Hearing triangles: perceptual clarity, opacity, and symmetry of spectrotemporal sound shapes

Post-print version; the following article has been accepted by The Journal of the Acoustical Society of America. After it is published, it will be found at http://asa.scitation.org/journal/jas.

Sven-Amin Lembke[1]

Music, Technology and Innovation – Institute for Sonic Creativity (MTI^2),

De Montfort University, Leicester LE1 9BH, United Kingdom[a)]

(Dated: 31 July 2018)

In electroacoustic music, the spectromorphological approach commonly employs analogies to non-sonic phenomena like shapes, gestures, or textures. In acoustical terms, sound shapes can concern simple geometries on the spectrotemporal plane, for instance, a triangle that widens in frequency over time. To test the auditory relevance of such triangular sound shapes, two psychoacoustic experiments assessed if and how these shapes are perceived. Triangular sound-shape stimuli, created through granular synthesis, varied across the factors grain density, frequency and amplitude scales, and widening vs. narrowing orientations. The perceptual investigation focused on three auditory qualities, derived in analogy to the visual description of a triangle: the *clarity* of the triangular outline, the *opacity* of the area enclosed by the outline, and the *symmetry* along the vertical dimension. These morphological qualities seemed to capture distinct perceptual aspects, each linked to different acoustical factors. Clarity of shape was conveyed even for sparse grain densities, while also exhibiting a perceptual bias for widening orientations. Opacity varied as a function of grain texture, whereas symmetry strongly depended on frequency and amplitude scales. The perception of sound shapes could relate to common perceptual cross-modal correspondences and share the same principles of perceptual grouping with vision.

PACS numbers: 43.66.Ba, 43.66.Jh, 43.66.Lj, 43.66.Mk, 43.75.Cd, 43.75.Tv

[a] sven-amin.lembke@dmu.ac.uk

I. INTRODUCTION

Describing musical or acoustical parameters commonly borrows labels from other sensory modalities by employing metaphors or analogies. For instance, the association of pitch with spatial elevation (or the vertical dimension) finds a high, consistent prevalence across languages in that the labels "low" and "high" are used to describe opposite ends of the pitch continuum (Stumpf, 1883). This pitch-to-elevation correspondence has also been a widely studied in cross-modal perception between vision and audition (Evans and Treisman, 2010; Spence, 2011). Whereas the previous example is limited to a single dimension per sensory modality, audio-visual correspondences have similarly been discussed for multidimensional scenarios, such as two-dimensional shapes or gestalts (Köhler, 1947). A well-known example concerns the spoken sounds "maluma" vs. "takete" (Köhler, 1947) or "bouba" vs. "kiki" (Ramachandran and Hubbard, 2001) being consistently associated with corresponding rounded vs. jagged visual shapes, respectively, and subsequently found to apply to instrumental timbre as well (Adeli *et al.*, 2014). These findings bear the significance that clear correspondences can also exist for rather complex, multidimensional representations of stimuli in both the visual and auditory modalities. The current article concerns an exploration into psychoacoustic factors underlying the perception of two-dimensional geometric shapes projected onto the spectrotemporal plane, motivated by how these relate to the notion of *sound shapes* (Smalley, 1997) in electroacoustic music.

Previous findings for pitch-to-elevation correspondence may in fact have the shortcoming that they were studied using sine tones as opposed to complex sounds. In sine tones, pitch is indistinguishable from timbre, because the sinusoidal frequency serves as the sole perceptual cue for both auditory qualities. Notably, when pitch remains the same, even differences along spectral brightness can evoke correspondences to elevation: square waves exhibit brighter spectra than sine tones and were also linked to higher elevations than the latter (Parise and Spence, 2012). In a similar way, pitch and brightness contours can also be reliably associated with each other if both evolve along low-to-high continua (McDermott *et al.*, 2008). The association with spatial elevation could therefore be related to a general effect of frequency height, as it affects both perceived pitch, which often relates to only the fundamental, and perceived timbre, which (not exclusively) depends on all partials in the spectrum. Based on this reinterpretation, even the multidimensional design of spectrograms may have a cross-modal underpinning, as its vertical dimension conventionally reflects a low-to-high mapping of frequency to elevation. Some spectrotemporal evolutions may therefore correspond to visual shapes on the time-vs.-frequency plane.

As a common theoretical framework within the genre of electroacoustic music, *spectromorphology* (Smalley, 1997) deals with how spectra evolve and are shaped over time. The description of such spectromorphologies lends itself to employing analogies to extra-sonic phenomena, such as gestures, motion, growth, or texture. For instance, in visualizations of spectromorphological processes that replace the role of traditional music notation, simple geometric shapes are sometimes used (Blackburn, 2011; Smalley, 1997), employing the analogous notion of *sound shapes* that result from an interplay between sound *gestures* and *textures*. These visualizations commonly imply sound shapes to evolve on the spectrotemporal domain: the horizontal dimension represents time; the vertical axis describes the frequency spectrum, while spectral amplitude may only be vaguely specified. Acoustical assumptions are even more clearly implied when these geometric shapes are used as visual annotations, resembling or even superimposed onto spectrograms (e.g., *EAnalysis* software, Couprie, 2014). To the same literal extent, mapping visual shapes onto the spectrotemporal plane is also applied in computer interfaces for sound manipulation (e.g., *AudioSculpt*, IRCAM, 2013) or ones governing spectrotemporal synthesis (e.g., Xenakis' *UPIC* system).

Gesture and texture are understood as the two form-bearing principles of spectromorphology (Smalley, 1997), which for simple geometric shapes presumably involves texture being framed by gesture. Importantly, this concerns both the acoustical characteristics of the sound shape, i.e., related to how it occupies the spectrotemporal plane, and the evoked perceptual qualities. For the auditory perception of geometric shapes, the relevant morphological qualities remain largely unknown, also in terms of how they would represent gestural or textural properties. The association of these auditory qualities to acoustic factors likely relates to psychoacoustics. Furthermore, these perceptual qualities will also depend on auditory-grouping processes (Bregman, 1990), possibly sharing the same grouping principles that apply to visual shapes (e.g., *proximity, good continuation*, Wertheimer, 1923). Given that the discussion of spectromorphologies in musical works often employs analogies to extra-sonic phenomena, the intended auditory perceptions could inherently rely on common cross-modal correspondences (Spence, 2011), which could in fact concern rather literal morphological analogies between vision and audition.

This presents the point of departure for the current study, which focuses on possibly the simplest case of sound shapes: a triangle. Such a geometrical shape may delineate a spectrotemporal evolution in which two sides of a triangle diverge in frequency over time, as illustrated in Figure 1 (left and center panels). In terms of morphological attributes (right panel), the perceptual *clarity* of the shape's outline could be implied by the diverging sides' trajectories alone, but also the spectral content enclosed therein could bear some morphological significance, for instance, in terms of its transparency or *opacity*. Based on the notion of sound shapes resulting from gesture-framed texture (Smalley, 1997), clarity and opacity would concern gestural and textural properties, respectively, although, alternatively, texture could even be wholly unrelated to shape. Another morphological quality could concern the *symmetry* of the two diverging sides of the triangle relative to the point or frequency of origin, as either being perfectly balanced, titled upward or downward. Given this literal analogy of mapping a visual triangle onto the spectrotemporal domain, this study aims to investigate if and how this translates to analogous perceptions of clarity, opacity, and symmetry in the auditory realm.

A range of acoustic factors could influence these three auditory qualities. For instance, as the schematic triangle depicted in Figure 1 (right panel) exhibits linear sides, how would this linearity be best translated into the perceived sound shape? Human perception is known to favor logarithmic, relative dependencies for both frequency (e.g., Attneave and Olson, 1971; Moore and Glasberg, 1983; Stevens *et al.*, 1937) and amplitude (e.g., Fletcher and Munson, 1933). Thus, psychoacoustically derived scales or weightings for these two physical dimensions could be presumed more suited for conveying a perceptually more balanced or symmetric shape. On the other hand, many software applications' default settings offer linearly scaled frequency axes (e.g., *EAnalysis*, *AudioSculpt*), owing to the equal-spaced frequency resolution of the underlying FFT. Similarly, software interfaces often feature linear ramps, for instance, to dynamically control a filter's center or cutoff frequency. As this high prevalence of 'linear' settings in audio-production applications may have established certain listening habits, one should also consider whether they affect judgments on sound-shape symmetry.

Whereas the characteristics of the triangle's sides can be hypothesized to mainly influence the shape's clarity and symmetry, the degree of perceived opacity would probably concern the spectrotemporal content enclosed inside the outline. A granular representation of this content, i.e., with the shape composed of many individual sound grains, allows for a number of acoustic variables to be investigated, yielding spectrotemporal content that span sparse to seamless granular textures (e.g., Figure 1, left vs. center panel). For textures to be perceived as seamless or continuous, the granularity would need to lie below the detection thresholds for temporal gaps: while for noises (Moore, 2013) and constant-frequency sinusoids (Moore *et al.*, 1993) temporal gaps below 10 ms can be detected, the detection thresholds for temporal gaps involving a change in frequency typically fall between 10 and 20 ms for sinusoids (Smith *et al.*, 2006) and bandlimited noise (Phillips *et al.*, 1997). Thus, a sufficiently high granular density would ensure the perception of seamless as opposed to sparser, more granular textures, in line with what auditory grouping principles would predict (Bregman, 1990). At the same time, these varying degrees of granularity could be assumed to also affect sound perception as a whole, for example, if only textural properties were relevant.

Apart from granular density affecting the texture as a whole, the presence of a wider gap in the spectrum could also influence the perceived opacity. As narrower gaps may in fact remain inaudible due to spectral masking, such spectral gaps would need to exceed at least the equivalent-rectangular bandwidth (ERB, Moore and Glasberg, 1983) to become perceptible. Finally, the role of the temporal orientation of the triangular sound shape as either widening or narrowing in frequency across time (e.g., Figure 1, left vs. center panel) could also affect the perceived clarity, opacity, or symmetry, similar to how the time orientation of sounds with ramped amplitudes are known to affect perceived loudness differently (e.g., Neuhoff, 2001; Susini *et al.*, 2007).

Based on an exploratory approach, this diverse range of potentially relevant acoustic factors, which spanned all spectrotemporal dimensions, were investigated. The main aim was to establish general dependencies that described how and to what extent acoustic factors influenced the shape-related properties *clarity*, *opacity*, and *symmetry*. As sound shapes were expected to rely on both gestural or textural properties (Smalley, 1997), the perception along a non-morphological, purely textural dimension (*homogeneity*, Grill *et al.*, 2011) complemented the investigation to aid in distinguishing between gestural and textural contributions. The exploration involved multifactorial designs in two experiments, presented in Sections II and III, respectively, and followed by their joint discussion in Section IV.

II. EXPERIMENT 1

Experiment 1 explored the perceptual relevance of the morphological qualities *clarity*, *opacity*, and *symmetry* in face of two factors that characterized the temporal composition of triangular sound shapes . With these triangles composed of granular content, the *density* of sound grains served as the first factor under investigation. The second factor compared

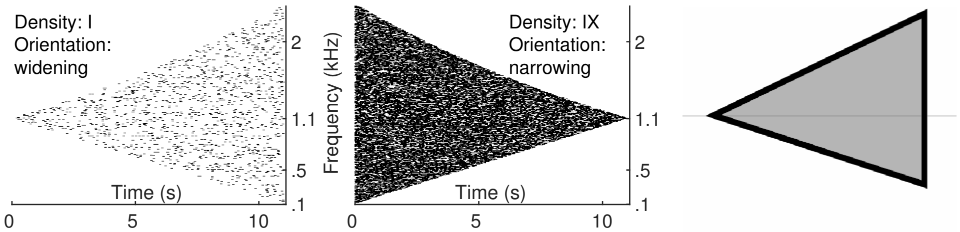

FIG. 1. The left and center panels display spectrograms of two triangular sound shapes composed out of sinusoidal grains. They correspond to Experiment 1's stimuli for lowest and highest grain density and widening and narrowing orientation, respectively. The right panel served as part of the graphical interface for participants, illustrating the morphological qualities: the black outline corresponds to *clarity*, the filled grey area to *opacity*, and the balance between the top and bottom ends of the triangle relative to the grey, horizontal axis to *symmetry*.

triangular sound shapes *orientation* as either widening or narrowing over time. Higher grain density was expected to influence both the clarity and opacity in that greater density could yield clearer and more solid sound shapes. As no frequency or amplitude aspects were manipulated here, this experiment allowed symmetry to be investigated for possible covariation with density or orientation, although no particular effect was anticipated beforehand. Shape orientation did also not entail a priori hypotheses, but its inclusion would allow the identification of potential perceptual asymmetries. A special interest lay in observing if and how differences between the morphological qualities would manifest themselves.

A. Method

a. Procedure. The experiment took place in a relatively absorbent sound-isolated booth (volume: 15.4 m^3, reverberation time: $T_{30} = 0.45$ s). The booth was primarily used as a

5.1-surround sound editing and mixing suite and, apart from the loudspeakers, was equipped with two computer flat screens, mouse, and keyboard, standing on a table situated in the center of the room. Participants faced the center loudspeaker on-axis at a distance of about 1.2 m. The experiment took around 60 minutes to complete.

During the experiment, participants were presented sound-shape stimuli that varied in their acoustic properties. In each experimental trial, a single sound shape stimulus was presented, and participants had to provide five responses through a computer interface. To characterize the perception of sound shapes, several perceptual qualities were considered and measured through continuous rating scales. As visualized in Figure 1 (right panel), these qualities were analogous to the visual description of a triangle, namely, the *clarity* of the defining triangular outline or contour (black), the *opacity* of the therein enclosed area (grey), and the *symmetry* of the shape relative to the triangle's tip (grey horizontal axis). The corresponding textual description for the rating scales was as follows:

- "How clearly is the shape outlined?", framed by the verbal anchors *faintly* to *clearly*, arranged left and right, respectively.
- "How transparent is the area inside the shape?", ranging from *transparent* to *solid*, again arranged horizontally.
- "How symmetric is the shape?", spanning from *tilted upwards* to *titled downwards*, arranged vertically from top to bottom, respectively.

In order to provide participants with a more intuitive sense of the rated qualities, the computer interface was interactive in that the visualized triangle dynamically adjusted the analogous clarity (line width of black outline), opacity (varying shades of grey), and symmetry (tilt relative to the horizontal axis) based on the current ratings.

In addition, a fourth rating was conducted on the overall impression of the sound's *homogeneity*: "How homogeneous is the overall sound?", involving the labels *heterogeneous* (left) and *homogeneous* (right). This measure was unrelated to shape and described a common textural property (Grill *et al.*, 2011), providing further insight into how texture and gesture contribute to sound shapes. Participants provided an additional response on identifying the orientation of the sound shape as either *becoming wider* or *becoming narrower* over time, which was exclusively used to monitor the proportion of correct classifications (96% across all stimuli and participants), serving as an indirect measure of participants' attention on the task.

b. Stimuli. All triangular sound shapes had a duration of 11 s and evolved along two frequency trajectories over time. As shown in Figure 1, a triangular sound shape could begin at the tip, centered on a single frequency, and widen toward its remaining two corners, the latter two spanning a bandwidth of frequencies and occurring at the same point in time. Conversely, a sound shape could begin at the wide end and narrow down toward the tip. Asynchronous granular synthesis composed the triangular sound shapes out of many individual 100-ms sinusoidal grains, each occurring at particular times and frequencies falling inside the triangular outline. For all individual grains, the amplitude exhibited ramped-cosine envelopes at the onsets and ends, with each taking up one third of the 100-ms grain duration.

The stochastic process governing the granular synthesis operated within certain constraints. In terms of frequency, the tip was always anchored at 1100 Hz; the trajectory toward lower frequencies followed linear frequency in Hz down to 100 Hz, while the upward trajectory followed ERB rate (equivalent-rectangular-bandwidth, Moore and Glasberg, 1983) up to 2434 Hz, spanning a maximum bandwidth of 2334 Hz.[1] As to time, the onsets of sinusoidal grains could occur anywhere along a time grid of 5 ms resolution, which lies below the lowest detection thresholds for temporal gaps (Moore, 2013). An iterative process created the granular sound-shape stimuli based on the above constraints, yielding higher grain densities with increasing iterations. Within these constraints, the onset times and frequencies were randomly assigned, while the amplitudes remained constant.

With regard to the investigated acoustic factors, sound shapes either widened or narrowed in frequency towards the end, with this difference in *orientation* representing the first of two independent variables (IVs). The second IV involved nine different levels of grain *density*. In sum, the two IVs resulted in a total of 18 experimental conditions (2×9). From an initial pool of 999 randomized iteration sequences, 72 sound shapes were selected as stimuli, classified into the nine distinct levels of grain density, each class represented by eight similar instances (9×8).

Grain density was quantified as the relative area of the triangular shape that was covered by grains, measured on a linearly scaled spectrotemporal reference grid (resolution: 5 ms time, 5 Hz frequency). Figure 2 shows the percentage of covered triangular area for all 72 sound shapes, already grouped into nine density levels (x-axis) each comprising eight instances. The graph illustrates the clear separation among all classes concerning the quan-

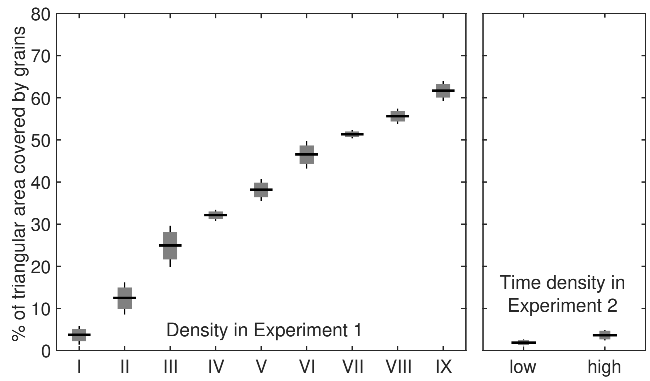

FIG. 2. Percentage of the triangular area covered by grains (y-axis) for the density levels investigated in Experiment 1 (left panel) and Experiment 2 (right panel). Box plots illustrate the distribution median (horizontal line), interquartile range (vertical box), and full range (thin vertical line) of all instances per class. For both experiments, each of the grain-density levels comprised eight instances.

tified percentage of grain density, overall, spanning a range from just below 5% to 60%. The two orientations (narrowing vs. widening) represented exact replica of the 72 conditions, i.e., each exemplar of the 72 sound shapes was replicated as a time-reversed copy. Overall, this yielded a total of 144 (72×2) experimental trials, presented in randomized order for each participant.

All sound-shape stimuli were equalized based on root-mean-square (RMS) amplitude and reproduced at equal gain. The sound stimuli were presented via a single Genelec *8040A* loudspeaker, representing the center speaker of the aforementioned 5.1-surround system. The listening level was on average 71 dB SPL at the wide side of the triangular shape, whereas the level at the tip was on average 61 dB SPL. An Avid *HD OMNI* audio interface processed the digital-to-analog conversion, based on the digital PCM format at 44.1 kHz sampling rate and 24-bit dynamic resolution.

c. Participants. 17 participants (15 male, two female) with a median age of 37 years (range: 19–54) completed the experiment. They had been recruited from the Music, Technology and Innovation community at De Montfort University, mainly represented by practitioners of electroacoustic music. In terms of musical expertise, participants exhibited a median of eight years of formal musical training, representing the maximum duration of training in any one of several musical subjects; 11 participants classified themselves as professional musicians. With regard to hearing deficiencies, one participant reported having tinnitus. Participation in the experiment involved informed consent, and the procedure had received prior approval by the Research Ethics Committee of De Montfort University. Participants were offered remuneration for their involvement, which some declined (mainly members of faculty).

B. Results

a. Data analysis. For clarity and opacity, the ratings spanned the values 0 to 1, corresponding to minimum and maximum clarity or opacity, respectively. Symmetry ratings were bi-polar: maximum symmetry represented the value 0; values of +1 and -1 corresponded to shapes being maximally tilted upwards or downwards, respectively. These rating measures served as dependent variables in three separate repeated-measures analyses of variance (ANOVA) with the two IVs *orientation* and *density*. In all cases, the within-subjects residuals across all experimental conditions did not indicate departures from normality (Shapiro-

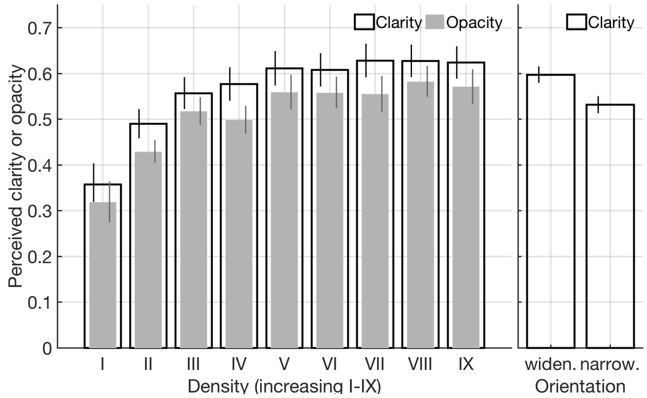

FIG. 3. Perceived clarity (black bars) and opacity (grey) for the nine density levels (left panel) and the widening vs. narrowing orientations (right panel) investigated in Experiment 1. Bars correspond to the group means of perceptual ratings, with the corresponding standard errors depicted in matching colors.

Wilk test). A criterion significance level of $\alpha = .05$ was assumed for all hypothesis tests. Where applicable, violations of sphericity (Mauchly's test) led to adjustments of the degrees of freedom based on the Greenhouse-Geisser correction (ε). Effect sizes concern generalized eta-squared η_G^2 (Bakeman, 2005) for ANOVA and Cohen's d for post-hoc t-tests.

b. Clarity measure. As shown in Figure 3 (left panel), increasing levels of grain density yielded corresponding gains in clarity ratings, $F(1.5, 24.3) = 7.3$, $\varepsilon = .19$, $p < .01$, $\eta_G^2 = .14$. Perceived clarity reached a plateau beyond level V, suggesting that shapes with greater grain density ceased to affect perceived clarity further. The two lowest density levels, I and II, evoked the largest perceived change in clarity. Interestingly, clarity was also perceived to be slightly higher for triangular sound shapes widening over time than for the reverse orientation, $F(1, 16) = 5.8$, $p = .03$, $\eta_G^2 = .02$, as illustrated in Figure 3 (right panel).

c. Opacity measure. As for clarity, also shown in Figure 3 (left panel), the ratings for opacity exhibited comparable gains with increasing grain density, $F(1.4, 21.8) = 6.3$, $\varepsilon = .17$, $p = .01$, $\eta^2_G = .14$. Again, ratings ceased to increase above density level V, and the perceived difference was most pronounced between the two lowest levels I and II. Unlike clarity, however, orientation of the sound shape did not appear to affect opacity.

d. Symmetry measure. No effects for symmetry ratings were observed, providing no indication that the chosen conditions for grain density and shape orientation affected perceived symmetry. Given the asymmetric use of scales for the upward and downward frequency trajectories, however, it should be noted that the global distribution of symmetry ratings ($N = 306$) across all conditions and participants was skewed. The median rating of 0.02 (lower quartile: -0.04, upper quartile: 0.21) was greater than zero, $z = 4.51, p < .01$ (Wilcoxon's signed-rank test), suggesting a slight asymmetric tilt upward and that ERB rate (upper trajectory) may have dominated over linear frequency in Hz (lower trajectory) in some participants' symmetry judgments.

e. Correlation among measures. Rank-correlation coefficients (Spearman's ρ) assessed the degree to which the shape-related measures exhibited similar rating profiles across conditions. Medians of participants' ratings across all experimental conditions ($N = 144$) were compared. As shown in the top-right half of the correlation matrix in Table I, the clarity and opacity ratings were moderately correlated, whereas correlations with symmetry ratings were either nearly absent for clarity or of opposite polarity for opacity. In addition, the non-morphological measure homogeneity exhibited clear correlations with clarity and opacity but hardly any with symmetry.

	clarity	opacity	symmetry	homogeneity	
clarity	—	.67	-.01	.84	←
opacity	.37	—	-.30	.74	Exp. 1
symmetry	.07	-.16	—	-.08	←
homogeneity	.37	.73	.28	—	
	↑	Exp. 2	↑		

TABLE I. Correlation matrix of averaged clarity, opacity, symmetry, and homogeneity ratings for Experiment 1 (top-right half, relative to diagonal), and Experiment 2 (bottom-left half). Rank correlations (Spearman's ρ) were computed across all experimental conditions.

III. EXPERIMENT 2

Experiment 2 explored a range of acoustic factors related to time, frequency, and amplitude that could influence perceived *clarity*, *opacity*, and *symmetry* of shape in different ways. Here, the investigation of *grain density* considered separate parametric variations along time and frequency. Clarity and opacity were expected to increase with greater grain density along both time and frequency, with the density oriented at levels that revealed the clearest perceptual differences in Experiment 1. However, the inclusion of additional factors was expected to also elucidate specificities for clarity and opacity. For instance, sound shapes exhibiting *spectral gaps* were expected to be perceived as more transparent, thus yielding lower opacity, while no similar effect was expected for clarity. Furthermore, differences between *frequency scales* and *amplitude weightings* explored their influence on a shape's symmetry, in which psychoacoustically derived functions were expected to yield differences to linear physical continua.

A. Method

Many aspects of the experimental procedure and stimulus presentation were the same for both experiments. Therefore, only differences to Experiment 1 are addressed in the following sections.

a. Procedure. Participants provided the same responses as in Experiment 1, except for the need to identify the orientation of the sound shape. For greater illustrative value, the aforementioned verbal anchors for the qualities clarity, opacity, and symmetry were complemented by the following additional labels *thin–bold*, *hollow–filled*, and *low–high*, respectively. The venue and technical setup for the experiment remained the same. The experiment took around 30 minutes to complete.

b. Stimuli. All sound shapes had a duration of 7 s, and only the widening orientation was considered. With regard to the frequency constraints delimiting the triangular shape, the tip was again anchored at 1100 Hz, while the opposite side exhibited a constant bandwidth of 2000 Hz over all conditions, as shown in Figure 4. The experimental design involved five IVs, namely, *time density*, *frequency density*, *frequency function*, *frequency fill*, and *amplitude weighting*. Each IV occurred at two treatment levels, resulting in 32 different conditions (2^5).

Unlike Experiment 1, the stochastic process governing the creation of sinusoidal grains involved separate parametric control over grain density in time and frequency, based on two reference vectors for each parameter. Two stages of random processes were used to generate the triangular composition of sinusoidal grains. First, randomized time vectors, i.e., a set of time values for the onsets of grains, were obtained from sampling a uniform distribution of time values without replacement. The time grid was based on a 5-ms resolution. Likewise, vectors of randomized frequencies falling within the maximal bandwidth were obtained by the same random-sampling technique, based on either linear frequency in Hz or ERB rate. The complete vector of frequencies corresponded to the maximum number of just-noticeable differences (JNDs) in frequency that the triangular bandwidth accommodated; the lowest known JND of 0.2% frequency deviation was used (Moore, 2013). As the second stage, the intersection of the triangular shape with the discretized grid of sampled frequencies and times yielded the spectrotemporal composition of grains. More specifically, for each point of the sampled time vector, a single element in the frequency vector was selected by uniform random sampling with replacement, and (only) if the frequency fell within the outline of the triangular shape, a grain was created at that frequency and time point.

Figure 4 provides representative examples for the five investigated IVs, compared to a reference condition displayed in the bottom-center panel. The above mentioned stochastic procedure was applied to implement two IVs based on varying density levels for both time and frequency, i.e., sampling time or frequency using either the complete vectors or only half the number of randomly selected values (top-left and bottom-left panels).

Another IV configured the two diverging trajectories of the triangle to follow frequency functions along either linear frequency in Hz or psychoacoustic ERB rate (equivalent-rectangular-bandwidth, Moore and Glasberg, 1983). Paired with the constant maximal

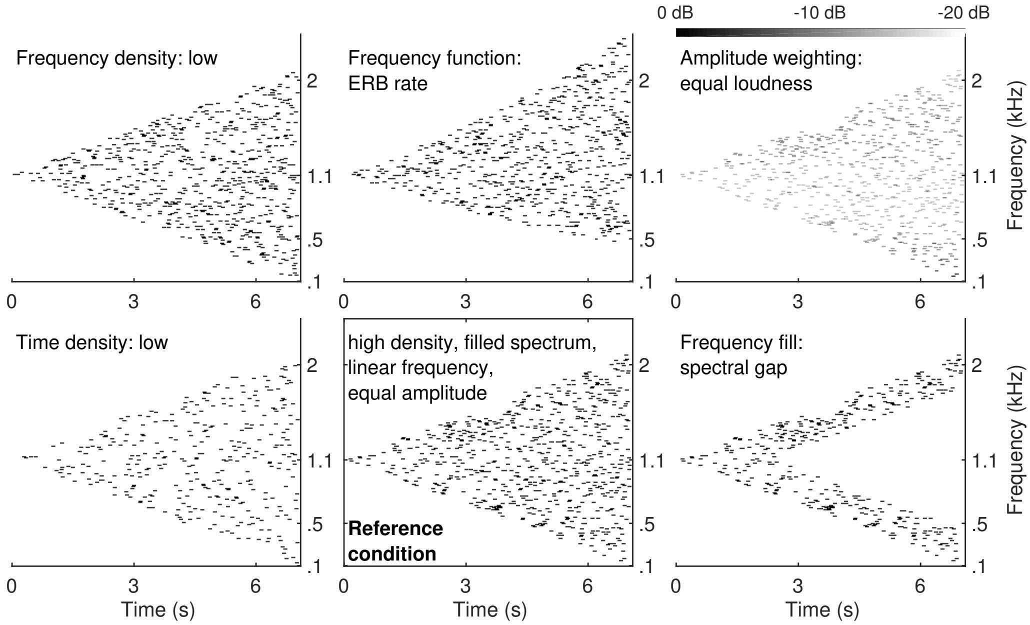

FIG. 4. Spectrograms of six example stimuli from Experiment 2, with the triangular sound shapes being composed of sinusoidal grains. The bottom-center panel serves as a reference condition which each of the surrounding panels compares to, across the factors *time density* (bottom-left), *frequency density* (top-left), *frequency function* (top-center), *amplitude weighting* (top-right), and *frequency fill* (bottom-right). Variation in amplitude is visualized in relative power level in dB; see legend at the top-right; amplitudes below -20 dB are not visualized.

bandwidth of 2000 Hz, this combination of frequency functions, however, introduced an unresolvable problem. More specifically, none of the triangle's two ends, i.e., its tip or its wide end, could be controlled in frequency without introducing misaligned frequencies on the opposite end. This irreconcilable issue arose from inherently divergent frequency functions given the additional constraint of maintaining a constant bandwidth. As a compromise, the tip was considered as the more important anchor, because its frequency served as the reference on which triangular symmetry was defined. In addition, the tip represented the dominant frequency that sounded throughout the (solid) shapes . Relative to the 1100 Hz frequency at the tip, these two functions therefore led to the maximum frequency limits of [100, 2100] Hz (bottom-center panel) and [434, 2434] Hz (top-center), respectively.

An additional frequency-related IV compared shapes that were completely filled with grains to ones exhibiting a widening spectral gap around the center frequency 1100 Hz (bottom-right panel). This widening gap occurred at a delay designed to reach a bandwidth of one ERB at 40% of the 7-s duration, thus becoming increasingly perceptible, and followed the same frequency scale as the main triangular trajectories. Finally, the fifth IV determined the amplitudes of individual sinusoidal grains. The first case considered equal amplitudes across all frequencies, whereas the second (top-right panel) used a psychoacoustic dependency and weighted amplitudes based on the frequency-dependent equal-loudness contours (Fletcher and Munson, 1933; ISO, 2003). Given that individual sinusoidal grains at the 1100 Hz tip exhibited about 60 dB SPL, the amplitude weightings were based on the 60-Phon contour.

For each of the 32 conditions, two different versions were tested in the experiment, resulting in a total of 64 experimental trials. These two versions were presented in two separate blocks; in each block, the 32 conditions were randomized in order. Furthermore, the order of the blocks was counterbalanced across all participants by alternation. To ensure that across all conditions the randomly generated sound shapes exhibited comparable distributional properties, a total of 999 versions for each condition had been generated initially, out of which two versions were selected that exhibited the closest fit to a reference distribution.

The condition for the highest frequency and time densities served as the reference (Figure 4, bottom-center panel).

With regard to how these sound-shape stimuli compared to those of Experiment 1, their grain density exhibited values in the bottom range of the previous experiment. As shown in Figure 2 (right panel), the percentage of the triangular area covered by the grains varied between 2.5% and 5% for the low and high time density levels, respectively, whereas frequency density did not affect the percentage of covered triangular area, for which reason those conditions are not displayed separately. This quantification used a reference grid with the same spectrotemporal resolution as for Experiment 1 and comprised eight instances per time density.

c. Participants. 20 participants (16 male, four female) with a median age of 41.5 years (range: 21–57) completed the experiment. Participants had a median of eight years of formal musical training (quantified as for Experiment 1); 12 participants classified themselves as professional musicians. Five participants reported having tinnitus, while another participant reported hearing difficulty at mid-range frequencies but only for the left ear.[2] As these hearing deficiencies seem rather common among practitioners of electroacoustic music and the reported deficiencies were not deemed a severe hindrance to the evaluation of the investigated shape-related qualities, no participants were excluded from the further analysis.

B. Results

a. Clarity measure. Clarity ratings did not yield any statistically significant effects across all acoustic factors, although Figure 5 (left panel) suggests a trend for a slight increase

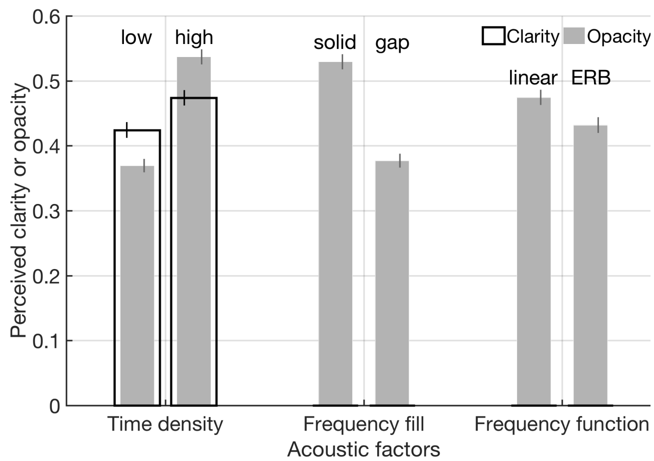

FIG. 5. Perceived clarity or opacity across different levels of time density (left panel), frequency fill (center), and frequency function (right) investigated in Experiment 2. See Figure 3 for complete legend.

in clarity (black bars) for higher time density. Given that Experiment 2 included sound shapes exhibiting spectral gaps that, however, did not occur in Experiment 1, a separate analysis on only the solid sound shapes was conducted to further investigate the anticipated influence of grain density. Indeed, in a paired t-test comparing all conditions involving low time density against those of high time density, greater time density again led to higher perceived clarity, $t\,(200) = -3$, $p < .01$, $d = -.24$.

b. Opacity measure. As shown in Figure 5 (left panel), opacity ratings (grey bars) increased for greater time density of grains, $F(1, 19) = 25.8$, $p < .01$, $\eta^2_G = .18$, being markedly more pronounced than the similar trend observed for clarity. Similarly, also the presence of a widening gap in the sound shape resulted in a marked reduction of perceived opacity (center panel), $F(1, 19) = 33.8$, $p < .01$, $\eta^2_G = .15$. This was complemented by the type of frequency scale also affecting opacity (right panel), in that linear frequency in Hz yielded slightly higher opacity ratings than ERB rate, $F(1,19)=5.8$, $p=.03$, $\eta^2_G=.01$. However, the latter effect did not seem to apply for conditions of low frequency and high time density, as suggested by a three-way interaction with these factors, $F(1,19)=5.2$, $p=.04$, $\eta^2_G<.01$.

c. Symmetry measure. The symmetry of shape did become relevant in this experiment. The strongest factor influencing symmetry was the kind of frequency function. As shown in Figure 6 (ratings on the left), sound shapes following linear frequency in Hz (green) were perceived as tilted downward, whereas those based on ERB rate (red) were judged as titled upward relative to complete symmetry (zero value), $F(1,19) = 90.0$, $p < .01$, $\eta^2_G = .45$. A number of two-way interactions with this factor provide more insight. Interactions with time density, $F(1,19)=8.2$, $p=.01$, $\eta^2_G=.01$, and frequency density, $F(1,19)=5.5$, $p=.03$, $\eta^2_G<.01$, suggest that the difference between frequency scales simply became slightly more pronounced for greater grain density.

An interaction between frequency scales and amplitude weightings, $F(1,19) = 13.8$, $p < .01$, $\eta^2_G = .02$, provides a more nuanced view on the total of four versions of frequency and amplitude scalings. As illustrated in Figure 6, different amplitude weightings did not appear to affect the symmetry ratings for the conditions involving ERB rate (both red). By contrast, for linear frequency, the conditions involving equal amplitudes (dark green), as opposed to equal loudness (light green), yielded ratings closer to complete symmetry (zero). A single post-hoc test between these two subsets ascertained a difference, $t\,(200)=4$, $p<.01$, $d=.32$. For the sake of completeness, a one-sample t-test for the linear-frequency-and-equal-amplitude subset against a mean of zero confirmed that these conditions still appeared to not be judged as completely symmetric, $t\,(200)=-7$, $p<.01$, $d=-.53$.

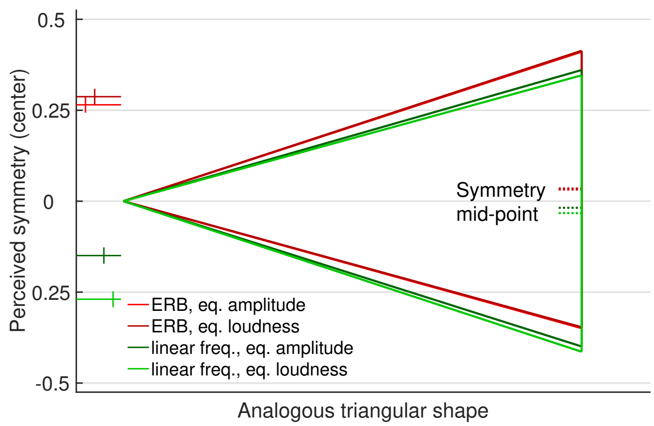

FIG. 6. Perceived symmetry across four combinations of two frequency functions (linear frequency, ERB rate) and two amplitude weightings (equal amplitude, equal loudness) investigated in Experiment 2. The y-axis represents the rating scale, with zero signifying complete symmetry. Horizontal lines intersecting the axis correspond to the group means, the corresponding intervals to standard errors. On the right, the four triangles illustrate the analogous degree of visual asymmetry the computer interface displayed to participants. The marked symmetry mid-points on the far right highlight the visual asymmetries.

Unrelated to frequency scale, an additional two-way interaction concerned the two density factors across time and frequency, $F(1,19) = 5.0$, $p = .04$, $\eta^2_G = .01$. This weak interaction resulted from slightly more positive symmetry ratings for the conditions that comprised a high and a low level from each factor, as opposed to the low-low and high-high density combinations.

d. Correlation among measures. As in Experiment 1, rank correlations were employed to assess the interrelatedness of all measures' rating profiles ($N = 64$). As shown in the bottom-left half-matrix in Table I, the same patterns of correlations as in Experiment 1 emerged, however, weaker in magnitude. Clarity was still moderately correlated with opacity, while hardly correlated with symmetry. Opacity and symmetry exhibited a weak negative correlation. The measure homogeneity was clearly correlated with all shape-related measures.

IV. DISCUSSION

Two listening experiments were conducted to explore the feasibility of perceiving the shape of sounds forming a triangle on the spectrotemporal plane. In direct analogy to the visual description of a triangle, three qualities were assessed as to their perceptual relevance, namely the *clarity* of the implied triangle's outline, the *opacity* (or inversely transparency) of the enclosed area, and the *symmetry* of the triangle relative to its tip on one (temporal) end. A number of acoustic factors were considered to study their potential effect on the perception of triangular sound shapes and their individual contributions to the three shape-related qualities.

As the composition of sound shapes relied on granular synthesis, the density among sound grains was expected to become perceptually relevant. As hypothesized, higher grain density appeared to lead to stronger perceptions of sound shape. In Experiment 1, both perceived clarity and opacity increased as a function of grain density. Importantly, however, clarity and opacity ceased to increase for sufficiently high grain densities, even though the observed perceptual plateau began in a region of grain density in which only about 40% of the spectrotemporal area was covered by sound grains (see density level V in Figure 2, left panel). This percentage may therefore represent a threshold for grain density above which no further increase in perceived clarity and opacity is achievable. On the other end, even very low density levels seemed to convey the clarity of shape sufficiently well, as average ratings amounted to at least a third of the scale range above the lowest possible clarity.

Experiment 2 distinguished between grain density along either time or frequency, moreover, studying this in the region of lowest grain density from Experiment 1, which yielded the greatest perceptual differences. Apart from an absence of main effects, frequency density only contributed to weaker interactions with time density. It should be acknowledged, at least, that the parametric variation of frequency density alone did not actually affect the percentage of triangular area covered by grains; this parameter therefore only influenced the selection of available frequencies involved, still to little effect on the shape-related qualities. To the contrary, time density yielded clearer effects, suggesting that primarily the temporal density of grains contributes to sound-shape perception. Further differences between perceived clarity and opacity emerged in Experiment 2 in that, except for a single post-hoc comparison, clarity was not overall affected by acoustic factors. Again, this suggests that the triangular outline was sufficiently well conveyed and that its perceived clarity remained robust to the investigated acoustical variables.

With regard to how clarity and opacity might differ in terms of perceptual grouping (Bregman, 1990), clarity could be presumed to rely on the Gestalt principle of *good continuation*, given that the perceptual system seems able to infer the complete triangular shape based on only a discontinuous, granular rendition of the outline. Conversely, opacity seemed more sensitive and varied more clearly as function of the general granularity of the spectrotemporal texture. As grain density directly relates to the spacing among grains on the spectrotemporal plane, this likely concerns the Gestalt principle of *proximity*, with greater proximity assumed to enhance the perceptual association or cohesion of the ensemble of grains. In an analogous manner, the same Gestalt grouping principles (Wertheimer, 1923) would likely also apply to perceiving a granular rendition of a visual triangle. However, the obtained results do not allow to deduce the exact nature of the auditory grouping, i.e., whether the triangular shape is perceptually grouped into a single entity or whether it would correspond to two or more auditory streams. After the experiments, some participants reported that they had noticed concurrent ascending and descending trajectories, which points toward the perception of at least two independent streams.

Spectral gaps emerging around the center frequency and widening over time had a unique effect on the perceived opacity of sound shapes. The occurrence of such gaps, which were designed to exceed the perceptual detection threshold of at least one ERB (equivalent rectangular bandwidth, Moore and Glasberg, 1983), went in line with a clear decrease in opacity compared to sound shapes exhibiting completely filled spectra. In other words, listeners perceived shapes exhibiting such gaps as more transparent. Perceived opacity or transparency therefore appears to vary as a function of both the degree of textural density and the presence of wider gaps in the spectral texture. Whereas these two factors contributed to effects of comparable magnitude, the type of frequency function also affected opacity. The psychoacoustic ERB scale (Moore and Glasberg, 1983) led to a slight decrease in opacity compared to linear scaling in Hz. One possible explanation is that the 2000 Hz bandwidth for linear frequency accommodated a larger number of JNDs and in turn also sampled frequencies, which could have contributed to a perceptually somewhat fuller coverage of the bandwidth compared to that for ERB rate. Alternatively, since the 2000 Hz bandwidth for the linear frequency was about 300 Hz lower than for ERB rate (compare bottom to top center panels in Figure 4), this frequency difference could have also contributed to the slight difference in opacity. Overall, the influence of frequency functions on opacity was still markedly smaller than the influence related to the textural properties grain density and spectral gaps.

Psychoacoustic scalings of both frequency (Moore and Glasberg, 1983) and amplitudes (Fletcher and Munson, 1933; ISO, 2003) were initially seen at a potential advantage in rendering sound shapes as perceptually more balanced. Despite the type of frequency function strongly influencing the symmetry of shape, however, none of the two frequency scales were judged as symmetric. Instead, sound shapes following ERB rate were rated as upward asymmetric, regardless of how the corresponding amplitudes were weighted. Conversely, shapes exhibiting linear frequency trajectories in Hz were rated as downward asymmetric. Since a total of four conditions for the two-by-two combinations of frequency and amplitude weightings were considered, these combinations covered a range of four distinct options among which differences in symmetry still emerged: the sound shapes based on linear frequency paired with equal amplitudes exhibited ratings closest to symmetry (0), although still clearly downward asymmetric.

The interpretation of the results on symmetry has to consider the known limitations inherent in the stimulus design. The misalignment of frequencies at the wide end between ERB-rate and linear-frequency functions was an irreconcilable consequence of controlling the frequency at the triangle's tip. As the misalignment manifested itself towards the end of the sound shapes, this may have biased participants judgment towards attending to the frequency differences along the symmetry continuum instead of evaluating the sound shapes against the 'ideal' point of symmetry. Therefore, no reliable estimate for the point of complete symmetry can be deduced from the current results. Nonetheless, given the exploratory aim to associate the symmetry quality to perceptually relevant acoustic factors, the findings indeed support that frequency functions and amplitude weightings affected perceived symmetry. Moreover, the directionality of the upward and downward tilts agrees with the common frequency-to-elevation correspondence (Evans and Treisman, 2010; Spence, 2011). While beneficial to the exploratory aims, the multifactorial experimental design was less suited to determine precise perceptual thresholds along individual parameters. A separate experiment investigating only the factors frequency function and amplitude weighting across more gradations and possibly even frequency ranges and sound levels is necessary for further clarification. Such an experiment could also address the limitation of the current study of using a constant absolute bandwidth of 2000 Hz to compare linear frequency with ERB rate, whereas an alternative means of normalization based on constant relative bandwidth (e.g., octaves) could have been chosen.

The only case where sound-shape orientation, i.e., whether a triangle widened or narrowed over time, became relevant concerned perceived clarity. In Experiment 1, triangles beginning at the tip and widening over time were perceived with greater clarity than their time-reversed replicas. This unanticipated finding points toward a perceptual asymmetry or bias that arises despite there being no spectral difference between the time orientations. This bias draws parallels to previously observed perceptual asymmetries, such as a loudness bias known for sounds with either increasing or decreasing amplitude ramps. In these studies, despite both ramp orientations exhibiting identical sound-level ranges, loudness perception was consistently overestimated for sounds with increasing as opposed to decreasing ramps. For instance, global loudness for sinusoids with increasing ramps is perceived higher than for the opposite orientation (Ponsot *et al.*, 2015; Susini *et al.*, 2007). As these findings were based on retrospective ratings, the bias for increasing ramps may arise from a recency effect of the high terminating sound level (Susini *et al.*, 2007). Another explanation concerns an ecological context (Neuhoff, 2001) in that increasing or decreasing sound-level ramps may signify approaching or receding sound sources or objects, respectively, with the perceptual looming bias for the former presumed to represent an advanced warning mechanism allowing for more time to react to a potential threat. As increasing and decreasing ramps also evoke similar biases in reaction times, neurophysiological and emotional responses (Bach *et al.*, 2009; Tajadura-Jiménez *et al.*, 2010), these findings lend further support to an ecological, if not even adaptive, relevance of the loudness bias.

In terms of similarities to the sound shapes studied in Experiment 1, loudness asymmetry has been observed for 20-dB ramps of up to 20 s duration (Susini *et al.*, 2007), which compares to the triangular sound shapes spanning 10 dB level change over a 11 s duration, and the bias also applies to broadband signals (e.g., noise, Neuhoff, 2001). Thus, even the observed asymmetries in clarity for triangular sound shapes could be related to an effect arising from loudness perception. However, it remains unclear whether clarity varied merely as a function of perceived differences in loudness or whether the time-reversed spectral evolution may have also contributed to the effect. Notably, this would represent another example of a loudness-related asymmetry affecting a different perceptual quality or process, such as for the previously mentioned findings for reaction times or emotion responses. In the same vein, perceived clarity may therefore be enhanced for widening sound shapes because of some perceptual or cognitive predisposition. With regard to music, this perceptual bias may even explain the observed asymmetry in the use of dynamics contours (Dean and Bailes, 2010), e.g., *crescendo* vs. *decrescendo*, potentially attributing a greater perceptual or cognitive salience to contours based on rising sound level.

This investigation focused on the visual-analogous qualities *clarity*, *opacity*, and *symmetry*, under the assumption that these represent separate aspects to the perception of triangular sound shapes. Indeed, a clear degree of separability became evident across the variation of a number of acoustic factors. The observed differences on how grain density influenced clarity and opacity and the consideration of related principles of auditory grouping lend support to both perceptual qualities being conceptually distinct, which is further supported by spectral gaps having solely affected opacity. The remaining quality symmetry assumes a distinct role in that it varied as a function of frequency and amplitude scaling, with higher grain density only enhancing the observed tendencies. Correlational analyses (see Table I) provide further insight into possible interdependencies among the three qualities, their patterns reflecting the distinct links between qualities and acoustic factors. Some degree of covariation is apparent between clarity and opacity, e.g., accounting for 45% and 15% of explained variance for Experiments 1 and 2, respectively. It should be noted that the latter likely reflects the general relationship between both qualities more, due to involving a greater variety of acoustic factors. With the investigated sound shapes presumably relating to spectromorphologies of texture framed by gesture (Smalley, 1997), clarity and symmetry appear to represent features related to gesture, whereas opacity seems to primarily account for textural properties. Given that texture could also describe a global sound property that is wholly unrelated to shape, sound homogeneity vs. heterogeneity (see Grill *et al.*, 2011) was also considered. Notably, homogeneity ratings exhibited correlations with all shape-related properties, suggesting that a single perceptual measure fails to achieve a more nuanced differentiation of aspects relating to shape. Of the three qualities, opacity expectedly exhibits the highest correlation with homogeneity (54% of explained variance), explained by their common link to texture. In sum, the findings argue for the notion of sound shape to concern a number of morphological qualities, with the three investigated ones seeming appropriate for the case of triangular shapes.

V. CONCLUSION

Dealing with how spectra evolve and are shaped over time, the theory of spectromorphology (Smalley, 1997) often alludes to extra-sonic phenomena like shape, gesture, texture, or motion, serving as a source for musical expression and discourse. The notion of sound shapes draws rather literal analogies onto a two-dimensional representation such as the spectrotemporal plane. Importantly, this notion also presumes the visual analogy to translate to auditory perception. For the common sound-shape geometry of a triangle (Blackburn, 2011; Smalley, 1997), three morphological qualities derived from vision seem to also apply to the auditory modality. The clarity of the triangular outline, the opacity of the enclosed area within, and the symmetry along the vertical/frequency dimension capture different aspects of the perceived sound shape, moreover, related to relatively distinct contributions of acoustic factors. The perception of sound shapes appears to therefore be multifaceted, whereas limiting its assessment to a single sound attribute (e.g., homogeneity) appears to conflate different shape-related properties, while also failing to differentiate between gestural and textural properties.

Given myriad possible arbitrary audiovisual mappings, attempts have been undertaken to identify those mappings of special value to electroacoustic-music practice (e.g., Giannakis, 2006). Such effective mappings could in fact draw on common, widespread cross-modal correspondences (Spence, 2011), and indeed, triangular symmetry seems related to one of the most widespread correspondences, that between frequency and elevation. Likewise, the observed multifaceted nature of shape perception probably extends to implicit associations between complex sounds and two-dimensional visual shapes (Adeli *et al.*, 2014; Köhler, 1947; Ramachandran and Hubbard, 2001). Similarly, clear parallels can also be observed between the auditory and visual realms sharing the same perceptual grouping principles for granular, pointillistic shapes (e.g., *proximity*, *good continuation*, Bregman, 1990; Wertheimer, 1923). In sum, it is conceivable that extra-sonic references to gestures, textures or motion could generally involve predispositions linked to cross-modal perception.

Considering the variety of ways in which sound shapes could be used in music, the findings of the current study have limitations that should be addressed. Obtained through an inherently exploratory approach, these findings confirm the perceptual relevance of the three morphological qualities in characterizing sound shapes, and they jointly assessed their relevance across a number acoustic factors related to musical practice. Although the observed influence of grain density, spectral fill, frequency, and amplitude functions on the morphological qualities should therefore be assumed valid, they provide only rough estimates concerning psychoacoustic thresholds or dependencies, requiring dedicated psychometric experiments for comprehensive characterization and validation. Furthermore, given the granular nature of the sound shapes, the identified links between the investigated acoustic factors and morphological qualities will only extend to cases involving similar degrees of textural homogeneity, whereas the perception of composite shapes that comprise sub-components varying in textural or gestural properties (e.g., *micro-composites,* Blackburn, 2011) could affect sound-shape perception differently.

The sound shapes investigated here considered literal mappings of two visual dimensions onto two spectrotemporal dimensions, based on how common software implementations associate shapes with spectrograms (e.g., *EAnalysis*, *AudioSculpt*). Although the time-vs-frequency mapping seems the most plausible approach implied by spectromorphology, the vertical visual dimension may not always be understood as referring exclusively to frequency, as amplitude is also integral to the spectrum. As a result, visualizations of sound shapes may in fact include some degree of ambiguity, by possibly confounding the vertical dimension for both frequency and amplitude, which similarly applies to other examples of graphic scores for music. In certain cases, two-dimensional visualizations could entail conceptual hybrids that concern waveform representations (time-domain) at the local scale of the vertical dimension, while its grand scale involves relationships along frequency. Yet on another level, the relationship between visual and auditory shape does not even have to rely on a direct mapping of visual to acoustic representations but could still involve a translation via an intermediate representation such as motion. For instance, listeners are able to identify visual shapes based on the sonification of velocity profiles of drawing gestures (Thoret *et al.*, 2014). Overall, sound shapes may therefore concern scenarios that are already less related to its implied meaning within spectromorphology, although these alternatives may similarly evoke shared notions like gesture and motion. Still, all these scenarios seem to most likely draw on implicit associations between the sensory modalities. Exploring these cross-modal correspondences (Spence, 2011) in the future as to their potential utility to electroacoustic music could lead to developing perceptually informed tools or control strategies for sound synthesis and processing that operate along relevant amodal morphological parameters.

ACKNOWLEDGMENTS

The author would like to thank colleagues and students from the Music, Technology and Innovation Research Centre for their participation and exchange in developing this study. Participant fees were partly paid through funds made available through the VC2020 Lectureship scheme at De Montfort University.

[1] This choice of frequency scales resulted in anticipation of the role of frequency function being investigated in Experiment 2, based on which Experiment 1 spanned the maximum bandwidth covered by both investigated frequency scales (linear frequency in Hz and ERB rate).

[2] As six out of 20 participants reported hearing issues, separate ANOVAs on the data from the remaining 14 participants were also evaluated. There was no clear indication that reported hearing issues compromised the interpretability of the results obtained from all 20 participants. All medium to large effects were confirmed at about the same level of statistical significance. Only the weakest effects ($\eta^2_G \leq .01$) did not attain statistical significance.

Adeli, M., Rouat, J., and Molotchnikoff, S. (**2014**). "Audiovisual correspondence between musical timbre and visual shapes," Frontiers in human neuroscience **8**, 352.

Attneave, F., and Olson, R. K. (**1971**). "Pitch as a Medium: A New Approach to Psychophysical Scaling," The American Journal of Psychology **84**(2), 147.

Bach, D. R., Neuhoff, J. G., Perrig, W., and Seifritz, E. (**2009**). "Looming sounds as warning signals: the function of motion cues," International Journal of Psychophysiology **74**(1), 28–33.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behavior Research Methods **37**(3), 379–384.

Blackburn, M. (**2011**). "The visual sound-shapes of spectromorphology: an illustrative guide to composition," Organised Sound **16**(01), 5–13.

Bregman, A. S. (**1990**). *Auditory scene analysis: the perceptual organization of sound* (MIT Press, Cambridge, MA).

Couprie, P. (**2014**). "EAnalysis (version 1) [software]," http://logiciels.pierrecouprie.fr (Last viewed November 16, 2017).

Dean, R. T., and Bailes, F. (**2010**). "A rise-fall temporal asymmetry of intensity in composed and improvised electroacoustic music," Organised Sound **15**(2), 147–158.

Evans, K. K., and Treisman, A. (**2010**). "Natural cross-modal mappings between visual and auditory features," Journal of vision **10**(1), 6.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," Bell System Technical Journal **12**(4), 377–430.

Giannakis, K. (**2006**). “A comparative evaluation of auditory-visual mappings for sound visualisation,” Organised Sound **11**(3), 297–307.

Grill, T., Flexer, A., and Cunningham, S. (**2011**). “Identification of perceptual qualities in textural sounds using the repertory grid method,” in *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, Coimbra, pp. 67–74.

IRCAM (**2013**). “AudioSculpt (version 3.4.0) [computer program],” http://anasynth.ircam.fr/home/english/software/audiosculpt (Last viewed November 16, 2017).

ISO (**2003**). *Acoustics: normal equal-loudness-level contours* (International Organization for Standardization, Geneva), ISO 226:2003.

Köhler, W. (**1947**). *Gestalt psychology* (Liveright, New York).

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2008**). “Is Relative Pitch Specific to Pitch?,” Psychological Science **19**(12), 1263–1271.

Moore, B. C. J. (**2013**). *An introduction to the psychology of hearing*, 6th ed. (Brill, Leiden).

Moore, B. C. J., and Glasberg, B. R. (**1983**). “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” Journal of the Acoustical Society of America **74**(3), 750–753.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1993**). “Detection of temporal gaps in sinusoids: effects of frequency and level,” Journal of the Acoustical Society of America **93**(3), 1563–1570.

Neuhoff, J. G. (**2001**). “An adaptive bias in the perception of looming auditory motion,” Ecological Psychology **13**(2), 87–110.

Parise, C. V., and Spence, C. (**2012**). "Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test," Experimental Brain Research **220**(3-4), 319–333.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). "Detection of silent intervals between noises activating different perceptual channels: some properties of "central" auditory gap detection," Journal of the Acoustical Society of America **101**(6), 3694–3705.

Ponsot, E., Susini, P., and Meunier, S. (**2015**). "A robust asymmetry in loudness between rising- and falling-intensity tones," Attention, perception & psychophysics **77**(3), 907–920.

Ramachandran, V. S., and Hubbard, E. M. (**2001**). "Synaesthesia – A window into perception, thought and language," Journal of Consciousness Studies **8**(12), 3–34.

Smalley, D. (**1997**). "Spectromorphology: explaining sound-shapes," Organised Sound **2**(2), S1355771897009059.

Smith, N. A., Trainor, L. J., and Shore, D. I. (**2006**). "The development of temporal resolution: between-channel gap detection in infants and adults," Journal of speech, language, and hearing research **49**(5), 1104–1113.

Spence, C. (**2011**). "Crossmodal correspondences: a tutorial review," Attention, perception & psychophysics **73**(4), 971–995.

Stevens, S. S., Volkmann, J., and Newman, E. B. (**1937**). "A Scale for the Measurement of the Psychological Magnitude Pitch," Journal of the Acoustical Society of America **8**(3), 185–190.

Stumpf, C. (**1883**). *Tonpsychologie: Band 1 (Tone psychology: Volume 1)* (Hirzel, Leipzig).

Susini, P., McAdams, S., and Smith, B. K. (**2007**). "Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings," Acta Acustica united with Acustica **93**(4), 623–631.

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., and Västfjäll, D. (**2010**). "Embodied auditory perception: the emotional impact of approaching and receding sound sources," Emotion **10**(2), 216–229.

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J.-L., and Ystad, S. (**2014**). "From sound to shape: auditory perception of drawing movements," Journal of Experimental Psychology: Human Perception and Performance **40**(3), 983–994.

Wertheimer, M. (**1923**). "Untersuchungen zur Lehre von der Gestalt: Teil II (Investigations on the study of Gestalt: Part II)," Psychologische Forschung **4**(1), 301–350.