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

Charalampos Saitis, Stefan Weinzierl, Katharina von Kriegstein, Sølvi Ystad

Acoust. Sci. & Tech.

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] and Sølvi Ystad[4]

[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] *Aix Marseille Université, CNRS, PRISM (Perception, Representations, Image, Sound and Music), 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France*

Keywords: Timbre semantics, crossmodal correspondences, conceptual metaphors

PACS number: 43.66.Jh

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 visual brightness, tactile roughness, or gustatory sweetness, can offer a new way of addressing old questions about the perceptual and neurocognitive mechanisms of timbre 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 human auditory cognition 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 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). These metaphorical linguistic structures are central to the process of conceptualizing timbre by allowing listeners to communicate subtle acoustic variations in terms of other, more commonly shared sensory experiences (nonauditory or auditory-onomatopoeic) and abstract conceptions.

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. 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 [2]. 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 brightness, roughness, and fullness [3–5]. The first two dimensions appear to be associated with spectral energy distribution and fine spectrotemporal modulations, respectively, while the third refers to impressions of overall spectral content.

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 [6]. Such analyses have provided additional insight regarding particular factors that contribute to the salient semantic dimensions of timbre (e.g., [7, 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 grow-

* e-mail: charalampos.saitis@campus.tu-berlin.de

ing body of studies documenting the behavior of pitch-based associations (e.g., pitch-height and -brightness; see [11] for a review) but similar research on timbre is still very limited [12–15]. In addition, there are currently very few published neuroscientific studies explicitly looking at auditory-nonauditory correspondences [16–18].

Observing certain crossmodal mappings in preverbal infants [19, 20] 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 [21, 22]. 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.

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 [23]. 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 [24]. For instance, describing a sound as bright may be rooted in a supramodal concept of brightness rather than visual brightness per se. In other words, the quality of bright may not be anchored in the visual modality, but in a supramodal representation responsive to certain stimulus features regardless of modal content (cf. [25]).

Neuroimaging data studied across a variety of semantic tasks, including crossmodal correspondences, demonstrates 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 where perceptual information streams from different modalities are known to converge (auditory-visual correspondences [16–18]; auditory brightness [26]; auditory size [27]; voice recognition [28]; general conceptual processing [29–31]). These include zones within the inferior parietal lobe, the lateral and ventral temporal cortex, and the prefrontal cortex, among others. 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 [32].

3. A 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 same 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, an extensive examination of auditory-nonauditory correspondences is needed, including amassing 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 [14]. 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., [33]). 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 [34]. 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 [35].

A systematic investigation of crossmodal correspondences between timbre and nonauditory perceptual dimensions therefore necessitates auditory stimuli that can be manipulated along intrinsic continuous dimensions of timbre [36], and nonlinguistic nonauditory stimuli designed along perceptually gradient scales to facilitate the matching of auditory-nonauditory sensory experiences that may evoke the same concepts (e.g., [9,14]). Leveraging advancements in sound synthesis and morphing, visual signal processing, haptic displays, and virtual reality, such research can bring a new perspective into understanding the sensations of sound.

References

[1] S. McAdams, "Musical timbre perception," in *The Psychology of Music*, D. Deutsch, Ed. (Academic Press, London, 2013), pp. 35–67.

[2] C. E. Osgood, "The nature and measurement of meaning," *Psychological bulletin*, **49**, 197–237 (1952).

[3] W. H. Lichte, "Attributes of complex tones.," *J. Exp. Psychol.*, **28**, 455–480 (1941).

[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 inter-language 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] V. Rioux and D. Västfjäll, "Analyses of verbal descriptions of the sound quality of a flue organ pipe," *Musicae Scientiae*, **5**, 55–82 (2001).

[8] C. Traube, *An Interdisciplinary Study of the Timbre of the Classical Guitar.* (PhD thesis, Dept. of Music Research, McGill University, Montreal, Canada, 2004).

[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] 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).

[13] 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).

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

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

[16] 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).

[17] 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).

[18] 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).

[19] 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).

[20] 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).

[21] O. Deroy and C. Spence, "Are we all born synaesthetic? examining the neonatal synaesthesia hypothesis," *Neurosci. Biobehav. Rev.*, **37**, 1240–1253 (2013).

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

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

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

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

[26] 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).

[27] 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).

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

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

[30] J. R. Binder, R. H. Desai, W. W. Graves and L. L. Conant, "Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies," *Cereb. Cortex*, **19**, 2767–2796 (2009).

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

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

[33] 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).

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

[35] 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*, 79–80 (2018).

[36] A. Merer, M. Aramaki, S. Ystad and R. Kronland-Martinet, "Perceptual characterization of motion evoked by sounds for synthesis control purposes," *ACM Trans. Appl. Percept.*, **10**, 1 (2013).