Sound-gesture identification in real-world sounds of varying timbral complexity

Sven-Amin Lembke
Music, Technology and Innovation - Institute for Sonic Creativity (MTI),
De Montfort University, Leicester, United Kingdom
sven-amin.lembke@dmu.ac.uk

Introduction

Real-world sounds commonly allow the identification of their physical source or cause, whereas they also convey qualitative timbral features. Source-related *categorical* properties of timbre can detract listeners' evaluation of timbral *qualities* (e.g., Lemaitre et al., 2010), where it can be assumed that both aspects of timbre may compete for listeners' attention.

In electroacoustic music, the acousmatic tradition and also the theory of spectromorphology (Smalley, 1997) place greater importance on the intrinsic sound qualities, based on which morphologies like *sound gestures* or *textures* can arise. A sound gesture concerns a process in which one or more spectral properties (e.g., spectral frequency or amplitude) vary over time, bearing either direct or metaphorical associations to underlying spatio-kinetic action(s) of gestural kind.

Spectromorphology also considers the notion of *source bonding* (Smalley, 1997), which acknowledges listeners' natural tendency to focus on extrinsic links, like the source or cause when they are identifiable, which in turn might impair the perception of the more important intrinsic features. The current study investigated the role of source bonding on the perception of sound gestures based on whether the source or cause could be identified and also across range of real-world sounds of varying timbral complexity.

Method

Twenty listeners (age range: 18-65, 12 female, 8 male) were asked to identify sound gestures inherent in real-world sounds. Gestures concerned variations along the auditory parameters timbral brightness or loudness. The latter varied along the temporal amplitude envelope (e.g., tennis ball bouncing, metal coin spinning). In many real-world sounds, loudness variation may also affect timbre. Gestures along timbral brightness varied along frequency trajectories, which could concern tonal or filtered noise components in the original real-world sounds (e.g., vacuum cleaner turned on and off, filling a water carafe). Although the underlying spectral features may have affected both timbral brightness and a more abstract notion of pitch, in the context of gestural contours, both can be assumed equivalent (e.g., McDermott et al., 2008).

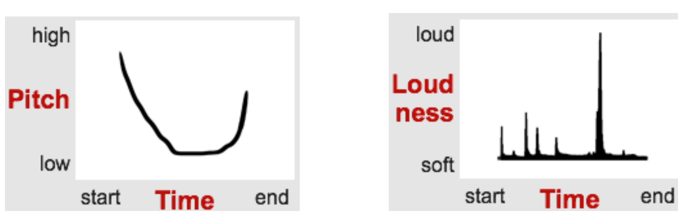

Figure 1: Examples of visual sketches based on which brightness gestures (left, here labeled pitch) or loudness gestures (right) gestures were identified.

For each parameter, participants evaluated 14 different sounds to 1) identify the embedded sound gesture and 2) also attempt to identify the underlying real-world source or cause by both providing verbal descriptions for source and cause and rating their confidence on this source/cause identity. The identification of sound gestures was achieved by selecting one correct visual sketch out of four options. As illustrated in Figure 1, the sketches were visual analogues of the underlying gestural brightness (labeled "pitch" for greater clarity to untrained listeners) or amplitude features extracted from the real-world sounds. Incorrect visual-sketch options concerned reversals or inversions of the correct gesture and/or alternatives from similar sounds.

Furthermore, gesture-identification accuracy was studied by taking the role of source bonding and repeated listening into account. To remove source bonds while retaining gestural cues, gestures were resynthesised as bandpass-filtered noise. For brightness, the filter's centre frequency followed the gestural shape, whereas for loudness, the gesture concerned variations of the amplitude envelope. Participants listened to the same gestures three times and in the following order: I. as the *original* real-world sounds, II. as the *noise*-based variants, and III. as a repeated presentation of the original. This III. presentation, however, either involved the *original* sound or a *hybrid* between the original and noise-based sounds.

Results

As shown in Figure 2, brightness (left) and loudness (left) gestures yielded 50% and more correct identifications for half the gestures investigated, suggesting that gestures can indeed be perceived in real-world sounds. Compared to gestures in original sounds (I.), however, noise-based gestures (II.) appeared to be identified more accurately, as the entire distribution of 14 gestures yielded 5-10% higher identification. Overall, these trends suggest that gestures presented in isolation were identified more reliably, i.e., when the source or cause was less identifiable (see relatively lower confidence for II. in Figure 3). As the noise-based gestures (II.) also represented a repeated listening, however, familiarity with the gesture could have also factored in. If repeated listening alone increased identification accuracy, then the identification rate in condition III. would have surpassed both previous presentations in I. and II. Indeed, the III. presentation of *hybrid* gestures appeared to increase identification rate, whereas the repetition of *original* sounds exhibited a lower median identification rate compared to the noise-based sounds in II. In sum, although gesture identification appeared to be aided by listening repetition, identification improved further when gestural cues were readily apparent, supported by the greater salience of gestural cues in hybrid sounds compared to original sounds.

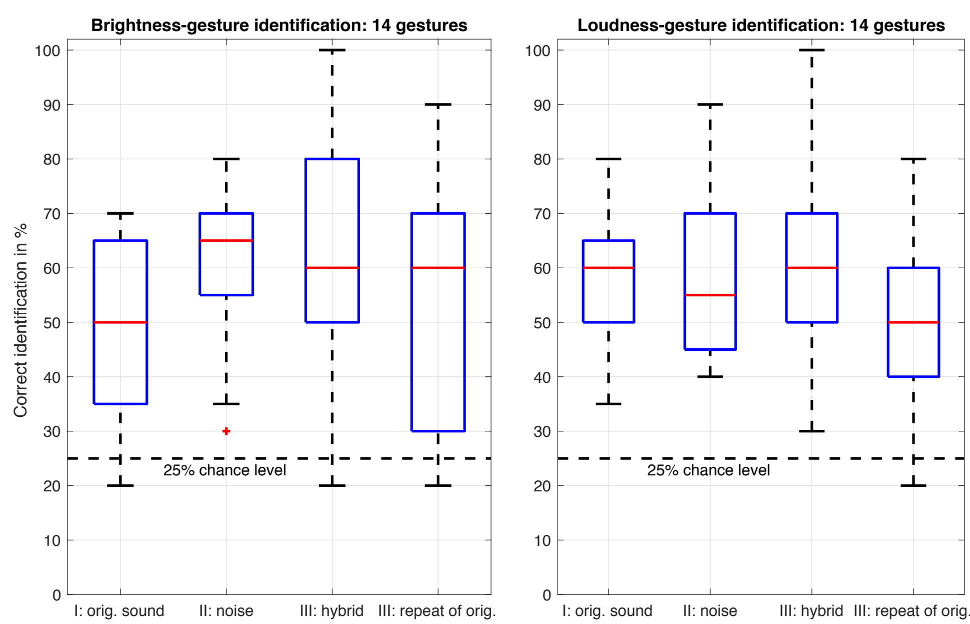

Figure 2: Identification accuracy of 14 brightness (left) and 14 loudness gestures (right).

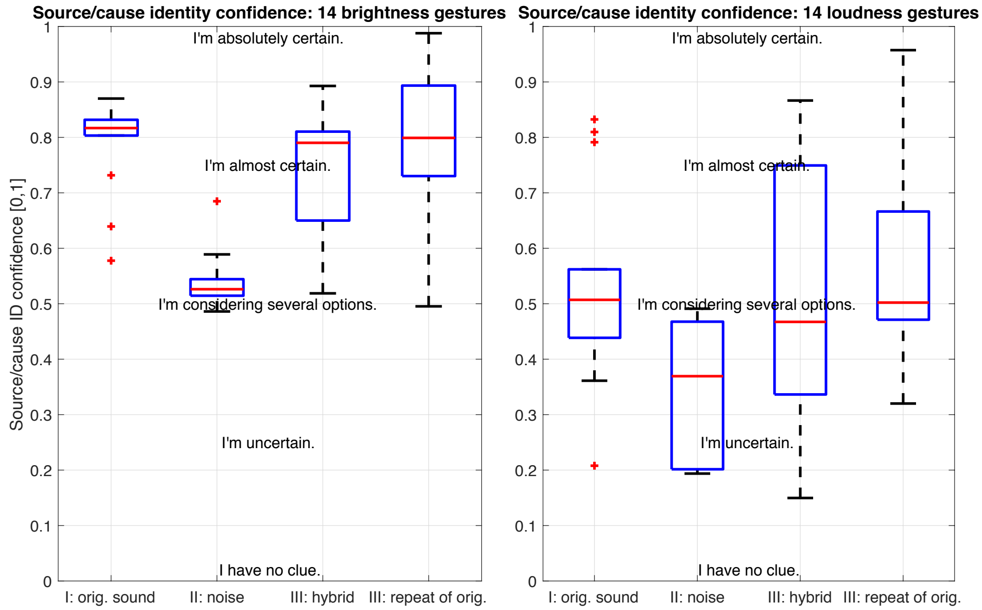

Figure 3: Ratings of confidence on source/cause identity of 14 brightness (left) and 14 loudness gestures (right).

Across all gestures, a wider variation of identification accuracy still became apparent and may relate to a range of factors, which will be studied using partial least-squares regression in time for the conference. A wider set of predictors will be considered, spanning source/cause confidence, timbre descriptors, and variables related to gestural features (e.g., shape, orientation, duration).

Discussion

The perception of time-variant timbral features like sound gestures relates to theories of embodied perception (Godøy, 2006), which presumably could extend to the perception of conventional instruments (e.g., note articulations involving timbral brightness contours). With regard to electroacoustic music, the central role of timbre has motivated the review of established distinctions between parameters (e.g., “the pitch *within* timbre”, Smalley, 1994, p. 40), while this broader notion of timbre may also benefit the study of other, more timbre-reliant musical practices (e.g., popular, non-Western).

References

Godøy, R. I. (2006). Gestural-Sonorous Objects: embodied extensions of Schaeffer’s conceptual apparatus. *Organised Sound*, *11*(2), 149–157.

Lemaitre, G., Houix, O., Misdariis, N., & Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, *16*(1), 16–32.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*(12), 1263–1271.

Smalley, D. (1994). Defining timbre - Refining timbre. *Contemporary Music Review*, *10*(2), 35–48.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound*, *2*(2), 107–126.