifically to spatial quality (e.g. location, width, envelopment). Often, and in the context of this research and report, timbre also excludes musicological descriptors and higher-level cognitive (e.g. hedonic and emotional) attributes.

There have been a number of studies identifying the timbral attributes that can be used to describe audio. However, the attributes identified in each of these studies are usually only relevant for a particular use case. For example, Gabrielsson and Sjögren [1979] conducted a series of experiments to identify the attributes that can be used to describe the main perceived differences between loudspeakers, whereas Pearce et al. [2016] undertook this for microphones. Many studies which intend to identify perceptual attributes opt for a similar approach: discerning the attributes which are the most prominent/important for distinguishing between particular stimulus types. This is often conducted with a form of indirect elicitation experiment and analysis, collecting preference/similarity/distance measurements between stimuli and performing a multidimensional scaling (MDS) or principal component analysis (PCA) in order to identify orthogonal components. These orthogonal components can then be labeled with an appropriate perceptual attribute using a direct elicitation experiment.

This section collates the perceptual attributes identified in previous studies, describes the methods used to remove unsuitable descriptors, and reports on group discussions conducted to categorise and structure the attributes into a suitable hierarchy.

2.1 Literature attributes

As discussed, the perceptual attributes that are identified in any study of sonic descriptors will often pertain primarily to a particular stimulus type. In order to collate the full range of timbral attributes, a wide range of studies was considered. Some of these focused on particular types of sounds, such as environmental sounds, speech, musical instruments, concert halls, or sound recording and reproduction systems, though some covered multiple types.

In total, 1140 attributes were identified. The number of attributes from each paper is shown in Table 1, along with the general topic of the paper.

Table 1: Number of attributes from each source		
Source	Number of attributes	Topic
Bagousse et al. [2010]	28	Spatial/Timbral
Cano [2006]	10	Sound description
Choisel [2007]	8	Multichannel reproduction
Davies et al. [2013]	49	Environmental/Soundscape





Disley et al. [2006]	17	Musical instrument
Gabrielsson et al. [1979]	13	Loudspeakers
Handel [1995]	16	Timbre
Hermes [2014]	105	Mix quality
Jensen [N.D.]	8	Timbre
Koivuniemi and Zacharov [2001]	12	Spatial sound
Levandier et al. [2008]	3	Loudspeakers
Lokki et al. [2011]	10	Concert halls
Lorho [2005]	16	Headphones
Mattila [2001a, 2001b, 2002, 2003] (Summarised in Bech and Zacharov [2006])	27	Speech
Michaud et al. [2015]	4	Loudspeakers
Pearce et al. [2016]	40	Microphones
Pedersen [2008]	647	Reproduced sound
Pedersen [2015]	42	Reproduced sound
Staffeldt [1974]	38	Loudspeakers
Wrzeciono and Marasek [2010]	11	Musical instrument
Zacharov and Pedersen [2015]	34	Reproduced sound
Zwicker and Fastl [2007]	2	Timbre perception

It can be seen from Table 1 that the paper by Pedersen [2008] reports on a particularly high number of perceptual attributes. This was due to it incorporating an appendix of the raw terms used prior to any attribute grouping experiments.

2.2 Attribute reduction

Within the 1140 attributes identified, there is likely to be a degree of redundancy, with multiple papers identifying the same attribute or variations of it (e.g. *brightness*, *bright*, and *brighter*). Additionally, there may be terms that describe aspects of sound that are not timbral (e.g. those to do with loudness, pitch, or spatial or musicological characteristics).





To remove this redundancy and the non-timbral terms, an automated removal of redundancy was performed, followed by a subjective removal of non-timbral attributes. A final step converted each attribute to its adjectival form.

2.2.1 Automated redundancy reduction

Redundancy in the data was automatically reduced by four methods: tokenizing, direct comparisons, lemmatization, and stemming. Firstly, the 1140 attributes were *tokenized*, a natural language processing expression meaning that each timbral descriptor was broken down into its component words, hereafter referred to as *terms*. This was conducted using the WordNet package's tokenizer in Python 3.5 (<http://www.nltk.org/api/nltk.tokenize.html>) [Princeton, 2010]. The tokenizing algorithm does not split compound adjectives.

Secondly, automated direct comparisons were made between all terms within the list, discarding any duplicates.

Thirdly, the remaining terms were *lemmatized* using Wordnet. Lemmatization is a lexicographical transformation of a word to a common form. For example, the words *sound's*, *sounds*, and *sound* would all be lemmatized to the word *sound*. Lemmatization was followed by the removal of any duplicate lemmatized terms.

Finally, terms were *stemmed*. Stemming is a more crude form of lemmatization, removing the suffixes of words to leave the base form of a word. For example, *brightness*, *brighter*, and *brightest* would all be stemmed to *bright*. However, stemming can result in a word that is spelt incorrectly or has no meaning. For example, "precise" will be stemmed to "precis" [sic]. To prevent this, terms were stemmed and duplicates of the stemmed terms were removed, but an un-stemmed version of the term was retained for the dictionary.

Using these four methods, the 1140 attributes were reduced to 683 terms.

2.2.2 Removal of non-timbral terms

Following the automated removal of redundant terms, a manual approach was taken to remove non-timbral terms. This was completed with two tasks. Firstly, each of the three authors evaluated each of the 683 terms against this criterion:

Does the term (or the adjectival form of it) describe a timbral characteristic of sound?

Secondly, for each retained term the authors replaced the term with its adjectival form. For example, *brightness* was replaced by *bright*.

Both tasks were completed by all three authors independently. All three sets of results were then compared. Any terms deemed by all three experimenters to fail to meet the retention criterion were rejected. This left 224 terms that two or more of the authors agreed to retain, and 131 terms that only one author suggested retaining.





2.2.3 Group discussion

A group discussion was held with the three authors to consider further each of the 349 retained terms. During this discussion, more detailed criteria for removing terms were developed and applied:

A term will be removed where it:

1. relates to loudness, pitch, or a spatial attribute;
2. refers to a musicological attribute;
3. is a hedonic or emotional term;
4. has meaning only with reference to another non-specified sound (e.g. natural, realistic); or
5. can only refer to the relationship between a sequence of sounds.

Where at least two authors agreed that a term failed one of the removal criteria, it was removed. Under these criteria, the 335 terms were reduced to 296. These 296 attributes form the dictionary that was then used for identifying timbral descriptors within the Freesound search history data. The terms within this dictionary are listed in [APPENDIX 1](#).

2.3 Hierarchical grouping

The remaining 296 terms were then grouped and organised manually during a group discussion between the three authors. As in the work of Pearce et al. [2016] and Pedersen and Zacharov [2015], the terms were structured into a hierarchical format. This structuring had two benefits: (i) it allows for terms that relate to the same timbral attribute to be grouped together; and (ii) it will allow for the frequency-of-use of each of the terms within the Freesound search data to be summed hierarchically. For example, the number of searches for *bright*, *dull*, or *dark* might each be fairly low, but if these terms are all part of a common container attribute, labelled *brightness*, across which the frequencies can be summed, this container attribute might be found to have significant value.

The hierarchy was structured as shown in Figure 1, with the location of each term within the hierarchy shown in [APPENDIX 1](#). A high-resolution version of the file can be downloaded from <http://www.audiocommons.org/assets/files/AC-WP5-SURREY-D5.1-Fig1.png>.

2.4 Inclusion of D2.1

As part of the AudioCommons project, deliverable D2.1 discussed the results of a survey intended to learn what users would expect of a system developed within AudioCommons. This survey included questions pertaining to the perceptual attributes that users might expect to search for.

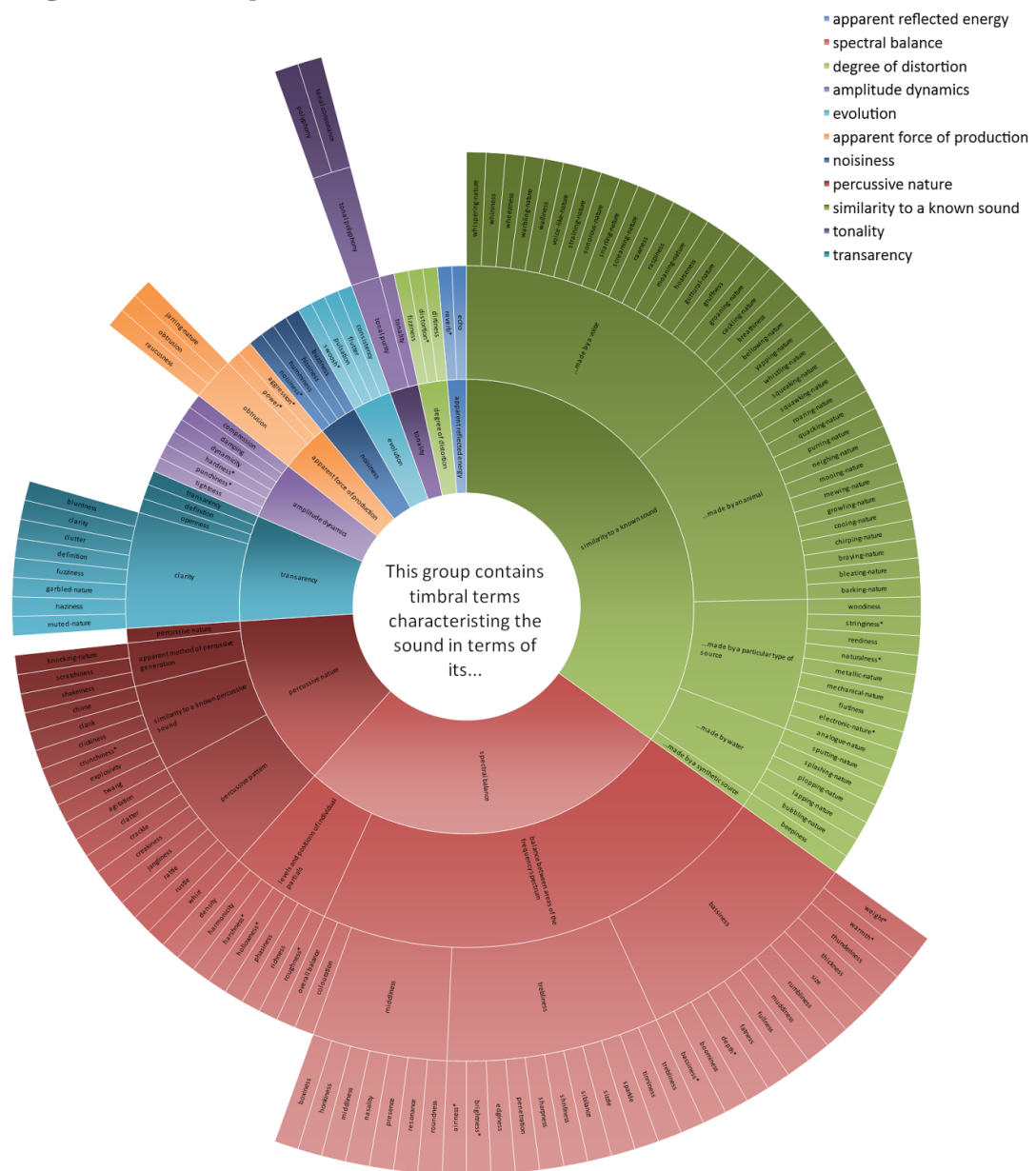
The 59 perceptual attributes from D2.1 were included within the group discussion. The removal criteria specified in [Section 2.2.3](#), were applied and, of the 59 terms, 31 were identified as non-timbral and removed. The remaining 28 were placed within the hierarchy. The hierarchical groups that contain one of the terms from D2.1 are denoted by an asterisk in Figure 1.

The explicit terms from D2.1 (after being converted to the closest adjectival form) are denoted with an asterisk in [APPENDIX 1](#).





Figure 1 - Sunburst plot of the hierarchical structure of the timbral attributes





3 Search term analysis

[Section 2](#) reported on the generation of a hierarchical dictionary of 249 terms, in adjective form, which can be used to describe the timbral attributes of audio signals. In order to determine which of these terms has potential to add value as an automatically added tag, the search history from *Freesound* (a website which hosts Creative Commons licensed audio samples) was compared against the dictionary of terms.

Freesound retains the most recent month's search term history, and this provided a database of 8,154,586 searches (equivalent to 263,000 per day or 183 per minute), 879,976 of which were unique.

3.1 Search term frequency

For each unique search, the entire search was tokenized using the WordNet tokenizer to split the searches into individual words. Each word was then compared against the dictionary of terms for an identical match. If a match was found, the frequency (occurrences per month) of the entire unique search was added to the total for the matching term.

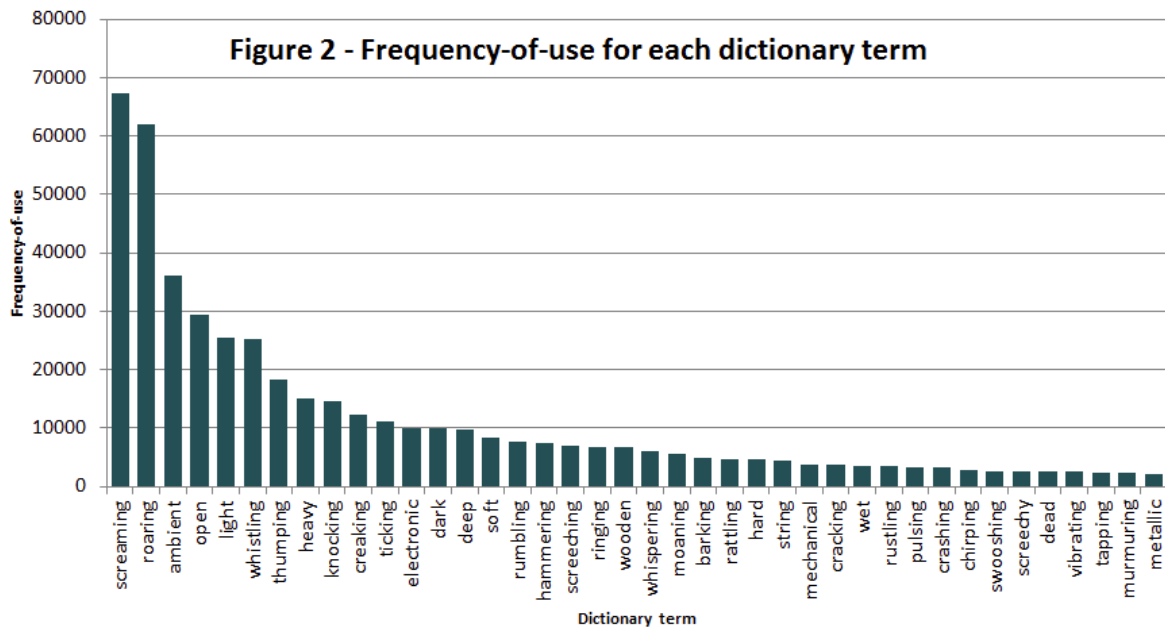
If no match was found, the similarity between the word and each dictionary term was calculated using the WordNet Wu Palmer metric [Wu and Palmer, 1994]: a measure of word similarity, ranging from 1.0 (perfect match) to 0.0 (no similarity). This metric is based on the distance between the two words within the WordNet taxonomy.

When conducting a Wu Palmer analysis in WordNet, the words under analysis must be converted to *SynSets*: the specific definition of the word being assessed. For example, the word *light* could be interpreted as either: (i) the absence of darkness, similar to bright; or (ii) the antonym of heavy, relating to the apparent weight of an object. The Wu Palmer similarity can therefore be at least partly dependent on the specific SynSet that is used. There is no automated method for determining the appropriate SynSet. The most common method, as employed in this analysis, is to use the SynSet that represents the most common usage of the word.

A threshold for the Wu Palmer similarity was set at 0.95, ensuring that only very similar terms would be grouped. If the similarity of a search term to a term in the dictionary was over 0.95, the frequency of searches for that search term was added to the total for the matching dictionary term.

Figure 2 shows the frequency-of-use of the 40 most frequently searched terms.





From Figure 2, it can be seen that the term *screaming* is the most frequently searched term. However, closer inspection of the search phrases in which this term occurred revealed that it was commonly being used not as a timbral descriptor (e.g. “screaming electric guitar tone”) but as a verb (e.g. “woman screaming”). To remove the matching searches where the dictionary term is not being used as a timbral descriptor, the matching searches were manually filtered.

3.2 Manual filtering

There were 66,694 matches between search terms and dictionary terms. For each term in the dictionary, up to the fifty most frequent unique searches were manually inspected. This task was completed by the three authors, with the instructions:

Include a search only if the term is used unambiguously as the timbral descriptor intended by the hierarchical grouping.

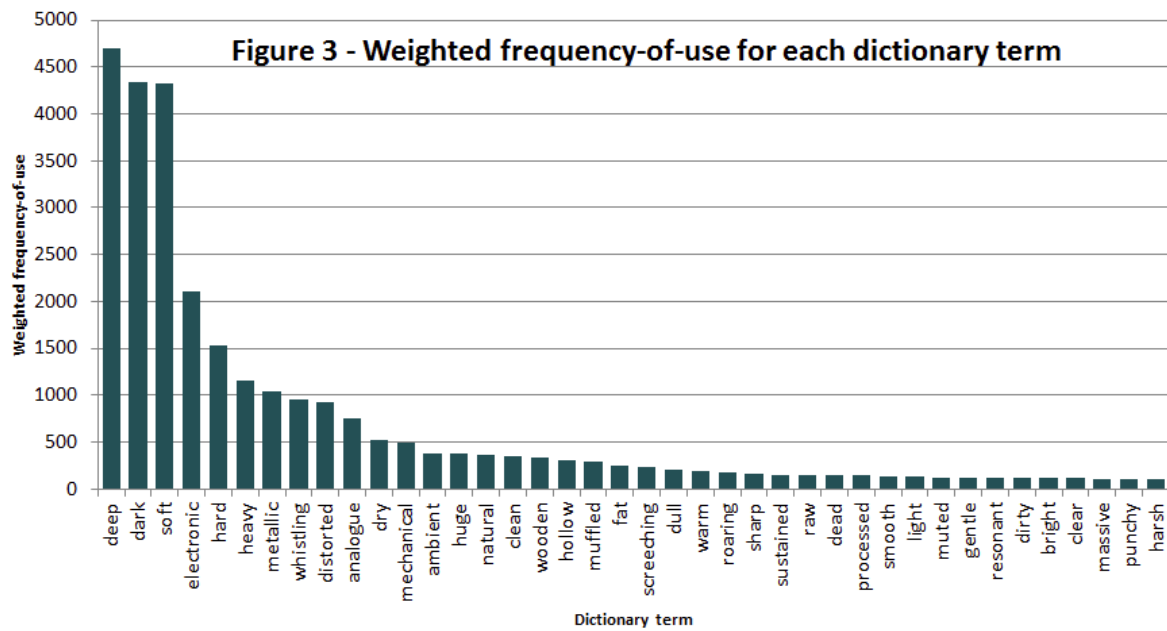
Ambiguity can result from:

- *The word being used in isolation (e.g. “screaming”);*
- *The word being potentially used as a verb (e.g. “woman screaming”);*
- *The word being potentially used as a noun (e.g. “female scream”); or*
- *The word being used as an adjective meaning something different from what our hierarchy intends (e.g. “noisy children”).*

Since it was the unique searches that were analysed, each had a different frequency-of-use. The total frequency-of-use of each retained term was found, and divided by the total frequency for all analysed occurrences of that term. This gave an estimate of the proportion of searches containing a particular term that were likely to be using the term as a timbral attribute. From this, the total frequency-of-use of the term was weighted by this proportion to give a more representative frequency-of-use for each timbral term within the dictionary.



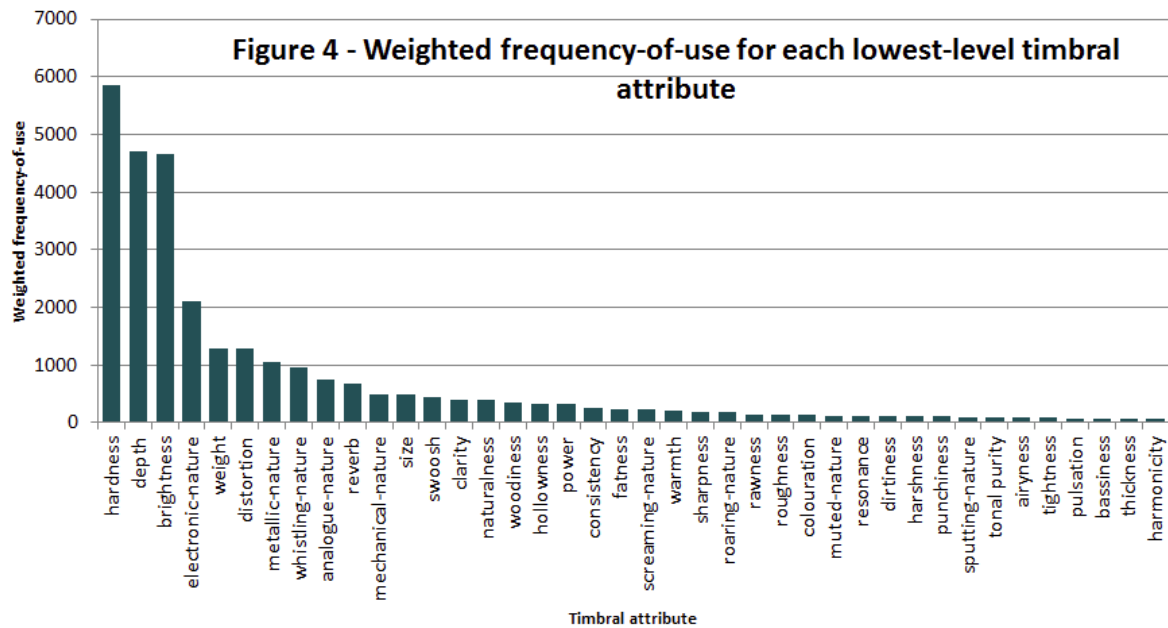
The post-filtering frequencies of the 40 most frequently searched terms is shown in Figure 3. It can be seen that the most commonly searched for term is deep, followed by dark, soft, and electronic.



3.3 Hierarchical frequency summing

In some cases, multiple terms may relate to the same perceptual attribute. For example, *brightness* is a timbral attribute identified in the hierarchy and contains the dictionary terms *bright*, *dark*, and *dull*. In order to effectively assess the frequency-of-use of each timbral attribute, the frequencies of each term for each attribute were summed.

This was accomplished by summing the post-filtered frequencies-of-use for each level within the hierarchy. Figure 4 shows the 40 most frequently searched lowest-level timbral attributes from each branch in the hierarchy.



From Figure 4, it can be seen that the *hardness* timbral attribute was the most searched for, with 5849 searches. This was closely followed by the *depth* and *brightness* attributes, each achieving 4697 and 4648 searches respectively.

The full hierarchy of timbral attributes is shown in Figure 5 with each segment's size determined by the number of searches. A high-resolution version of the file can be downloaded from <http://www.audiocommons.org/assets/files/AC-WP5-SURREY-D5.1-Fig5.png>, and an interactive version can be found at <http://iosr.uk/projects/AudioCommons/sunburst.php>.

From this figure, it can be seen that the *spectral balance* attribute group contains the most searches, with 13,123 searches. This is followed by the *similarity to a known sound* attribute group, and the *amplitude dynamics* attribute group, with 6919 and 6102 searches respectively.



4 Conclusion

In this deliverable, timbral terms and attributes from multiple previous studies have been collated, parsed, and organised to create a dictionary of timbral terms, organised into a hierarchical structure. This dictionary was then compared against the Freesound search history in order to identify the frequency-of-use for each perceptual term. The term *deep* was the most frequently used, followed by *dark*, *soft*, and *electronic*.

To allow for the fact that multiple terms may relate to the same perceptual attribute, the frequency-of-use for each term was summed according to the hierarchical attribute structure. This showed that the *depth* attribute was only the third most searched for, with the *hardness* attribute being the most frequently searched for.

The timbral attributes used most frequently in *Freesound* searches are likely to make useful tags. The findings presented in this document can therefore guide future research in WP5, which aims to develop automatic tagging tools for timbral attributes likely to add value to uploaded audio.





5 References

Bagousse, S., Paquier, M., and Colomes, C., 2010: 'Families of sound attributes for assessment of spatial audio', in 129th Convention of the Audio Engineering Society (San Francisco, USA).

Bech, S. and Zacharov, N, 2006: *Perceptual Audio Evaluation: Theory, Method and Application* (West Sussex, England: Wiley).

Cano P., 2006: 'Content-based audio search: from fingerprinting to semantic audio retrieval', PhD Thesis, Pompeu Fabra University.

Choisel, S., 2007 : 'Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference', *The Journal of the Acoustical Society of America*, Vol. 121(1), pp. 388–400.

Davies, W., Adams, M., Bruce, N., Cain, R., Carlyle, A., Cusack, P., Hall, D., Hume, K., Irwin, An., Jennings, P., Marselle, M., Place, C., Poxon, J., 2013: 'Perception of soundscapes: An interdisciplinary approach', in *Applied Acoustics*, No. 74, pp. 224–231.

Disley, A., Howard, D., and Hunt, A., 2006: 'Timbral description of musical instruments', in *Proceedings of 9th International Conference on Music Perception and Cognition* (Bologna, Italy), pp. 61–68.

Gabrielsson, A. and Sjögren, H., 1979: 'Perceived sound quality of sound-reproducing systems', *The Journal of the Acoustical Society of America*, Vol. 65(4), pp. 1019–1033.

Handel, S., 1995: 'Chapter 12: Timbre perception and auditory object identification' in Moore, B. ed. *'Hearing'* (Academic press), pp. 425–461.

Hermes, K., 2014: 'Towards measuring and modelling the perceived quality of music mixes', PhD Confirmation Report, University of Surrey, Institute of Sound Recording.

Jensen, K., N.D.: 'The timbre model', unpublished, University of Copenhagen, Music informatics laboratory, available: <http://www.musanim.com/pdf/JensenTimbreModel.pdf> .

Koivuniemi, K., and Zacharov, N., 2001: 'Unraveling the perception of spatial sound reproduction: Language development, verbal protocol analysis and listener training', in 111th convention of the Audio Engineering Society, (New York, USA).

Lokki, T., Pätynen, J., A., Kuusinen, Vertanen, H., and Tervo, S., 2011: 'Concert hall acoustics assessment with individually elicited attributes', *The Journal of the Acoustical Society of America*, Vol. 130, No. 2, pp. 835–849.

Lorho, G., 2005: 'Evaluation of spatial enhancement systems for stereo headphone reproduction by preference and attribute rating', in 118th convention of the Audio Engineering Society (Barcelona, Spain).

Lavandier, M., Meunier, S., and Herzog, P., 2008: 'Identification of some perceptual dimensions underlying loudspeaker dissimilarities', *The Journal of the Acoustical Society of America*, Vol. 123, No. 6, pp. 4186 – 4198.

Mattalia, V., 2001a: 'Descriptive analysis of speech quality in mobile communications: descriptive language development and external preference mapping' in 111th convention of the Audio Engineering Society, (New York, USA).

Mattila, V., 2001b: 'Perceptual analysis of speech quality in mobile communications', PhD Thesis, Tampere University of Technology, Tampere Finland.





Mattila, V., 2002: 'Descriptive analysis and ideal point modelling of speech quality in mobile communications' in 113th convention of the Audio Engineering Society, (Los Angeles, USA).

Mattila, V., 2003: 'Semantic analysis of speech quality in mobile communications: descriptive language development and mapping to acceptability' in Food quality and preference, Vol. 14, pp.441–453.

Michaud, P., Lavandier, M., Meunier, S., and Herzog, P. (2015). "Objective characterization of perceptual dimensions underlying the sound reproduction of 37 single loudspeakers in a room," Acta Acustica with Acustica, Vol. 101, pp. 603–615.

Pearce, A., Brookes, T., Mason, R., and Dewhurst, M., 2016: 'Eliciting the most prominent perceived differences between microphones' The Journal of the Acoustical Society of America, Vol. 139, No. 5, pp. 2970–2981.

Pedersen, T., 2008: 'The semantic space of sounds: Lexicon of sound-describing words': unpublished, Delta Labs, available: http://assets.madebydelta.com/docs/share/Akustik/The_Semantic_Space_of_Sounds.pdf

Pedersen, T., and Zacharov, N., 2015: 'The development of a sound wheel for reproduced sound' in 138th convention of the Audio Engineering Society (Warsaw, Poland).

Princeton University, 2010: 'About wordnet'. Available: <http://wordnet.princeton.edu>

Staffeldt, H., 1974: 'Correlation between subjective and objective data for quality loudspeakers' in 47th convention of the Audio Engineering Society (Copenhagen, Denmark).

Wrzeczono, P. and Marasek, K., 2010: 'Violin sound quality: expert judgements and objective measures' in Advances in Music Information Retrieval, pp. 237–260.

Wu, Z., and Palmer, M., 1994: 'Verb semantics and lexical selection' in 32nd annual meeting of the association of computational linguistics (Las Cruces, New Mexico), pp. 133–138.

Zacharov, N., and Pedersen, T., 2015: 'Spatial sound attributes: development of a common lexicon' in 139th convention of the Audio Engineering Society (New York, USA).

Zwicker, E. and Fastl, H., 1999: 'Psycho-acoustics: Facts and models' (Berlin: Springer), 2nd edn.





APPENDIX 1: Dictionary of timbral terms and hierarchical structure

Table A1 lists the 294 timbral terms that were retained from the attribute reduction described in [Section 2.2](#). The hierarchical location of each term is represented by its container groups, as discussed in [Section 2.3](#).

Table A1: Dictionary of timbral terms and hierarchical grouping				
Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
echoey	apparent reflected energy			echo
echoic	apparent reflected energy			echo
dead	apparent reflected energy			reverb
intimate	apparent reflected energy			reverb
roomy	apparent reflected energy			reverb
reverberant	apparent reflected energy			reverb
dry *	apparent reflected energy			reverb
wet *	apparent reflected energy			reverb
balanced	spectral balance	balance between areas of the frequency spectrum		overall balance
unbalanced	spectral balance	balance between areas of the frequency spectrum		overall balance
bassy *	spectral balance	balance between areas of the frequency spectrum	bassiness	bassiness
booming	spectral balance	balance between areas of the frequency spectrum	bassiness	boominess
boomy	spectral balance	balance between areas of the frequency spectrum	bassiness	boominess
deep *	spectral balance	balance between areas of the frequency spectrum	bassiness	depth





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
fat *	spectral balance	balance between areas of the frequency spectrum	bassiness	fatness
full	spectral balance	balance between areas of the frequency spectrum	bassiness	fullness
muddy	spectral balance	balance between areas of the frequency spectrum	bassiness	muddiness
rumbling	spectral balance	balance between areas of the frequency spectrum	bassiness	rumbliness
huge	spectral balance	balance between areas of the frequency spectrum	bassiness	size
massive	spectral balance	balance between areas of the frequency spectrum	bassiness	size
thick	spectral balance	balance between areas of the frequency spectrum	bassiness	thickness
thin	spectral balance	balance between areas of the frequency spectrum	bassiness	thickness
thundering	spectral balance	balance between areas of the frequency spectrum	bassiness	thunderiness
cold	spectral balance	balance between areas of the frequency spectrum	bassiness	warmth
warm *	spectral balance	balance between areas of the frequency spectrum	bassiness	warmth
heavy *	spectral balance	balance between areas of the frequency spectrum	bassiness	weight
light *	spectral balance	balance between areas of the frequency spectrum	bassiness	weight
weighty	spectral balance	balance between areas of the frequency spectrum	bassiness	weight





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
colored	spectral balance	balance between areas of the frequency spectrum		colouration
colorless	spectral balance	balance between areas of the frequency spectrum		colouration
coloured	spectral balance	balance between areas of the frequency spectrum		colouration
colourless	spectral balance	balance between areas of the frequency spectrum		colouration
equalised	spectral balance	balance between areas of the frequency spectrum		colouration
even	spectral balance	balance between areas of the frequency spectrum		colouration
flat	spectral balance	balance between areas of the frequency spectrum		colouration
neutral	spectral balance	balance between areas of the frequency spectrum		colouration
processed	spectral balance	balance between areas of the frequency spectrum		colouration
uncoloured	spectral balance	balance between areas of the frequency spectrum		colouration
uneven	spectral balance	balance between areas of the frequency spectrum		colouration
boxy	spectral balance	balance between areas of the frequency spectrum	midliness	boxiness
honky	spectral balance	balance between areas of the frequency spectrum	midliness	honkiness
middly	spectral balance	balance between areas of the frequency spectrum	midliness	midliness





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
nasal	spectral balance	balance between areas of the frequency spectrum	midcliness	nasality
present	spectral balance	balance between areas of the frequency spectrum	midcliness	presence
resonant	spectral balance	balance between areas of the frequency spectrum	midcliness	resonance
round	spectral balance	balance between areas of the frequency spectrum	midcliness	roundness
rounded	spectral balance	balance between areas of the frequency spectrum	midcliness	roundness
airy *	spectral balance	balance between areas of the frequency spectrum	trebliness	airiness
bright *	spectral balance	balance between areas of the frequency spectrum	trebliness	brightness
brilliant	spectral balance	balance between areas of the frequency spectrum	trebliness	brightness
dark *	spectral balance	balance between areas of the frequency spectrum	trebliness	brightness
dull *	spectral balance	balance between areas of the frequency spectrum	trebliness	brightness
polished	spectral balance	balance between areas of the frequency spectrum	trebliness	brightness
shining	spectral balance	balance between areas of the frequency spectrum	trebliness	brightness
edgy	spectral balance	balance between areas of the frequency spectrum	trebliness	edginess
penetrating	spectral balance	balance between areas of the frequency spectrum	trebliness	penetration





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
piercing	spectral balance	balance between areas of the frequency spectrum	trebliness	penetration
blunt	spectral balance	balance between areas of the frequency spectrum	trebliness	sharpness
sharp	spectral balance	balance between areas of the frequency spectrum	trebliness	sharpness
shrill	spectral balance	balance between areas of the frequency spectrum	trebliness	shrillness
sibilant	spectral balance	balance between areas of the frequency spectrum	trebliness	sibilance
sizzling	spectral balance	balance between areas of the frequency spectrum	trebliness	sizzle
sparkly	spectral balance	balance between areas of the frequency spectrum	trebliness	sparkle
tinny	spectral balance	balance between areas of the frequency spectrum	trebliness	tinniness
trebly	spectral balance	balance between areas of the frequency spectrum	trebliness	trebliness
dense	spectral balance	levels and positions of individual partials		density
harmonic	spectral balance	levels and positions of individual partials		harmonicity
inharmonic	spectral balance	levels and positions of individual partials		harmonicity
grating	spectral balance	levels and positions of individual partials		harshness
harsh *	spectral balance	levels and positions of individual partials		harshness
hollow	spectral balance	levels and positions of individual partials		hollowness
mellow *	spectral balance	levels and positions of individual partials		hollowness





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
phasey	spectral balance	levels and positions of individual partials		phasiness
rich	spectral balance	levels and positions of individual partials		richness
rough	spectral balance	levels and positions of individual partials		roughness
smooth *	spectral balance	levels and positions of individual partials		roughness
dirty	degree of distortion			dirtiness
clean *	degree of distortion			distortion
clipped	degree of distortion			distortion
distorted *	degree of distortion			distortion
fizzy	degree of distortion			fizziness
compressed	amplitude dynamics			compression
damped	amplitude dynamics			damping
staccato	amplitude dynamics			damping
transient	amplitude dynamics			damping
dynamic	amplitude dynamics			dynamicity
hard *	amplitude dynamics			hardness
pillowy	amplitude dynamics			hardness
soft *	amplitude dynamics			hardness
attacking	amplitude dynamics			punchiness
punchless	amplitude dynamics			punchiness
punchy *	amplitude dynamics			punchiness
tight	amplitude dynamics			tightness
consistent	evolution			consistency
continuous	evolution			consistency
droning	evolution			consistency
evolving	evolution			consistency
homogenous	evolution			consistency
monotonous	evolution			consistency





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
stable	evolution			consistency
steady	evolution			consistency
sustained	evolution			consistency
varied	evolution			consistency
intermittent	evolution			consistency
fluctuating	evolution			flutter
modulated	evolution			flutter
modulating	evolution			flutter
rippling	evolution			flutter
vibrating	evolution			flutter
wavering	evolution			flutter
fluttery	evolution			flutter
trilling	evolution			flutter
beating	evolution			pulsation
pulsating	evolution			pulsation
pulsing	evolution			pulsation
throbbing	evolution			pulsation
thrumming	evolution			pulsation
murmuring	evolution			pulsation
sweeping	evolution			swoosh
swishy	evolution			swoosh
swooshing	evolution			swoosh
whooshy	evolution			swoosh
ambient *	evolution			swoosh
atmospheric	evolution			swoosh
ghostly	evolution			swoosh
soughing	evolution			swoosh
aggressive *	apparent force of production			aggression





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
jarring	apparent force of production	obtrusion		jarring-nature
obtrusive	apparent force of production	obtrusion		obtrusion
clamorous	apparent force of production	obtrusion		raucusness
clangorous	apparent force of production	obtrusion		raucusness
raucous	apparent force of production	obtrusion		raucusness
blaring	apparent force of production			power
blasting	apparent force of production			power
delicate	apparent force of production			power
gentle	apparent force of production			power
gutsy	apparent force of production			power
powerful *	apparent force of production			power
powerless	apparent force of production			power
strident	apparent force of production			power
strong	apparent force of production			power
subdued	apparent force of production			power
tender	apparent force of production			power
weak	apparent force of production			power
fierce	apparent force of production			power
buzzing	noisiness			buzziness
buzzy	noisiness			buzziness





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
hissy	noisiness			hissiness
hummy	noisiness			humminess
noisy *	noisiness			noisiness
hammering	percussive nature	apparent method of percussive generation		knocking-nature
knocking	percussive nature	apparent method of percussive generation		knocking-nature
tapping	percussive nature	apparent method of percussive generation		knocking-nature
pounding	percussive nature	apparent method of percussive generation		knocking-nature
thudding	percussive nature	apparent method of percussive generation		knocking-nature
thumping	percussive nature	apparent method of percussive generation		knocking-nature
grinding	percussive nature	apparent method of percussive generation		scratchiness
scraping	percussive nature	apparent method of percussive generation		scratchiness
scratchy	percussive nature	apparent method of percussive generation		scratchiness
shaking	percussive nature	apparent method of percussive generation		shakiness
agitated	percussive nature	percussive pattern		agitation
chattering	percussive nature	percussive pattern		clatter
clattering	percussive nature	percussive pattern		clatter
pattering	percussive nature	percussive pattern		clatter
crackling	percussive nature	percussive pattern		crackle
creaking	percussive nature	percussive pattern		creakiness
jangling	percussive nature	percussive pattern		janglingness
jingling	percussive nature	percussive pattern		janglingness
rattling	percussive nature	percussive pattern		rattle
rattly	percussive nature	percussive pattern		rattle
rustling	percussive nature	percussive pattern		rustle





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
whirring	percussive nature	percussive pattern		whirr
percussive	percussive nature			percussive nature
bonging	percussive nature	similarity to a known percussive sound		chime
chiming	percussive nature	similarity to a known percussive sound		chime
pinging	percussive nature	similarity to a known percussive sound		chime
pingy	percussive nature	similarity to a known percussive sound		chime
ringing	percussive nature	similarity to a known percussive sound		chime
tinkling	percussive nature	similarity to a known percussive sound		chime
tinkly	percussive nature	similarity to a known percussive sound		chime
clanking	percussive nature	similarity to a known percussive sound		clanky
clanky	percussive nature	similarity to a known percussive sound		clanky
clinking	percussive nature	similarity to a known percussive sound		clanky
clunky	percussive nature	similarity to a known percussive sound		clanky
clanging	percussive nature	similarity to a known percussive sound		clanky
clunking	percussive nature	similarity to a known percussive sound		clanky
clicky	percussive nature	similarity to a known percussive sound		clickiness
cracking	percussive nature	similarity to a known percussive sound		clickiness
snapping	percussive nature	similarity to a known percussive sound		clickiness
ticking	percussive nature	similarity to a known percussive sound		clickiness
crunchy *	percussive nature	similarity to a known percussive sound		crunchiness





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
crashing	percussive nature	similarity to a known percussive sound		explosivity
explosive	percussive nature	similarity to a known percussive sound		explosivity
popping	percussive nature	similarity to a known percussive sound		explosivity
twanging	percussive nature	similarity to a known percussive sound		twang
analogue	similarity to a known sound	made by a particular type of source		analogue-nature
electronic *	similarity to a known sound	made by a particular type of source		electronic-nature
flutey	similarity to a known sound	made by a particular type of source		flutiness
mechanical	similarity to a known sound	made by a particular type of source		mechanical-nature
metallic	similarity to a known sound	made by a particular type of source		metallic-nature
artificial	similarity to a known sound	made by a particular type of source		naturalness
natural *	similarity to a known sound	made by a particular type of source		naturalness
reedy	similarity to a known sound	made by a particular type of source		reediness
string *	similarity to a known sound	made by a particular type of source		stringiness
stringy	similarity to a known sound	made by a particular type of source		stringiness
wooden	similarity to a known sound	made by a particular type of source		woodiness
barking	similarity to a known sound	made by an animal		barking-nature
barky	similarity to a known sound	made by an animal		barking-nature
bleating	similarity to a known sound	made by an animal		bleating-nature
braying	similarity to a known sound	made by an animal		braying-nature





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
cheeping	similarity to a known sound	made by an animal		chirping-nature
chirping	similarity to a known sound	made by an animal		chirping-nature
cooing	similarity to a known sound	made by an animal		cooing-nature
growling	similarity to a known sound	made by an animal		growling-nature
mewing	similarity to a known sound	made by an animal		mewing-nature
mooring	similarity to a known sound	made by an animal		mooring-nature
neighing	similarity to a known sound	made by an animal		neighing-nature
purring	similarity to a known sound	made by an animal		purring-nature
quacking	similarity to a known sound	made by an animal		quacking-nature
quacky	similarity to a known sound	made by an animal		quacking-nature
roaring	similarity to a known sound	made by an animal		roaring-nature
squawking	similarity to a known sound	made by an animal		squawking-nature
squeaking	similarity to a known sound	made by an animal		squeaking-nature
whistling	similarity to a known sound	made by an animal		whistling-nature
yapping	similarity to a known sound	made by an animal		yapping-nature
yelping	similarity to a known sound	made by an animal		yapping-nature
beepy	similarity to a known sound	made by a synthetic source		beepiness
peeping	similarity to a known sound	made by a synthetic source		beepiness
bipping	similarity to a known sound	made by a synthetic source		beepiness





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
bleeping	similarity to a known sound	made by a synthetic source		beepiness
bellowing	similarity to a known sound	made by a voice		bellowing-nature
breathy	similarity to a known sound	made by a voice		breathiness
cackling	similarity to a known sound	made by a voice		cackling-nature
groaning	similarity to a known sound	made by a voice		groaning-nature
gruff	similarity to a known sound	made by a voice		gruffness
guttural	similarity to a known sound	made by a voice		guttural-nature
hoarse	similarity to a known sound	made by a voice		hoarseness
moaning	similarity to a known sound	made by a voice		moaning-nature
rasping	similarity to a known sound	made by a voice		raspiness
raspy	similarity to a known sound	made by a voice		raspiness
raw	similarity to a known sound	made by a voice		rawness
screaming	similarity to a known sound	made by a voice		screaming-nature
screeching	similarity to a known sound	made by a voice		screaming-nature
screechy	similarity to a known sound	made by a voice		screaming-nature
squealing	similarity to a known sound	made by a voice		screaming-nature
snarling	similarity to a known sound	made by a voice		snarling-nature
sonorous	similarity to a known sound	made by a voice		sonorous-nature
strained	similarity to a known sound	made by a voice		straining-nature





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
voice-like	similarity to a known sound	made by a voice		voice-like-nature
wailing	similarity to a known sound	made by a voice		wailiness
warbling	similarity to a known sound	made by a voice		warbling-nature
wheezing	similarity to a known sound	made by a voice		wheeziness
wheezy	similarity to a known sound	made by a voice		wheeziness
whining	similarity to a known sound	made by a voice		whininess
whiny	similarity to a known sound	made by a voice		whininess
whispering	similarity to a known sound	made by a voice		whispering-nature
babbling	similarity to a known sound	made by water		bubbling-nature
bubbling	similarity to a known sound	made by water		bubbling-nature
gurgling	similarity to a known sound	made by water		bubbling-nature
lapping	similarity to a known sound	made by water		lapping-nature
popping	similarity to a known sound	made by water		popping-nature
plashing	similarity to a known sound	made by water		splashing-nature
splashing	similarity to a known sound	made by water		splashing-nature
splashy	similarity to a known sound	made by water		splashing-nature
sputtering	similarity to a known sound	made by water		sputting-nature
clashing	tonality	tonal purity	tonal polyphony	tonal consonance
consonant	tonality	tonal purity	tonal polyphony	tonal consonance
discordant	tonality	tonal purity	tonal polyphony	tonal consonance
harmonious	tonality	tonal purity	tonal polyphony	tonal consonance





Table A1 (continued): Dictionary of timbral terms and hierarchical grouping

Dictionary term	Level-1 group	Level-2 group	Level-3 group	Timbral attribute
polyphonic	tonality	tonal purity	tonal polyphony	polyphony
impure	tonality			tonal purity
complex	tonality			tonal purity
pure	tonality			tonal purity
tonal	tonality			tonality
transparent	transparent			transparency
blurry	transparent	clarity		blurriness
focused	transparent	clarity		blurriness
unfocused	transparent	clarity		blurriness
clear	transparent	clarity		clarity
muffled	transparent	clarity		clarity
unclear	transparent	clarity		clarity
cluttered	transparent	clarity		clutter
crisp	transparent	clarity		definition
defined	transparent	clarity		definition
distinct	transparent	clarity		definition
fuzzy	transparent	clarity		fuzziness
garbled	transparent	clarity		garbled-nature
hazy	transparent	clarity		haziness
veiled	transparent	clarity		haziness
muted	transparent	clarity		muted-nature
detailed	transparent			definition
precise	transparent			definition
closed	transparent			openness
constricted	transparent			openness
open	transparent			openness
restricted	transparent			openness

