©2020 The Acoustical Society of Japan

ACOUSTICAL LETTER

Timbre semantics through the lens of crossmodal correspondences: A new way of asking old questions

Charalampos Saitis[1,*], Stefan Weinzierl[1], Katharina von Kriegstein[2,3], Sølvi Ystad[4] and Christine Cuskley[5]

[1]*Audio Communication Group, Technische Universität Berlin, Sekretariat E-N 8, Einsteinufer 17c, 10587 Berlin, Germany*
[2]*Faculty of Psychology, Technische Universität Dresden, Bamberger Str. 7, 01187 Dresden, Germany*
[3]*Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany*
[4]*PRISM (Perception, Representations, Image, Sound and Music), AMU-CNRS,*
31 Chemin Joseph Aiguier, 13402 Marseille, France
[5]*School of English Language, Literature and Linguistics, Newcastle University,*
Percy Building, NE1 7RU, Newcastle upon Tyne, United Kingdom

(*Received 30 September 2018, Accepted for publication 19 May 2019*)

Keywords: Timbre semantics, Crossmodal correspondence, Conceptual metaphor
PACS number: 43.66.Jh [doi:10.1250/ast.41.365]

1. Introduction

This position paper argues that a systematic study of the behavioral and neural mechanisms of crossmodal correspondences between timbral dimensions of sound and perceptual dimensions of other sensory modalities, such as brightness, roughness, or sweetness, can offer a new way of addressing old questions about the perceptual and neurocognitive mechanisms of auditory semantics. At the same time, timbre and the crossmodal metaphors that dominate its conceptualization can provide a test case for better understanding the neural basis of crossmodal correspondences and human semantic processing in general.

2. Motivation

Timbre is one of the most fundamental aspects of acoustical communication and yet it remains one of the most poorly understood. The remarkable ability of the brain to recognize the source of a sound — glass breaking, footsteps approaching, a singer's voice, a musical instrument — stems in part from a capacity to perceive and process differences in the timbre of sounds. Despite being an intuitive concept, however, timbre covers a very complex set of auditory attributes that are not accounted for by frequency, intensity, duration, spatial location, and the acoustic environment [1]. Furthermore, people lack a specific sensory vocabulary for sound. Instead, sound qualities are communicated primarily through sensory attributes from different modalities (e.g., bright, warm, sweet) but also through onomatopoeic attributes (e.g., ringing, buzzing, shrill) or through nonsensory attributes relating to abstract constructs (e.g., rich, complex, harsh).

Research in timbre semantics has long aimed to identify the few salient semantic substrates of linguistic descriptions of timbral impressions that can yield consistent and differentiating responses to different timbres, along with their acoustical correlates (see [2] for a comprehensive review). In the most commonly adopted approach, timbre is considered as a set of verbally defined perceptual attributes that represent the dimensions of a semantic space, derived through factor analysis of ratings along verbal scales known as semantic differentials [3]. The latter are typically constructed either by two opposing descriptive adjectives such as "bright–dull" or by an adjective and its negation as in "bright–not bright." Previous studies have identified three salient semantic dimensions for timbre, which can broadly be interpreted in terms of luminance, texture, and mass [4,5]. The first appears to be associated with the energy midpoint of the spectral distribution, the second with fine spectrotemporal modulations, and the third with the width of the spectral distribution.

The semantic differential method has been instrumental in advancing the scientific understanding of timbre. Yet the view that the complex multivariate character of meaning can be captured by a low-dimensional spatial configuration can be challenged. A different approach relies on cognitive categories emerging from psycholinguistically inferred semantic relations in free verbalizations of sound qualities. Such analyses have provided additional insight regarding particular factors that contribute to the salient semantic dimensions of timbre [6–8]. Still, both semantic differential scales and free verbalization tasks seem to miss an important point: sensory nonauditory attributes of timbre exemplify a more ubiquitous aspect of human cognition known as *crossmodal correspondences*: people tend to map between sensory experiences in different modalities (e.g., between color and touch [9]) or within the same modality (e.g., between pitch, timbre, and loudness [10]).

Our current understanding of crossmodal correspondences strongly resembles a "black box": there is ample evidence of consistently regular mappings between modalities but limited knowledge of both the psychophysics and higher cognitive processes that govern those mappings. In the case of sound, there is a growing body of studies documenting the behavior of associations between pitch and other modalities (e.g., pitch-height and -brightness; see [11] for a review) but similar research on timbre is still very limited [12–18]. In addition,

[*] Present address: Centre for Digital Music, Queen Mary University of London, United Kingdom. e-mail: c.saitis@qmul.ac.uk

there are currently very few published neuroscientific studies explicitly looking at auditory-nonauditory correspondences [19–21].

Observing certain crossmodal mappings in preverbal infants [22,23] suggests that they may reflect structural similarities shared across modality-specific sensory coding at a purely perceptual (i.e., prelinguistic or nonlinguistic) level. Such accounts may be extended to embodied conceptual representations grounded in perception and action and on the statistics of the environment. Pitch-height mappings, for example, may originate in bodily experience, because people's larynges rise when they produce higher pitches and descend when they produce lower pitches. Furthermore, a robust mapping seems to exist between the frequency of a sound and the average elevation of its source in the statistics of natural auditory scenes and in the filtering properties of the outer ear [24]. This is further supported by behavioural evidence showing a strong interaction between the pitch-elevation correspondence and auditory elevation [25].

However, strictly embodied explanations of concepts may be insufficient to explain all crossmodal associations, especially those observed in adults as well as children at least 5–9 years old where language is engaged to describe perceptions and which appear to emerge during late decisional rather than early perceptual processes [26]. Such evidence suggest that even if some crossmodal associations have their origins in perception and action, through continuous cultural learning they may become incorporated in language and thus mediated by semantic processes; moreover, they may arise from supramodal conceptual representations established after stimulus features have been recoded into an abstract semantic format common to perceptual and linguistic systems [27–30]. Accordingly, statistical regularities between frequency and elevation in the environment [24] might become incorporated in language so that both pitches and heights come to be described as high or low. It is also plausible that the brain might have evolved to develop mechanisms that internalize environmentally associated features as common neural codes that respond to certain stimulus features regardless of modal content or co-locate them in the respective modality-specific regions via direct communication.

Neuroimaging data demonstrate that semantic processing in the brain involves direct interaction and exchange of information between modality-specific sensorimotor areas, possibly through synchronized activity, but also recruits a large network of so-called supramodal regions (auditory-visual correspondences [19–21]; auditory brightness [31]; auditory size [32]; voice recognition [33]; "noisy/rough" timbres [34]; general conceptual processing [35–37]). According to a theory of "embodied abstraction," modality-specific perceptual systems may provide the primary mechanism for acquiring concepts and grounding them in the external world, while supramodal zones enable the gradual abstraction of unimodal sensorimotor simulations to facilitate highly schematic conceptual functions [36].

3. A research roadmap

In viewing timbre semantics through the lens of crossmodal correspondences, questions about the psychoacoustics and neural basis of the former can thus be reconsidered: What intrinsic timbral properties of sound evoke the analogous impression as touching a velvety surface or viewing a hollow object? Are perceptual attributes of different sensory experiences (e.g., a smooth surface, a sweet taste, and a rounded form) mapped to similar or distinct timbres? Do crossmodal timbral attributes (e.g., bright, warm, sweet) correspond to common, supramodal neural configurations, or do they trigger matching responses between the auditory and the respective modality-specific (e.g., visual, somatosensory, gustatory) areas? To address these questions, a comprehensive examination of auditory-nonauditory correspondences is needed, including the collection of behavioral and neuroimaging data from appropriate tasks.

Previous work has three important methodological limitations. First, the use of words to convey sensory attributes (e.g., using the word "sharp" instead of a sharp form) might have influenced the investigated associations because of analogous mappings existing between linguistic features of words and visual forms [15]. Second, stimuli (linguistic or physical) were often reduced to two values per modality with no grades in between. Such choices implicitly assume that crossmodal associations are purely context-sensitive and monotonic, but evidence of absolute or nonlinear mappings challenge such assumptions (e.g., [38]). Additionally, participants might have explicitly categorized stimuli in terms of opposing poles rather than based on the actual mapping of one sensory cue to another [39]. Third, pertaining only to the few timbre-based studies, sound stimuli tended to be limited to recorded notes from musical instruments, which may implicate source-cause categories [40].

A systematic investigation of crossmodal correspondences between timbre and nonauditory perceptual dimensions therefore first necessitates auditory stimuli that can be manipulated along intrinsic continuous dimensions of timbre. These can be obtained through a framework for analysis-synthesis that uses "abstract" sounds as the source material. An example of such sounds is what Pierre Schaeffer named *acousmatic*, where sound identities appear intentionally obscured or unconnected to their source, and which have been shown to activate crossmodal conceptual associations of the type proposed to investigate here [41–43]. We propose to follow an exploratory approach using well-established methods from experimental psychology and sound synthesis. Using the semantic differential, acousmatic sounds that evoke nonauditory sensory attributes can be obtained and subsequently subjected to dissimilarity ratings. The resulting "timbre space" [1] will form the basis for developing timbre morphing continua along its dimensions using granular synthesis. A granular representation is better suited for the kind of "abstract" sounds proposed here as the source material for analysis-synthesis [43].

Next, to facilitate the crossmodal matching of auditory-nonauditory sensory experiences that may evoke the same or analogous concepts, nonlinguistic nonauditory stimuli should be designed along perceptually gradient scales on the basis of experiencing actual real-world sensations. Nonauditory sensory attributes typically come from the modalities of vision, touch, and gustation. More precisely, we have identified 16 such attributes of timbre: (visual) *bright*, *dark*, *deep*, *thick*, *thin*, *hollow*, *full*, *round*, *rounded*, *sharp*; (tactile) *soft*, *hard*, *smooth*, *rough*, *warm*; and (gustatory) *sweet*. These can be grouped into pseudo-bipolar pairs (i.e., they do not necessarily represent true psychologically bipolar continua in any modality) to facilitate the construction of the stimuli. For the latter, tactile surfaces with carefully controlled dimensions (e.g., hardness, roughness) can be developed by means of either physical materials [9] or haptic synthesis [44]. Similarly, the use of three-dimensional visual modeling and virtual reality technologies may offer a promising approach for inducting individual attributes of visual stimuli (e.g., thickness [23]) in a highly controlled manner. With respect to gustatory stimuli, tastants [15] or mixtures [45] can be used to create different concentrations of basic tastes.

Crossmodal correspondences between timbre and perceptual dimensions of other modalities can be quantified in terms of shared mappings between physical or virtual visual/tactile/gustatory stimuli varying along their respective dimensions and synthesized sounds varying along the timbre morphing continua. The latter should remain the same across auditory-visual, -tactile, and -gustatory crossmodal matching designs in order to always investigate the same timbral dimensions. Next, as a first step towards investigating the neurobiological mechanisms of auditory-nonauditory crossmodal correspondences, congruent (i.e., systematically matched) auditory-nonauditory stimulus pairs derived from behavioral tasks can be used in neuroimaging designs involving contrasts between congruent and incongruent crossmodal matching.

In their study of crossmodal associations between sounds and tastes, Simner and colleagues [15] used formant synthesis to create four vowel quality continua (first formant, second formant, voice discontinuity, spectral balance). Listeners selected preferred positions along the continua to accompany each of the basic tastes of sweet, sour, bitter, and salty, each received at two different concentrations. It was found that mappings of sour, bitter, and salty generally patterned with each other, while sweet patterned away from all three other tastes. Indeed, in taste perception humans often confuse sour with salty and/or bitter, and occasionally salty with bitter, but they always discriminate between sweet and all other tastes [46]. It was further observed that, when participants matched sounds of low to mid to high frequency to different tastes, a corresponding hierarchy of sweet to bitter to sour emerged. These results demonstrate the potential of the methods proposed in this research roadmap to retrieve relations between timbral dimensions of sound and perceptual dimensions of other modalities.

Stimulating new concepts for sound synthesis, visual signal processing, haptic displays, and virtual reality, the proposed research can bring a new perspective into the study of auditory perception and communication, and open pathways to the development of new semantic technologies in music and speech. The use of brain imaging methods to study the neural substrates of crossmodal correspondences is especially timely. The first neuroimaging studies specifically looking at crossmodal correspondences have only recently started to appear, demonstrating the new and important insights that can be gained, but also the challenges in designing appropriate tasks, such as implicit multisensory attention [21].

Acknowledgments

The first author wishes to thank the Alexander von Humboldt Foundation for support through a Humboldt Research Fellowship (2016–2018).

References

[1] K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper and R. R. Fay, Eds., *Timbre: Acoustics, Perception, and Cognition* (Springer, Cham, 2019).

[2] C. Saitis and S. Weinzierl, "The semantics of timbre," in *Timbre: Acoustics, Perception, and Cognition*, K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper and R. R. Fay, Eds. (Springer, Cham, 2019), pp. 119–149.

[3] C. E. Osgood, "The nature and measurement of meaning," *Psychol. Bull.*, **49**, 197–237 (1952).

[4] G. von Bismarck, "Timbre of steady tones: A factorial investigation of its verbal attributes," *Acustica*, **30**, 146–159 (1974).

[5] A. Zacharakis, K. Pastiadis and J. D. Reiss, "An interlanguage study of musical timbre semantic dimensions and their acoustic correlates," *Music Percept.*, **31**, 339–358 (2014).

[6] C. Saitis, C. Fritz, G. P. Scavone, C. Guastavino and D. Dubois, "Perceptual evaluation of violins: A psycholinguistic analysis of preference verbal descriptions by experienced musicians," *J. Acoust. Soc. Am.*, **141**, 2746–2757 (2017).

[7] C. Saitis and S. Weinzierl, "Concepts of timbre emerging from musician linguistic expressions," *J. Acoust. Soc. Am.*, **141**, 3799 (2017).

[8] C. Saitis, H. Järveläinen and C. Fritz, "The role of haptic cues in musical instrument quality perception," in *Musical Haptics*, S. Papetti and C. Saitis, Eds. (Springer, Cham, 2018), pp. 73–93.

[9] V. U. Ludwig and J. Simner, "What colour does that feel? Tactile-visual mapping and the development of cross-modality," *Cortex*, **49**, 1089–1099 (2013).

[10] R. D. Melara and L. E. Marks, "Interaction among auditory dimensions: Timbre, pitch, and loudness," *Percept. Psychophys.*, **48**, 169–178 (1990).

[11] P. Walker, "Cross-sensory correspondences: A theoretical framework and their relevance to music," *Psychomusicology*, **26**, 103–116 (2016).

[12] L. E. Marks, "On colored-hearing synesthesia: Cross-modal translations of sensory dimensions," *Psychol. Bull.*, **82**, 303 (1975).

[13] J. Ward, B. Huckstep and E. Tsakanikos, "Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?" *Cortex*, **42**, 264–280 (2006).

[14] A.-S. Crisinel and C. Spence, "As bitter as a trombone: Synesthetic correspondences in nonsynesthetes between tastes/flavors and musical notes," *Atten. Percept. Psychophys.*, **72**, 1994–2002 (2010).

[15] J. Simner, C. Cuskley and S. Kirby, "What sound does that taste? Cross-modal mappings across gustation and audition," *Perception*, **39**, 553–569 (2010).

[16] M. Adeli, J. Rouat and S. Molotchnikoff, "Audiovisual correspondence between musical timbre and visual shapes," *Front. Hum. Neurosci.*, **8**, 352 (2014).

[17] C. Reuter, J. Jewanski, C. Saitis, I. Czedik-Eysenberg, S. Siddiq and M. Oehler, "Colors and timbres — consistent color-timbre mappings at non-synesthetic individuals," *Proc. 34th*

Jahrestagung der Deutschen Gesellschaft für Musikpsychologie: Musik im audiovisuellen Kontext, Gießen, Germany (2018).

[18] C. Cuskley, M. Dingemanse, S. Kirby and T. M. van Leeuwen, "Cross-modal associations and synaesthesia: Categorical perception and structure in vowel-colour mappings in a large online sample," *Behav. Res. Methods*, no. published online (2019).

[19] S. Sadaghiani, J. X. Maier and U. Noppeney, "Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing," *J. Neurosci.*, **29**, 6490–6499 (2009).

[20] N. Bien, S. ten Oever, R. Goebel and A. T. Sack, "The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study," *NeuroImage*, **59**, 663–672 (2012).

[21] K. McCormick, S. Lacey, R. Stilla, L. C. Nygaard and K. Sathian, "Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation," *Neuropsychologia*, **112**, 19–30 (2018).

[22] P. Walker, J. G. Bremner, U. Mason, J. Spring, K. Mattock, A. Slater and S. P. Johnson, "Preverbal infants' sensitivity to synaesthetic cross-modality correspondences," *Psychol. Sci.*, **21**, 21–25 (2010).

[23] S. Dolscheid, S. Hunnius, D. Casasanto and A. Majid, "Prelinguistic infants are sensitive to space-pitch associations found across cultures," *Psychol. Sci.*, **25**, 1256–1261 (2014).

[24] C. V. Parise, K. Knorre and M. O. Ernst, "Natural auditory scene statistics shapes human spatial hearing," *Proc. Natl. Acad. Sci. USA*, **111**, 6104–6108 (2014).

[25] Y. Jamal, S. Lacey, L. Nygaard and K. Sathian, "Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception," *Multisens. Res.*, **30**, 287–306 (2017).

[26] C. Spence, "Crossmodal correspondences: A tutorial review," *Atten. Percept. Psychophys.*, **73**, 971–995 (2011).

[27] G. Martino and L. E. Marks, "Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis," *Perception*, **28**, 903–923 (1999).

[28] M. Rakova, *The Extent of the Literal: Metaphor, Polysemy and Theories of Concepts* (Palgrave Macmillan, New York, 2003).

[29] V. Walsh, "Magnitudes, metaphors, and modalities: A theory of magnitude revisited," *Oxford Handbook of Synesthesia* (Oxford University Press, Oxford, 2013), pp. 1–20.

[30] C. Cuskley, J. Simner and S. Kirby, "Phonological and orthographic influences in the bouba–kiki effect," *Psychol. Res.*, **81**, 119–130 (2017).

[31] S. Reiterer, M. Erb, W. Grodd and D. Wildgruber, "Cerebral processing of timbre and loudness: fMRI evidence for a contribution of broca's area to basic auditory discrimination," *Brain Imaging Behav.*, **2**, 1–10 (2008).

[32] K. von Kriegstein, D. R. Smith, R. D. Patterson, D. T. Ives and T. D. Griffiths, "Neural representation of auditory size in the human voice and in sounds from other resonant sources," *Curr. Biol.*, **17**, 1123–1128 (2007).

[33] K. von Kriegstein and A.-L. Giraud, "Implicit multisensory associations influence voice recognition," *PLoS Biol.*, **4**(10), e326 (2006).

[34] Z. Wallmark, M. Iacoboni, C. Deblieck and R. A. Kendall, "Embodied listening and timbre: Perceptual, acoustical, and neural correlates," *Music Percept.*, **35**, 332–363 (2018).

[35] G. A. Calvert, "Crossmodal processing in the human brain: Insights from functional neuroimaging studies," *Cereb. Cortex*, **11**, 1110–1123 (2001).

[36] J. R. Binder and R. H. Desai, "The neurobiology of semantic memory," *Trends Cogn. Sci.*, **15**, 527–536 (2011).

[37] M. Kiefer and F. Pulvermüller, "Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions," *Cortex*, **48**, 805–825 (2012).

[38] E. Guzman-Martinez, L. Ortega, M. Grabowecky, J. Mossbridge and S. Suzuki, "Interactive coding of visual spatial frequency and auditory amplitude-modulation rate," *Curr. Biol.*, **22**, 383–388 (2012).

[39] C. V. Parise, "Crossmodal correspondences: Standing issues and experimental guidelines," *Multisens. Res.*, **29**, 7–28 (2016).

[40] C. Saitis and K. Siedenburg, "Exploring the role of source-cause categories in timbral brightness perception," *Proc. Timbre 2018: Timbre is a Many-Splendored Thing*, Montreal, Canada, pp. 79–80 (2018).

[41] D. Schön, S. Ystad, R. Kronland-Martinet and M. Besson, "The evocative power of sounds: Conceptual priming between words and nonverbal sounds," *J. Cogn. Neurosci.*, **22**, 1026–1035 (2009).

[42] T. Grill, *Perceptually Informed Organization of Textural Sounds*, PhD thesis, University of Music and Performing Arts Graz, Graz, Austria (2012).

[43] S.-A. Lembke, "Hearing triangles: Perceptual clarity, opacity, and symmetry of spectrotemporal sound shapes," *J. Acoust. Soc. Am.*, **144**, 608–619 (2018).

[44] S. Okamoto, S. Ishikawa, H. Nagano and Y. Yamada, "Spectrum-based synthesis of vibrotactile stimuli: Active footstep display for crinkle of fragile structures," *Virtual Real.*, **17**, 181–191 (2013).

[45] S. Bonnans and A. Noble, "Effect of sweetener type and of sweetener and acid levels on temporal perception of sweetness, sourness and fruitiness," *Chem. Senses*, **18**, 273–283 (1993).

[46] C. A. Mueller, K. Pintscher and B. Renner, "Clinical test of gustatory function including umami taste," *Ann. Otol. Rhinol. Laryngol.*, **120**, 358–362 (2011).