

Four distinctions for the auditory "wastebasket" of timbre

Kai Siedenburg^{1, 2*}, Stephen McAdams¹

¹Schulich School of Music, McGill University, Canada, ²Department of Medical Physics and Acoustics, Carl von Ossietzky University of Oldenburg, Germany

Submitted to Journal:
Frontiers in Psychology

Specialty Section:
Auditory Cognitive Neuroscience

Article type:
Opinion Article

Manuscript ID:
272043

Received on:
05 Apr 2017

Revised on:
07 Sep 2017

Frontiers website link:
www.frontiersin.org

In review

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

KS and SMC discussed ideas relevant to the topic. KS devised the first draft of the manuscript, which subsequently underwent several substantial revisions based on joint discussion among the co-authors.

Keywords

timbre perception, music cognition, Psychoacoustics, Conceptual framework, definitions

Funding statement

This work was supported by a grant from the Canadian Natural Sciences and Engineering Research Council (RGPIN 2015-05280) and a Canada Research Chair (950-223484) awarded to Stephen McAdams.

In review

Four distinctions for the auditory “wastebasket” of timbre*

Kai Siedenburg^{1,2,*} and Stephen McAdams¹

¹Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT),
Schulich School of Music, McGill University, Montreal, QC, Canada

²Department of Medical Physics and Acoustics, University of Oldenburg, Germany

Correspondence*:

Kai Siedenburg

kai.siedenburg@uni-oldenburg.de

KEYWORDS:

2 timbre perception, music cognition, psychoacoustics, conceptual framework, definitions

1 INTRODUCTION

3 If there is one thing about timbre that researchers in psychoacoustics and music psychology agree on, it is
4 the claim that it is a poorly understood auditory attribute. One facet of this commonplace conception is that
5 it is not only the complexity of the subject matter that complicates research, but also that timbre is hard to
6 define (cf., Krumhansl, 1989). Perhaps for lack of a better alternative, one can observe a curious habit in
7 introductory sections of articles on timbre, namely to cite a definition from the American National Standards
8 Institute (ANSI) and to elaborate on its shortcomings. For the sake of completeness (and tradition!) we
9 recall:

10 “Timbre. That attribute of auditory sensation which enables a listener to judge that two nonidentical
11 sounds, similarly presented and having the same loudness and pitch, are dissimilar [sic]. NOTE-Timbre
12 depends primarily upon the frequency spectrum, although it also depends upon the sound pressure
13 and the temporal characteristics of the sound.”(ANSI, 1960/1994, p. 35)

14 One of the strongest criticisms of this conceptual framing was given by Bregman (1990), commenting,

15 “This is, of course, no definition at all. [...] The problem with timbre is that it is the name for an
16 ill-defined wastebasket category. [...] I think the definition [...] should be this: ‘We do not know how
17 to define timbre, but it is not loudness and it is not pitch.’ [...] What we need is a better vocabulary
18 concerning timbre.”(pp. 92-93)

19 In an even more radical spirit, Martin (1999, p. 43) proposed, “[Timbre] is empty of scientific meaning,
20 and should be expunged from the vocabulary of hearing science.” Almost twenty years later, although the
21 notion is still part of the terminology, we are far from having reached a clearer taxonomy. One could even

*This manuscript is a revised version of a chapter from the doctoral dissertation of the first author (03/2016, McGill University, Ch. 2). A panel at the 2017 Berlin Interdisciplinary Workshop on Timbre discussed the same topic (see, https://www.youtube.com/playlist?list=PL9-Wvg1IK10jCMN3uEs4L7_aIt6B6GV1g).

22 ask: Can something useful be done with the wastebasket in the end? In what follows, we propose four
23 conceptual distinctions for timbre.

2 TIMBRE IS A PERCEPTUAL ATTRIBUTE

24 Already in the 19th century, the title of Helmholtz’s seminal treatise “On the sensations of tone as
25 a physiological basis for the study of music” (von Helmholtz, 1885/1954) distinguishes an external
26 physical sound event (the tone) from its internal perceptual representation (the sensation). The sensation
27 comprises subjective auditory attributes such as pitch, loudness, and timbre, but the physical tone does
28 not. Accordingly, the ANSI definition explicitly addresses sensory attributes. There are, unfortunately,
29 many examples of a different type of usage, where timbre is primarily used to refer to features of physical
30 sound events. These cannot only be found in adjacent academic disciplines such as music theory or music
31 information retrieval, but even in music psychology, where the term is at times used as a shorthand for a
32 sound event or complex tone, the relevant perceptual attribute of which is timbral in nature (e.g., “listeners
33 were presented with three timbres”). This shorthand usage is tempting but harmful. It encourages the
34 reader to equate the sound event and its timbre, which are in reality connected by a complex sequence
35 of information-processing steps in the human auditory system. It becomes particularly problematic in
36 conjunction with ecological views of perception, which often appear to circumvent the problem of
37 information transformation by proclaiming a direct correspondence between perception and the world. As
38 noted by Clarke (2005),

39 “The amplitude and frequency distribution of the sounds emitted when this piece of hollowed wood
40 is struck are a direct consequence of the physical properties of the wood itself—are an ‘imprint’
41 of its physical structure—and an organism does not have to do complex processing to ‘decode’
42 the information within the source: it needs to have a perceptual system that will resonate to the
43 information.” (p.18)

44 A crux of the belief that the perceptual system is attuned to the “perceptual invariants” of the environment
45 is, however, that “the detection of physical invariants, like image surfaces, is exactly and precisely an
46 information-processing problem, in modern terminology” (Marr, 1982, p. 30). We need to study the ways in
47 which auditory representations are robust to transformations of the acoustic signal given a specific context,
48 in order to understand the correspondence of tone and sensation.

49 One can even observe more hazardous attempts to rephrase timbre as not primarily depending on
50 perception. In a recent ANSI critique from a composer’s viewpoint, Roads (2015) states,

51 “[The ANSI definition] describes timbre as a perceptual phenomenon, and not as an attribute of a
52 physical sound. Despite this, everyone has an intuitive sense of timbre as an attribute of a sound
53 like pitch or loudness (e.g., ‘the bassoon timbre’ [...]). From a compositional point of view, we are
54 interested in the physical nature of timbre [...] in order to manipulate it for aesthetic purposes.”(p.
55 xviii)

56 On the contrary, we insist that timbre is a *perceptual* attribute, as are pitch and loudness. Furthermore, there
57 does not exist *the* bassoon timbre, but rather *a* bassoon timbre at a given pitch and dynamic, produced with a
58 specific articulation and playing technique (see Sec. 4). In order not to let the indispensable interdisciplinary
59 discourse around timbre disintegrate into terminological incoherence, we should resist tempting shorthands
60 right from the start and clearly separate physical sound events or tones and their morphologies (as well as

61 their representations via musical scores, sampled time-pressure audio signals, spectrotemporal analyses,
62 etc.) from the resulting auditory sensations. The three distinctions that follow consequently address timbre
63 as a perceptual attribute.

3 TIMBRE IS A QUALITY AND A CONTRIBUTOR TO SOURCE IDENTITY

64 There are two standard approaches in which timbre as a perceptual attribute is defined. Both approaches
65 consider timbre as a bundle of auditory sensory features, to which, however, subtly different functions are
66 ascribed. On the one hand, there is the (ANSI-like) definition by negation that encompasses all auditory
67 attributes that allow listeners to perceive differences between sounds of equalized pitch, loudness, and say,
68 spatial position. Here, the function of timbre attributes remains as vague as to allow listeners to engage in
69 dissimilarity ratings and discrimination tasks. In this approach, timbre is referred to as quality: Two sounds
70 can be declared qualitatively dissimilar without bearing semantic associations or without their source/cause
71 mechanisms being identified. On the other hand, timbre is indeed defined via this latter role, namely as
72 that collection of auditory sensory features that primarily contributes to the inference (or specification) of
73 sound sources and events (although timbral differences do not always correspond to differences in sound
74 sources, see below). Here the function ascribed to timbral attributes is tied to an identification task.

75 The difference between viewing timbre from the angles of qualitative comparison and source identification
76 is not always clearly articulated. Dissimilarity studies that investigate timbre as qualia and work with
77 acoustic stimuli may fail to account for the effects of source identification in dissimilarity ratings. In fact, the
78 latent structure that underlies dissimilarity ratings is modeled by acoustic properties, implicitly assuming
79 that dissimilarity ratings are solely based on the sensory representation of the sounds' acoustic features and
80 not influenced by semantic categories elicited by the features of sound sources. It is questionable whether
81 source identification can be neglected for acoustic stimuli, however, as one might argue that listeners "can't
82 help" but integrate semantic information into dissimilarity ratings of Western orchestral instrument tones
83 (Siedenburg et al., 2016b). In order not to conflate a study of sensory similarity with semantic factors, it is
84 important to take into account the distinction between timbre as a quality and timbre as a contributor to
85 source identity (also see, Lemaitre et al., 2010).

4 TIMBRE FUNCTIONS ON DIFFERENT SCALES OF DETAIL

86 When Helmholtz noted "By the quality of a tone [*Klangfarbe*] we mean that peculiarity which distinguishes
87 the musical tone of a violin from that of a flute or that of a clarinet or that of the human voice, when
88 all these instruments produce the same note at the same pitch."(von Helmholtz, 1885/1954, p. 10), he
89 (perhaps unwittingly) provided the textbook definition of timbre for the next 150 years. This sentence
90 operationalizes timbre via the perceptual differences based on the distinct acoustics of sound sources such
91 as the flute and clarinet, and, like the ANSI definition, only compares timbre across tones with the same
92 pitch, loudness, and duration.

93 Apart from the cul-de-sac in which this definition deprives any non-pitched sound of its timbre (Bregman,
94 1990, p. 92), the approach also neglects the fact that most pitched musical instruments can give rise to whole
95 palettes of distinct timbral qualities which covary with pitch and loudness. Not only do different playing
96 techniques and articulations affect physical and timbral properties of tones (e.g., Barthelet et al., 2010),
97 but a *fortissimo* comes with many pronounced partials (and a correspondingly bright timbre), whereas
98 a *pianissimo* yields significantly attenuated amplitudes of higher order partials (Meyer, 1995). A tone's
99 spectral content also covaries with fundamental frequency (F_0) and playing effort. Low-pitched registers

100 comprise many partial tones, higher tones do not. The acoustical covariance of F_0 and spectrotemporal
101 envelope shape appears to lead to small but systematic interactions between pitch and timbre (e.g., Marozeau
102 and de Cheveigné, 2007), and these relations appear to be supported by perceptual learning (Sandell and
103 Chronopoulos, 1997) and musical training (Steele and Williams, 2006). The corresponding pitch-timbre
104 “covariance matrices” are likely to be used as a valuable perceptual cue for source identification (Handel
105 and Erickson, 2004), although this research topic has been barely explored.

106 On an even more fine-grained scale, there can be differences between sounds from exemplars of the
107 same type of sound-producing objects or algorithms (such as a Stradivarius violin and an inexpensive
108 factory-made model). The ways in which this translates into audible timbral differences and how these
109 relate to judged instrument quality (in the sense of good vs. bad) is yet another research topic (cf., Saitis
110 et al., 2012).

111 In sum, it is misleading to suggest that one sound-producing object or instrument yields exactly one
112 timbre. Contrary to parlance of “the bassoon timbre”, there is no single timbre that *fully* characterizes the
113 bassoon. The timbre of a bassoon tone depends on pitch, playing effort, articulation, fingering, etc. In light
114 of a biological analogy, a single type of sound-producing object or sound-synthesis algorithm may give
115 rise to a timbral *genus* that can encompass various timbral *species*. These species may feature systematic
116 variation along various parameters, such as playing technique, covariance with pitch and loudness, or
117 expressive intent. Genera group into *families* (e.g., corresponding to the timbres from string vs. brass
118 instruments) and at some point into *kingdoms* (timbres related to, say, acoustic vs. electronic means of
119 sound production). Overall, this yields a “hierarchy of embedded distinctions” (Krumhansl, 1989, p. 45)
120 that encompasses scales of different timbral detail to which the ANSI definition is agnostic and the textbook
121 definition ignorant.

5 TIMBRE IS A PROPERTY OF FUSED AUDITORY EVENTS

122 Polyphonic music is the unequivocal target territory for timbre research. Consequently, studies are beginning
123 to explore the acoustic correlates of what has been called “polyphonic timbre” (Alluri and Toiviainen,
124 2010), “capturing the overall emerging timbre of a jazz ensemble, a rock concert, or a symphony”, and
125 thus the “global sound” of a piece of music.

126 Analogous to pitch and loudness, however, we view timbre as a perceptual property of perceptually fused
127 auditory events. If two or more auditory events do not fuse, they do not contribute to the same timbre.
128 Sounds from a bass-drum, a handclap, and a synth pad usually do not fuse into a single auditory image,
129 such that each of these sounds will possess an individual timbre in the mind of a listener. It is the emergent
130 property of the combination of the individual timbres that evokes hip-hop, but there is no a unitary “hip-hop
131 timbre”.

132 In fact, auditory scene analysis (ASA) principles do not provide a definitive borderline of where
133 segregation ends, because stream formation depends on the listener’s focus in the ASA hierarchy. Not
134 entirely fused (heterogeneous) musical lines can be heard as one stream or many, depending on auditory
135 focus and musical context. On the other hand, completely disregarding ASA processes by extracting
136 features from the audio mixture may contribute to the reported limitations in using music information
137 retrieval algorithms as perceptual models (cf., Siedenburg et al., 2016a). As perhaps best summarized by
138 Aucouturier and Pachet (2007, p. 659),

139 “Overall, this suggests that the horizontal coding of frames of data, without any account of source
140 separation and selective attention, is a very inefficient representation of polyphonic musical data, and
141 not cognitively plausible.”

142 A metaphor might drawn from the relation between pitch and harmony perception, where one can still
143 hear individual pitches (timbres), but there is another quality that emerges from the relations among the
144 pitches (timbres). Hence, rather than presupposing that polyphonic music gives rise to unitary auditory
145 images (which the notion of “polyphonic timbre” suggests), we believe that it is the combinatorial interplay
146 of timbres that is at the heart of the perception of polyphonic music.

6 CONCLUSION

147 By proposing four basic distinctions for the notion of timbre we hope to clear up some confusion around
148 what has been claimed to be the terminological wastebasket of music psychology and psychoacoustics—
149 musical timbre. In direct opposition to physical realists such as Isaac (2017), we propose to locate timbre
150 on the perceptual side of the “psychophysical divide”, i.e., in the mind of the listener instead of in physical
151 properties. We further argue that the notion is commonly viewed from different angles: as qualia and as
152 a contributor to source identity, but the language around this distinction needs to be clarified to avoid
153 confusion between them. We have illustrated that there may be large- or small-scale timbral differences
154 (e.g., arising from timbral families vs. species), and that timbre is a property of fused auditory events
155 instead of multi-stream auditory mixtures. We do not claim that this is an exhaustive categorization—
156 more fine-grained taxonomies must be developed in order to account for timbre’s perceptual richness.
157 Nonetheless, the four proposed distinctions may serve as a basic taxonomy to clarify discourse in future
158 inquiries into timbre. Furthermore, each distinction encompasses its own host of research questions that
159 subsequent empirical work may address. In any case, once a few layers of dust are removed, what we had
160 thought of as a wastebasket turns out to be a colorful umbrella(-term) upside down.

161 The composer Philippe Manoury (1991) observed that “One of the most striking paradoxes concerning
162 timbre is that when we knew less about it, it didn’t pose much of a problem.”(p. 293) This can also be put in
163 more optimistic terms: We already know much about timbre. We understand its plentiful, distinct colors are
164 real, and they won’t go away. It is time to let inadequate standards rest and start to focus on the specifics.

7 ACKNOWLEDGMENTS

165 We wish to thank the reviewer for insightful and stimulating comments. *Funding:* This work was supported
166 by a grant from the Canadian Natural Sciences and Engineering Research Council (RGPIN 2015-05280)
167 and a Canada Research Chair (950-223484) awarded to Stephen McAdams.

REFERENCES

- 168 Alluri, V. and Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre.
169 *Music Perception* 27, 223–241
- 170 ANSI (1960/1994). Psychoacoustic terminology: Timbre
- 171 Aucouturier, J.-J. and Pachet, F. (2007). The influence of polyphony on the dynamical modelling of musical
172 timbre. *Pattern Recognition Letters* 28, 654–661
- 173 Barthes, M., Guillemain, P., Kronland-Martinet, R., and Ystad, S. (2010). From clarinet control to timbre
174 perception. *Acta Acustica united with Acustica* 96, 678–689

- 175 Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA:
176 MIT Press)
- 177 Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*
178 (Oxford, UK: Oxford University Press)
- 179 Isaac, A. (2017). Prospects for timbre physicalism. *Philosophical Studies* published online, Feb 13 2017
- 180 Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In *Structure and perception of*
181 *electroacoustic sound and music*, eds. S. Nielzén and O. Olsson (Amsterdam, The Netherlands: Excerpta
182 Medica), vol. 846. 43–53
- 183 Lemaitre, G., Houix, O., Misdariis, N., and Susini, P. (2010). Listener expertise and sound identification
184 influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied* 16,
185 16–32
- 186 Manoury, P. (1991). Les limites de la notion de ‘timbre’. In *Le timbre: Métaphore pour la composition*, ed.
187 J.-B. Barriere (Christian Bourgois, Paris). 293–299
- 188 Marozeau, J. and de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness
189 dimension of timbre. *The Journal of the Acoustical Society of America* 121, 383–387
- 190 Marr, D. (1982). *Vision: A computational approach* (San Francisco, CA: W. H. Freeman & Co)
- 191 Martin, K. D. (1999). *Sound-source recognition: A theory and computational model*. Ph.D. thesis,
192 Massachusetts Institute of Technology
- 193 Meyer, J. (1995). *Akustik und musikalische Aufführungspraxis: Leitfaden für Akustiker, Tonmeister, Musiker,*
194 *Instrumentenbauer und Architekten* (Bergkirchen, Germany: Bochinsky)
- 195 Roads, C. (2015). *Composing Electronic Music: A New Aesthetic* (Oxford, UK: Oxford University Press)
- 196 Saitis, C., Giordano, B. L., Fritz, C., and Scavone, G. P. (2012). Perceptual evaluation of violins: A
197 quantitative analysis of preference judgments by experienced players. *The Journal of the Acoustical*
198 *Society of America* 132, 4002–4012
- 199 Sandell, G. J. and Chronopoulos, M. (1997). Perceptual constancy of musical instrument timbres;
200 generalizing timbre knowledge across registers. In *Proc. 3rd Triennial ESCOM Conference, Uppsala,*
201 *Sweden*, ed. A. Gabrielsson. 222–227
- 202 Siedenburg, K., Fujinaga, I., and McAdams, S. (2016a). A comparison of approaches to timbre descriptors
203 in music information retrieval and music psychology. *Journal of New Music Research* 45, 27–41
- 204 Siedenburg, K., Jones-Mollerup, K., and McAdams, S. (2016b). Acoustic and categorical dissimilarity
205 of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds. *Frontiers in*
206 *Psychology* 6, doi: 10.3389/fpsyg.2015.01977
- 207 Steele, K. M. and Williams, A. K. (2006). Is the bandwidth for timbre invariance only one octave? *Music*
208 *Perception* 23, 215–220
- 209 von Helmholtz, H. (1885/1954). *On the sensations of tone as a physiological basis for the theory of music*
210 (New York, NY: Dover), trans. by A. J. Ellis of 4th German ed., 1877, republ. 1954 edn.