More than any other musical parameter, the understanding of musical timbre requires the application of sophisticated audio analysis tools, so-called “audio features”, which help to specify the acoustic properties underlying the percept of timbre. Yet, a curious divide characterizes the usage of audio features for modeling musical timbre in the fields of Music Information Research (MIR) and Music Psychology (MP). While MIR usually makes use of a broad variety of features for timbre-related tasks such as instrument classification, MP only employs a highly constrained set of features to capture the perceptual dimensions of timbre. As outlined in Aucouturier & Bigand (J Intell Inf Syst (2013) 41:483–497), this divergence leads to substantial misunderstandings.

Here, we will characterize the two fields’ basic methodologies in more detail, their epistemic framework, and how they interpret the task of modeling timbre. Data on stimuli from classical perception studies will be presented that may clarify the discrepancy around the description of spectral envelopes, often based on 13–25 Mel–frequency cepstral coefficients in MIR, while only using the spectral centroid in MP. We will argue that Aucouturier & Bigand ignored methodological problems inherent in studies of computational neuroscience, cited as supposedly methodologically sound alternative to approaches in MIR.

Our main point of arrival will be that the usage of the technologies of timbre, as well as the broader discourse around the computational modeling of music requires explicit negotiations about the validity of different evaluation criteria. There is further need of an interdisciplinary agreement on the applicability of different forms of scientific parsimony.