Perceptual factors contributing to the identity of sound gestures

Sven-Amin Lembke
Music, Technology and Innovation – Institute for Sonic Creativity (MTI^2),
De Montfort University, Leicester
sven-amin.lembke@dmu.ac.uk

Abstract

Sound gestures assume an important role in electroacoustic music (Smalley 1997). Studies on sound-to-gesture relations have revealed valuable insights, but still more fundamental research related to electroacoustic practice is necessary and was here addressed by two perceptual experiments. Using the meso and micro timescales proposed by Gødoy (2006), two questions were explored: 1) what information does a gesture convey at the meso level, and 2) how do signal features at the micro level affect gestural properties?

Experimental findings illustrate how meso-level gestural identity depends on features encoded in the micro-level signal structure. For one, important properties of gestural shape and curvature were informed by the physical scales underlying micro-level processes. How gestures evolve over time can be argued to represent aesthetic aspects in their own right, which go beyond simple gestural categories like 'upward' or 'downward'. Furthermore, micro structure contains signal features that relate to both gestural and other potential signifiers. Whereas clear source bonds (Smalley 1997) do not rule out gestural perception, the experimental findings still suggest that such strong bonds can detract from identifying gestures in real-world sounds.

1. Introduction

In the conception and experience of electroacoustic music, the notion of the gesture assumes an important role, especially when drawing on the theory of spectromorphology (Smalley 1997). Gestural analogies to sound have also been widely discussed for music in general, often in support of theories of embodied cognition (Jensenius et al. 2010), where the underlying relationships between sound and extramusical dimensions further relate to research on cross-modal correspondences (Spence 2011). Although most music-related research has focused on performance gestures in instrumental practice and their underlying relationships, the special case of metaphorical gestures has been acknowledged (Jensenius et al. 2010). Like other gestures, they convey some form of information, and their communicative aim is often intentional. Notably, it can be argued that they only represent gestures when perceived as such.

Engrained in the audio signal, such sonic gestures (Van Nort 2009; Jensenius et al. 2010) or gestural sonorous objects (Godøy 2006) bear heavily on electroacoustic practice. In the latter, the proposed co-reliance of core Schaefferian concepts and the embodied realities of listening

provide an interesting frame for the study of perception and cognition in electroacoustic music. Studies on sound-to-gesture relations in visual or motion cues have indeed revealed valuable insights (e.g., Caramiaux, Bevilacqua, and Schnell 2010; Nymoen et al. 2012; Blackburn 2013; Lembke 2018), although more fundamental research related to electroacoustic practice and theories is necessary.

Godøy (2006) related the notion gestural objects to *meso* and *micro* timescales. Whereas gestural identities are only identified at the meso scale, their emergence depends on micro-level signal features, which motivates two modes of inquiry: 1) *what* information does a gesture convey at the meso level, and 2) *how* do signal features at the micro level affect gestural properties? This paper seeks to illustrate these modes using two recently conducted experiments.

2. Experimental findings

2.1. Experiment 1: Perception of sound-gesture shape

Conceivably the simplest of gestures, the uni-directional process undergoes a state change along a single trajectory. This change can apply to different auditory parameters, which here will be limited to pitch, loudness, and tempo.[1] As the most basic informational unit of the meso level, what is conveyed are pitch glides from low to high, loudness ramps from soft to loud, tempo changes from slow to fast or the respective inverse mappings. This categorical information fails to address how these changes unfold over time, by considering qualitative aspects related to morphology or shape, and in what ways the same change between states can be achieved in multiple ways.

Twenty participants evaluated differences in sound-gesture shape, as illustrated in Figure 1. Comparing three different gestures at a time, listeners related them to varying degrees of visual curvature (e.g., exponential, logarithmic or straight lines). All gestures were based on bandpass-filtered noise and explored wider ranges across frequency, sound level, and duration.

Figure 1: Graphical user interface (GUI) for Experiment 1, in which gesture shape was compared between three sound gestures and related to varying levels of curvature using a visual analogue.

[1] Arguably, these terms may seem out of place in the electroacoustic context. They were mainly used for the sake of clarity. Whereas the studied pitch variations using noise-based synthetic sounds would have similarly influenced timbral brightness, the former provides a more intuitive label to most listeners, especially those without specialised listening experience.

Pitch gestures involved upward or downward glides spanning two octaves by variation of the filter's centre frequency. Four acoustic scales were considered and concerned linear frequency in Hz, ERB rate (Moore and Glasberg 1983) as well as exponential and logarithmic scalings of frequency.

Loudness gestures concerned an increasing or decreasing loudness ramp spanning 18 dB. Four functions were considered: linear amplitude, linear sound level in dB, logarithmic amplitude, and squared-cosine amplitude (also known as *S-curve*). These represent standard control functions for amplitude fades in digital audio workstations (DAWs, e.g., Pro Tools, Reaper).

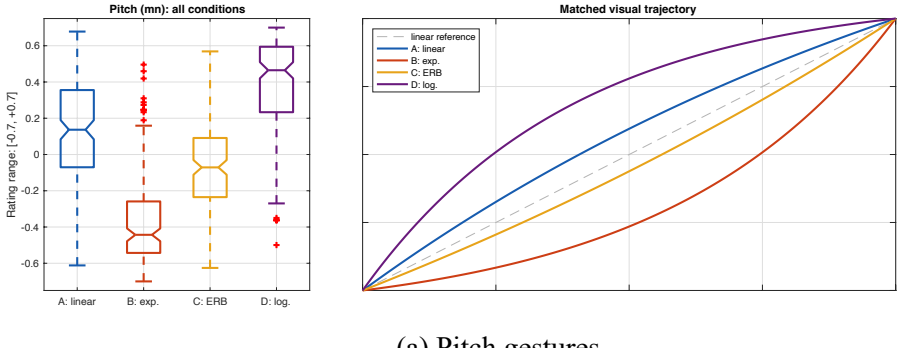
Tempo gestures concerned a sequence of 16 noise bursts whose tempo either accelerated or slowed down between beginning and end, by a factor of two. The functions were gathered from previous studies on tempo ritardandi, based on 1/q power expressions proposed by Friberg and Sundberg (1999) and the quadratic inter onset interval (IOI) by Repp (1992).

The clearest perceived differences relating to gestural shape were obtained for pitch gestures. Results are presented in Figure 2, comprising both boxplots[2] of the shape-related ratings and their corresponding visual analogues based on which listeners assigned the ratings. For pitch gestures (panel a), the ERB-rate scale matched the visual straight line most closely, whereas linear frequency scaling in Hz led to a slightly logarithmic curvature. The assumed 'linearity' of ERB rate agrees with psychoacoustic research on the logarithmic frequency scaling of the human auditory system, which corresponds to an exponential increase in Hz (Moore and Glasberg 1983). Only markedly stronger exponential scaling of frequency were matched to an analogous exponential line segment. Likewise, the logarithmic frequency scaling was matched to a similar curvature.

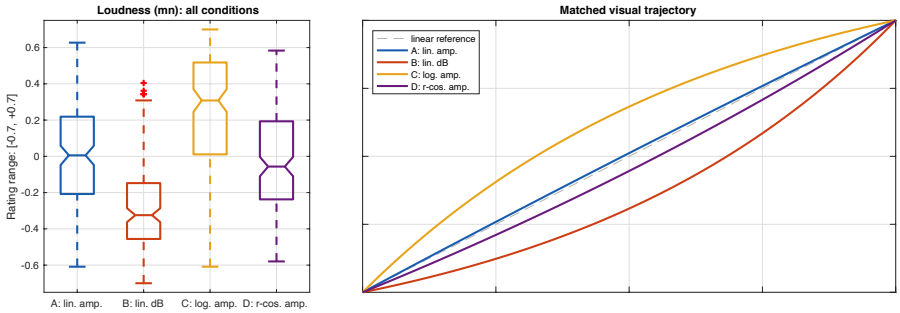
Loudness gestures (panel b) also yielded clear differences, in that gestures varying along linear amplitude were matched with a straight linear trajectory, whereas the raised-cosine amplitude (S curve) was perceived as slightly exponential. Somewhat surprisingly, a linear scaling along sound level in dB was matched to a markedly exponential trajectory. Whereas this finding could be specific to the context of the four scales studied, it is notable that the same four represent standard options for fades in DAW software.

Tempo gestures (panel c) exhibited the weakest and most ambiguous results. The curvature deviating most from a straight linear segment was obtained for the power coefficient $q = -1/2.5$. This function was considered as a logarithmic dependency to complement the remaining scales, which had been derived from perceptual research on ritardandi. Among the latter, however, differences were small. Still, the function for $q = 2.5$ approximated a straight line most closely and reflects a preferential region for power coefficients $q = 2–3$ that listeners have judged as most 'musical' (Friberg and Sundberg 1999).

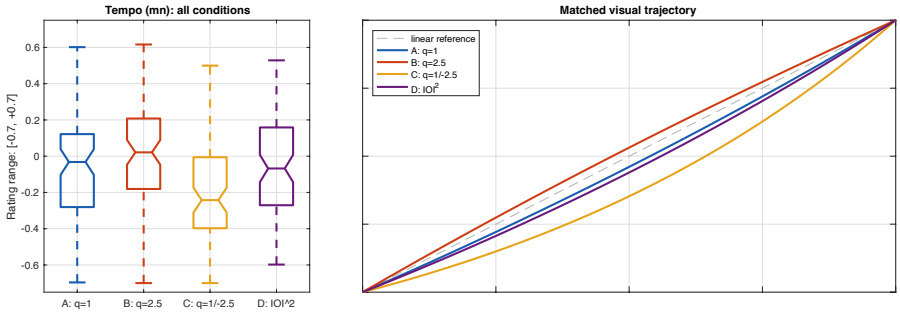
[2] Boxplots illustrate the distribution of *N* data points. The central horizontal line corresponds to the median, which splits the distribution into halves (50%). The box enclosing the median expands a quarter of the distribution above and below the median (75% and 25%). The thin whiskers reaching beyond the box encompass the entire distribution of *N* data points (excl. outliers, crosses).



(a) Pitch gestures



(b) Loudness gestures



(c) Tempo gestures

Figure 2: Boxplots (left) for gesture-shape ratings ($N = 160$) and visual analogues of the line segments (right) based on the boxplot medians that reflect what participants would have seen. Notches represent 95% confidence intervals for the medians.

Overall, listeners were able to evaluate and distinguish between gestural shape for pitch and loudness quite reliably, illustrated by the wider range of curvatures covered. Tempo gestures, by contrast, did not evoke clear morphological differences for those scales studied.

2.2. Experiment 2: Identification of auditory gestures

Although perceived at the meso level, listeners' ability to identify gestures will depend on whether the underlying micro-level signal features provide a sufficient degree of gestural salience. This touches on the notion of source bonding (Smalley 1997), which can be hypothesised to emphasise listeners' focus on extrinsic links, such as the sound source or cause, while detracting from those intrinsic qualities that convey the gestural identity.

Twenty listeners were asked to identify sound gestures inherent in recordings of real-world sounds. Gestures again concerned pitch or loudness variations, related to either frequency trajectories or the temporal amplitude envelope, respectively.

The identification of sound gestures was achieved by selecting one correct visual sketch out of four options. The sketches were visual analogues of the underlying gestural pitch or amplitude features extracted from the real-world sounds, as illustrated in Figure 3. Incorrect visual-sketch options concerned reversals or inversions of the correct gesture and/or alternatives from similar sounds.

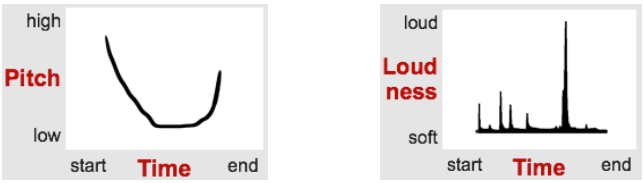


Figure 3: Examples of visual sketches for pitch (left) and loudness (right) gestures.

The identification rate of participants was studied by taking the role of source bonding and repeated listenings into account. In an attempt to remove source bonds while retaining gestural cues, the studied sound gestures were resynthesised as bandpass-filtered noise. For pitch gestures, the filter's centre frequency followed the pitch gestures; for loudness gestures, the amplitude envelope followed the loudness gesture. 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.

As shown in Figure 4, pitch and loudness gestures could be identified reliably, with 50% and more correct identifications for half the gestures investigated. This suggests that the sketches were understood as visual analogues of sound gestures and that the latter could indeed be perceived in real-world sounds. Notably, the identification rates for the most reliable quarter of sound gestures spanned from 60% to 100%, whereas only a small proportion of gestures was not identified above chance level.

Compared to gestures in original sounds (I.), noise-based gestures (II.) were identified more accurately, as most if not all gestures were identified above chance. For both pitch and loudness gestures, this trend applied to the entire range of gestures, leading to 5–10% higher identification. Overall, these trends suggest that gestures presented in isolation, i.e., without a readily identifiable source or cause, were identified more reliably.

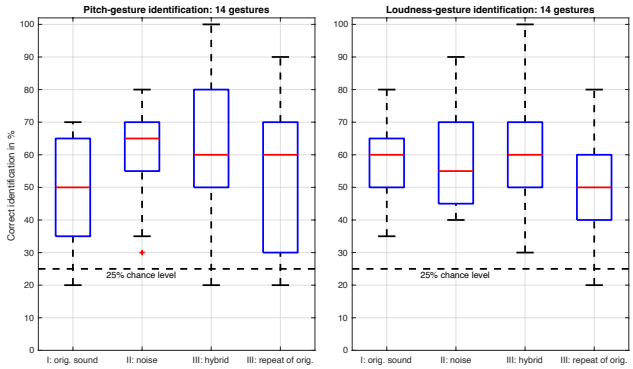


Figure 4: Boxplots for the correct identification of pitch and loudness gestures (left and right panels, respectively, $N = 14$). Participants identified the same gestures across three sound conditions, in the following order: I. *original* real-world sounds, II. bandpass-filtered *noise*, and III. either *hybrid* between original and noise or a repetition of the original. *Chance level* corresponds to the probability of all participants correctly identifying one out of four options at random.

Yet, the higher gestural identification for noise-based sounds could have resulted from greater familiarity with the gestures after already having evaluated the original sounds. This is where the III. presentation provides further insight: if repeated listening alone increased gesture identification, 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 overall, even reaching up to 100% correct identification for some gestures. The repetition of the original sounds, however, exhibited a lower median identification rate compared to the noise-based sounds in II. In sum, although listening repetitions seem to aid the identification of sound gestures, they are still more reliably perceived when gestural cues are readily apparent, further supported by the greater salience of gestural cues in hybrid sounds compared to original sounds.

3. Conclusion

Using the meso and micro timescales proposed by Godøy (2006) as a guide, the current experimental findings illustrate how meso-level gestural identity depends on features encoded in the micro-level signal structure. For one, important properties of gestural shape, e.g., resembling exponential vs. logarithmic curvature, are only informed by the scales underlying micro-level processes. How these gestures evolve over time can be argued to represent aesthetic aspects in their own right. What gestures convey therefore are no simple ‘upward’ or ‘downward’ trajectories. Rather, morphology may represent an integral part of the gestural identity.

On the other hand, the microstructure contains signal features that relate to both gestural and other potential signifiers. Whereas clear source bonds (Smalley 1997) do not rule out gestural perception, the current findings suggest that such strong bonds can detract from identifying gestures in real-world sounds. The degree to which source bonding may counteract gestural perception can further be assumed to vary across levels of listening expertise (Lemaitre et al. 2010).

Even without identifiable sources or causes, sounds may exhibit time-variant trajectories along multiple features, which each may convey gestures (Van Nort 2009). When these trajectories evolve independently, only the most salient feature may convey the meso gesture, e.g., pitch dominating over timbral brightness (Nymoen et al. 2012). In the future, similar studies focused on electroacoustic scenarios seem promising, which may inform creative practice through a better understanding of the perceptual and cognitive processes involved. From the interplay of gestures to the formation of sound shapes (Smalley 1997; Blackburn 2011), this would pave the way for studies on larger, macro-scale (Godøy 2006) musical contexts.

4. References

BLACKBURN, Manuella. 2011. "The Visual Sound-Shapes of Spectromorphology: An Illustrative Guide to Composition." *Organised Sound* 16 (01): 5–13.

BLACKBURN, Manuella. 2013. "Illustration and the Compositional Process: An Update on Pedagogical and Creative Uses." In *Proc. Tape to Typedef: Compositional Methods in Electroacoustic Music Symposium*. Sheffield.

CARAMIAUX, Baptiste, Frédéric BEVILACQUA, and Norbert SCHNELL. 2010. "Towards a Gesture-Sound Cross-Modal Analysis." In *Gesture in Embodied Communication and Human-Computer Interaction*, 158–70. Berlin: Springer.

FRIBERG, Anders, and Johan SUNDBERG. 1999. "Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners." *The Journal of the Acoustical Society of America* 105 (3): 1469–84.

GODØY, Rolf Inge. 2006. "Gestural-Sonorous Objects: Embodied Extensions of Schaeffer's Conceptual Apparatus." *Organised Sound* 11 (2): 149–57.

JENSENIUS, Alexander Refsum, Marcelo M. WANDERLEY, Rolf Inge GODØY, and Marc LEMAN. 2010. "Musical Gestures: Concepts and Methods in Research." In *Musical Gestures: Sound, Movement, and Meaning*, 12–35. New York: Routledge.

LEMAITRE, Guillaume, Olivier HOUIX, Nicolas MISDARIIS, and Patrick SUSINI. 2010. "Listener Expertise and Sound Identification Influence the Categorization of Environmental Sounds." *Journal of Experimental Psychology: Applied* 16 (1): 16–32.

LEMBKE, Sven-Amin. 2018. "Hearing Triangles: Perceptual Clarity, Opacity, and Symmetry of Spectrotemporal Sound Shapes." *The Journal of the Acoustical Society of America* 144 (2): 608–19.

MOORE, Brian C. J., and Brian R. GLASBERG. 1983. "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns." *The Journal of the Acoustical Society of America* 74 (3): 750–53.

NYMOEN, Kristian, Jim TORRESEN, Rolf Inge GODØY, and Alexander Refsum JENSENIUS. 2012. "A Statistical Approach to Analyzing Sound Tracings." In *Speech, Sound and Music Processing: Embracing Research in India*, edited by Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet, K. Jensen, and D. Mohanty, 7172:120–45. Berlin/Heidelberg: Springer.

REPP, Bruno H. 1992. “Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann’s ‘Träumerei’.” *The Journal of the Acoustical Society of America* 92 (5): 2546–68.

SMALLEY, Denis. 1997. “Spectromorphology: Explaining Sound-Shapes.” *Organised Sound* 2 (2): 107–26.

SPENCE, Charles. 2011. “Crossmodal Correspondences: A Tutorial Review.” *Attention, Perception & Psychophysics* 73 (4): 971–95.

VAN NORT, Doug. 2009. “Instrumental Listening: Sonic Gesture as Design Principle.” *Organised Sound* 14 (2): 177–87.