Congruency of fundamental frequency and spectral envelope statistics in auditory scene analysis

Kai Siedenburg[*1] and Simon Jacobsen[1]

[1]University of Oldenburg – Germany

Abstract

Fundamental frequency (F0) and spectral envelope (SE) are critical acoustical features of harmonic sounds and centrally affect auditory scene analysis (ASA). F0 and SE are also considered key acoustical correlates of pitch and timbre perception, respectively. However, distinctions between F0- and SE-based effects in auditory perception are hardly clear-cut. In acoustical analyses, it has been shown that there is substantial covariance between F0 and SE properties for a large number of musical instrument sounds. Accordingly, perceptual studies have shown strong interference effects between pitch and auditory brightness perception. Yet, most perceptual studies have used ad-hoc artificial sounds, which do not respect the statistics of natural sounds (some SE profiles go together with some F0s). It thus remains unclear how exactly F0 and SE interplay and affect ASA under realistic acoustical conditions. In the present work, we use an analysis-by-synthesis approach to measure aspects of sound pleasantness, auditory brightness perception, and interleaved melody recognition for assessing the role of congruency between F0 and SE statistics in ASA.

We constructed a model of the sound envelope with 13 lower-order cepstral coefficients derived from sounds' excitation patterns. A set of around 1,900 recorded harmonic tones from 50 different sustained orchestral instruments was analyzed. Principal component analysis (PCA) was applied on the matrix of cepstral coefficients and the first two PCA components explained 62 % of the variance with coefficients 2-13 (leaving out the DC components). Only considering the first two components mapped out spectral similarities in the sound set, where clusters could mainly be explained by instrument family membership. Some instruments (e.g., from the clarinet or flute classes) showed distinct F0-dependent trajectories in the PCA space, potentially corresponding to different F0-registers. For the subsequent experiments, points in the space were resynthesized by choosing F0s that co-occurred (congruent condition) or did not co-occur (incongruent condition) with the chosen SE points, alongside a condition with fixed F0.

In Exp. 1, listeners rate sounds generated in the space according to their pleasantness. Preliminary data suggests that sounds with congruent F0 and SE statistics are rated as more pleasant compared to sounds with incongruent statistics and that sounds from outer regions of the space (where no actual instrument sound is located) are rated as less pleasant compared to the inner space. In Exp. 2, listeners rate sounds along the first two components of the PCA according to their auditory brightness. Based on informal listening, we hypothesize that listeners perceive the first but not the second dimension of the PCA space to be closely related to brightness. The latter would suggest that auditory brightness may be a sound attribute that directly stems from the statistics of acoustical instrument sounds. In Exp. 3, an interleaved melody recognition paradigm is used to test the role of congruency in ASA. We expect higher recognition scores for the congruent condition compared to the incongruent condition, which would underline the importance of congruency of acoustical cues in ASA.

---

[*]Speaker