

Flexibility Conceiving Relationships between Timbres Revealed by Network Analysis

ROGER T. DEAN¹ , FELIX DOBROWOHL^{1,2} and YVONNE LEUNG^{1,3}

¹MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

²University of Potsdam, Germany

³University of New South Wales, Australia

Email: Roger.Dean@westernsydney.edu.au

Perceived relationships between timbres are critical in electroacoustic music. Most studies assume timbres have fixed inter-relationships, but we tested whether distinct tasks change these. Thirty short sounds were used, from five categories: acoustic instruments, impulse responses, convolutions of the preceding, environmental sounds and computer-manipulated instrumental sounds. In Task 1, 46 non-musicians formed a ‘cohesive’ sonic ordering of unlabelled icons (sounds attached). In Task 2, they categorised the icons into four boxes. In Task 3 listeners separately ordered the sounds from each of Task 2’s boxes using the approach of Task 1. Tasks 1 and 2/3 revealed distinct orderings, consistent with conceptual flexibility. To analyse the orderings, we replaced conventional distance by adjacency measures, and described each system as a network (rather than spatial positions), confirming that the two task outcomes were distinct. Network analyses also showed that the two systems were mechanistically distinct and allowed us to predict temporally changing networks, modelling the observed networks as successive perceptions. Further simulated networks generated with the temporal model readily encompassed all possible pairings between the sounds and not just those we observed. The temporal network model thus confirms conceptual flexibility even in untrained listeners, clearly suitable for a composer to use.

1. INTRODUCTION

Electroacoustic composers often use small-scale motivic structure, transforming it to create new relationships and hence the macrostructure of a piece (Xenakis 1971; Wishart 1985). Orchestral composers Schoenberg and Webern have also used timbral variation within their motives, as in *klangfarbenmelodie* (Rushton 2001), voicing a melody through a series of timbrally different instruments. Schoenberg also discussed ‘timbre structures’. Electroacoustic music (EAM) has increased attention to timbral morphing and juxtaposition, small or large scale. EAM composers thus must forge diverse and mutable relationships between timbres, most commonly as sequential relations. A case study of composer Roger Reynolds’s work illustrates this mutability (McAdams 2004), and reviews of cognitive processes in composition (Dean 2017) and EAM (Dean 2009) cohere with this. Such composerly conceptual flexibility suggests that

listeners need comparable flexibility in conceiving timbres: to identify recurrent and changing sequencing of timbral combinations in a piece.

In contrast to such fluidity, most studies of timbre perception map apparent fixities in their relationships (McAdams 1993; Bregman 1994; McAdams and Siedenburg 2019). Generally, they measure perceptual distances between pairs of short sounds, and then use multidimensional scaling (MDS) to provide a two- or higher dimensional representation, following classic ‘timbre space’ studies of instrumental sounds (Grey and Gordon 1978). Such maps can suggest analogies between sounds, where sounds *a* and *b* bear a similar relation as *x* and *y* (McAdams and Cunible 1992). Word ‘embedding’ techniques in machine learning and psychology similarly relate woman/man, countess/count or queen/king (Pennington, Socher and Manning 2014). Scalar acoustic features only predict a portion of the MDS mappings (Caclin, McAdams, Smith and Winsberg 2005): temporal factors are probably important in the remainder.

Thus kinetic trajectories of acoustic features related to timbre are predictors of detection of phrasing by untrained listeners, particularly in EAM (Olsen, Dean and Leung 2016), paralleling the ontogeny of language acquisition (Johnson 2016). Similarly, analyses of continuous response measures of musical change and affect (Bailes and Dean 2012; Dean, Bailes and Dunsmuir 2014a, 2014b) reveal the kinetic contribution of both acoustic intensity and spectral features. So the conception of a particular timbre (or phoneme) may be influenced by its context. For example, changing background noise levels changed the timbre space of a group of synthetic sounds (Zacharakis, Terrell, Simpson, Pasiadis and Reiss 2017).

Flexibility and mutability suggest viewing inter-sound relationships as transforming networks, in which any pair of sounds might be connected to varying degrees. This is analogous to networks between people or ideas, with interactions that are one- or two-directional and of varying intensity. ‘Network analysis’ has a major role in spheres such as sociology,

neuroscience and internet commerce (see McAndrew and Everett 2015 for a musical social network).

Network analysis has been little applied to musical sounds. Relationships between mean computational features in pop songs covering an approximate 50-year period since the 1950s have been used in a network analysis of their ‘evolution’, suggesting that keeping the spectral features (i.e., presumably the timbres) relatively invariant is successful (Serrà, Corral, Boguñá, Haro and Arcos 2012). We can find no network analysis of perceptual relations between short sounds.

So here we assess whether the composer’s need for fluidity in conception of timbral relationships can be fulfilled by non-musicians in two successive ordering tasks. We propose that different tasks create different orderings, and reveal different timbral networks. Early network analyses simply sought the best descriptive model of a set of relationships, but more recent generative models (Betzel and Bassett 2017), allow the prediction of populations of networks. This generative modelling allows us to assess whether a composerly flexibility in transforming timbral relationships might be readily available to them and to listeners.

2. MATERIALS AND PARTICIPANTS

The 30 sounds used (each about 3 seconds) were either open access, recorded manually, or produced by experimental (convolution) or compositional (digital signal processing, DSP) manipulation by the first author (referenced by his initials, RTD). Table 1 lists the sounds and their membership of five pre-chosen sound design groups (acoustic instruments, impulse responses, convolutions of the preceding two, environmental sounds and computer-manipulated instrumental sounds). The Table 1 note explains a subdivision of the environmental group addressed in the network analyses.

Participants ($n=46$) were recruited within Western Sydney University as being ‘untrained’ in music; psychology undergraduates received a course credit for participation. They completed the Goldsmiths’ Musical Sophistication Index (GMSI; Müllensiefen, Gingras, Musil and Stewart 2014), and reported on average 2.0 years training on an instrument (maximum 6) and a mean of 1.7 years in music theory (maximum 6). They were all below the median of the large UK population aggregate GMSI score, hence termed ‘non-musicians’. All participants self-reported as students. They reported 19 nationalities, but all except six spent their formative years in Australia. Mean age was 23.2 years (range 17–47); 32 reported their gender as female, 14 as male.

Max (v7.0, from Cycling74) was used to present the experiment and to collect the responses (screen positions).

3. PROCEDURE AND METHODS

After the GMSI, participants listened to the sounds on Beyerdynamic DT770 Pro headphones, seated in front of a computer screen.

The first task required participants to audition and order unlabelled sound icons ‘cohesively from left to right’ according to their principles (Figure 1). They could listen to the sounds and adjust their relative positions on a horizontal line as much as they wished. This task created one type of ‘demand’ for conceptual responses.

Task 2, following, was intended to provide different pressures: it required ‘grouping’ sounds. Figure 2 shows its interface, requiring sorting into four boxes based on ‘affinity between sounds’. The boxes were, from left to right, blue, green, orange and turquoise, but there was no request to order sounds across the screen or within a box. This was a categorisation task, where congeners would be placed in the same box. We expected it would be quite difficult, as there were five sound design groups but only four boxes, potentially requiring some reconceptualisation.

Next, Task 3 showed box by box the sound icons from each individual box, in the order blue, green, orange and turquoise, but with unboxed sounds randomly placed below a horizontal line, as in Task 1. Task 3 required a left to right ordering on the line of sounds from the presently available set, as for Task 1. After box 1, the screen was cleared and the process repeated for each box. So Task 3 made the same demand as Task 1, but with four subsets of the sounds. Tasks 2/3 cumulatively provided different conditions for relating the sounds from those of Task 1. The ‘two tasks’ below refer to Task 1 vs the cumulation of Tasks 2/3. We fixed the order of tasks, since Task 1 is more-open ended than 2. Furthermore, the likely influence of Task 1 on Task 2/3 is less perturbing than vice versa. In any case, Task 2 necessarily influences the possibilities available within 3.

After listening, participants briefly summarised their ordering strategy(ies); whether they would like to label the sound groups in Task 2; and whether they would like less or more boxes in Task 2.

Comparing the orderings indicated by the two tasks, person by person, requires taking account of the fact that Task 2 could be done with or without regard for the order of the boxes on the page. In contrast, Task 1 requires an unambiguous single ordering of all the sounds. Thus to determine whether the sequences in the two tasks as performed by each individual could or could not have been identical, all possible sequences of the boxes (categories) in Task 2 need to be used to assemble all the possible overall Task 2/3 orderings for each participant. Four different boxes provide 24 permutations of their orders. Note that the sounds are labelled by numbers, but the numbers are merely

Table 1. Subdivision of the environmental group addressed in the network analyses

Sound number/description	Design group	Environmental subgroup
1/AkkordionRTDprocessedv1	RTD	
2/Best_Helicopter_Pass-Mike_Koenig-1156203370	Env	Machine/Human
3/Convolution_Piano_Teufelsberg	Conv	
4/Convolution_Sax_Stadium	Conv	
5/Convolution_Synth_Tunnel	Conv	
6/Convolution_Trombone_Plate	Conv	
7/Guitar-RTDprocessedv1	RTD	
8/Humpback Whales-SoundBible.com-1557869991	Env	Natural
9/Ice-skating-daniel_simon	Env	Machine/Human
10/Impulse_MMA_Stadium_IR	Imp	
11/Impulse_Plate3s_IR	Imp	
12/Impulse_Teufelsberg_IR	Imp	
13/Impulse_Tunnel_IR	Imp	
14/Instrument_Piano-soundlab	Inst	
15/Instrument_Sax-soundlab	Inst	
16/Instrument_Synthesiser_dry	Inst	
17/Instrument_Trombone-soundlab	Inst	
18/Monsoon-SoundBible.com-382003327	Env	Natural
19/Pelican Chick-SoundBible.com-549990839	Env	Natural
20/Photocopy-machine-daniel_simon	Env	Machine/Human
21/Piano-RTDprocessedv1	RTD	
22/Rowing A Boat-SoundBible.com-2108783030	Env	Natural
23/Saxophone-RTDprocessedv4	RTD	
24/Sewing Machine 3-SoundBible.com-1495871530	Env	Machine/Human
25/Slot-machine-daniel_simon	Env	Machine/Human
26/Trombone-RTDprocessedv1	RTD	
27/Water-drops-daniel_simon	Env	Natural
28/Waterphone-daniel_simon	Inst	
29/Wind_1234-mike-koenig	Env	Natural
30/Windy-SoundBible.com-1165996801	Env	Natural

Note. Sound Design Groups: Conv, Convolved sounds; Env, Environmental sounds; Imp, Impulse Responses; Inst, Acoustic musical instrument; RTD, Computer music DSP-manipulated acoustic instrument sounds. In the final column, Environmental sounds are subcategorised: ‘Natural’ are those not requiring human intervention; ‘Machine/Human’ are either industrial or human activity sounds (both involving human intervention). Sound 22 is similar to the creaking of a tree. The Machine/Human subgroup is sometimes simply termed ‘Machine’. The RMS intensities of all sounds were equalised, but participants could adjust their listening level.

categorical markers, thus to assess the distance between two sequences of 30 categorical items, the only factor of interest is the number of steps along the sequences between sound x in Task1 and sound x in Task 2/3, a ‘transposition distance’. Thus the maximum distance between x in two such sequences is 29 (position 1 in one and position 30 in the other). The distance for item y in position 2 of one sequence is then maximised if in position 29 of the other sequence (i.e., distance 27) and so on. The calculation is illustrated below. Consequently, the maximum transposition distance feasible between two sequences of the same 30 items is 450, that is $(29 + 27 + 25 \dots + 1) \times 2$. The minimum is 0, when the two sequences are identical. Code was written to determine this transposition distance between every possible sequence of the four box orders in Task 2/3, and the single sequence of Task 1. Note that a result of 450 implies a precise retrograde (observed once).

The total transposition distance for this pair of sequences is obtained by adding all the individual distances. For each individual, all permutations of Task2/3 orderings were assessed against that participant’s Task 1 ordering. The values obtained are then assessed across all participants, by taking all the individual minima, and again determining their range, mean, median and standard deviation; and similarly for the individual participant maxima values.

All statistical analyses used R version 3.5, and the *igraph* and *Statnet* network analysis packages. These are explained further in the following section.

4. RESULTS

We first showed that the inter-item relationships of Task 1 and Task 2/3 were distinct. Network analysis then illustrated the distinction and demonstrated the relationship mutability.

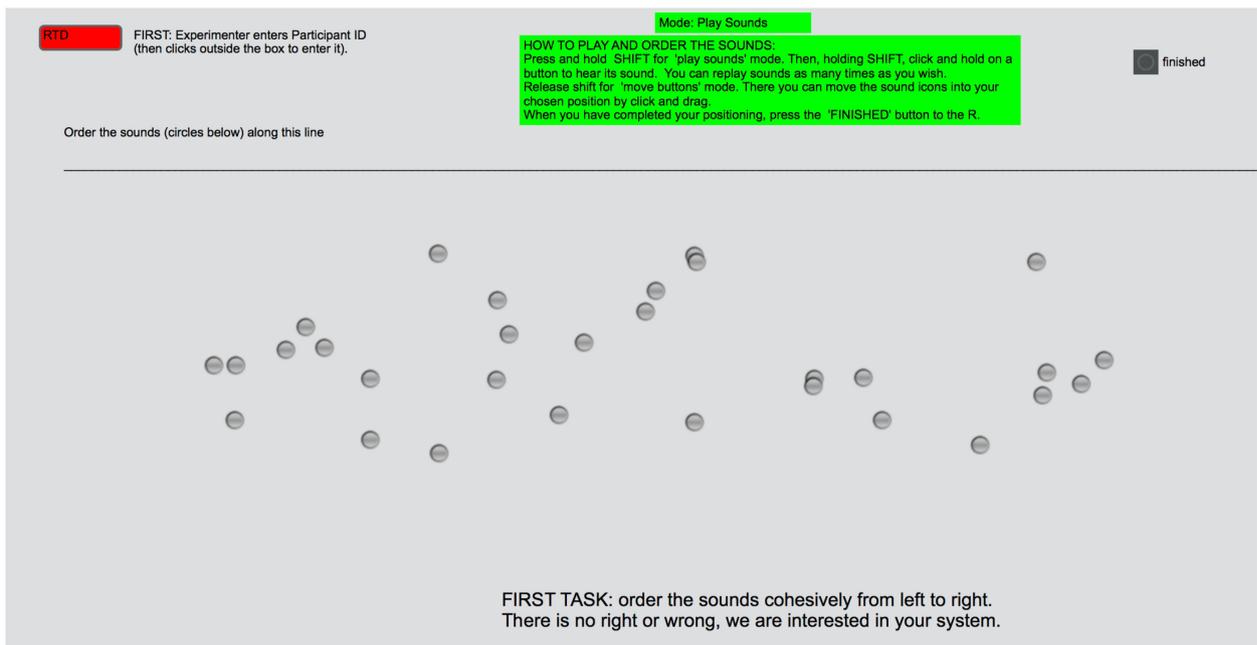


Figure 1. A screen dump of the interface for Task 1, coded in Max.

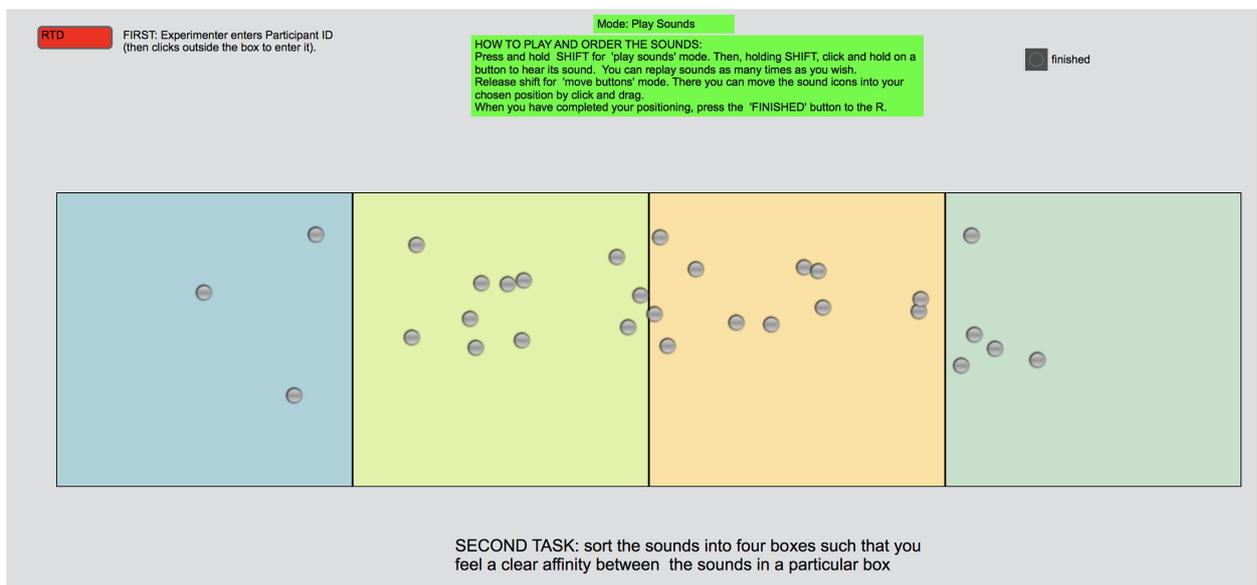


Figure 2. A screen dump of the interface for Task 2.

4.1. Cumulative orderings from Task 1 and Task 2/3 are distinct

We expected correlation between orderings in the two tasks as performed by any individual participant, since once used to conceive relationships a sonic feature is likely to remain in memory. But did participants nevertheless present different orderings in the two tasks, given the different demands, and as assessed by the

transposition distances? Forty-five participants used all four boxes in Task 2/3, while one used only three (22, Figure 3). This latter was accommodated within the analysis code (using only six permutations of the box order for this participant).

Every participant's Task 2/3 orderings (i.e., all the permutation orders) were all different from their ordering of Task 1. The mean, median and standard deviation of the minimum transposition distances,

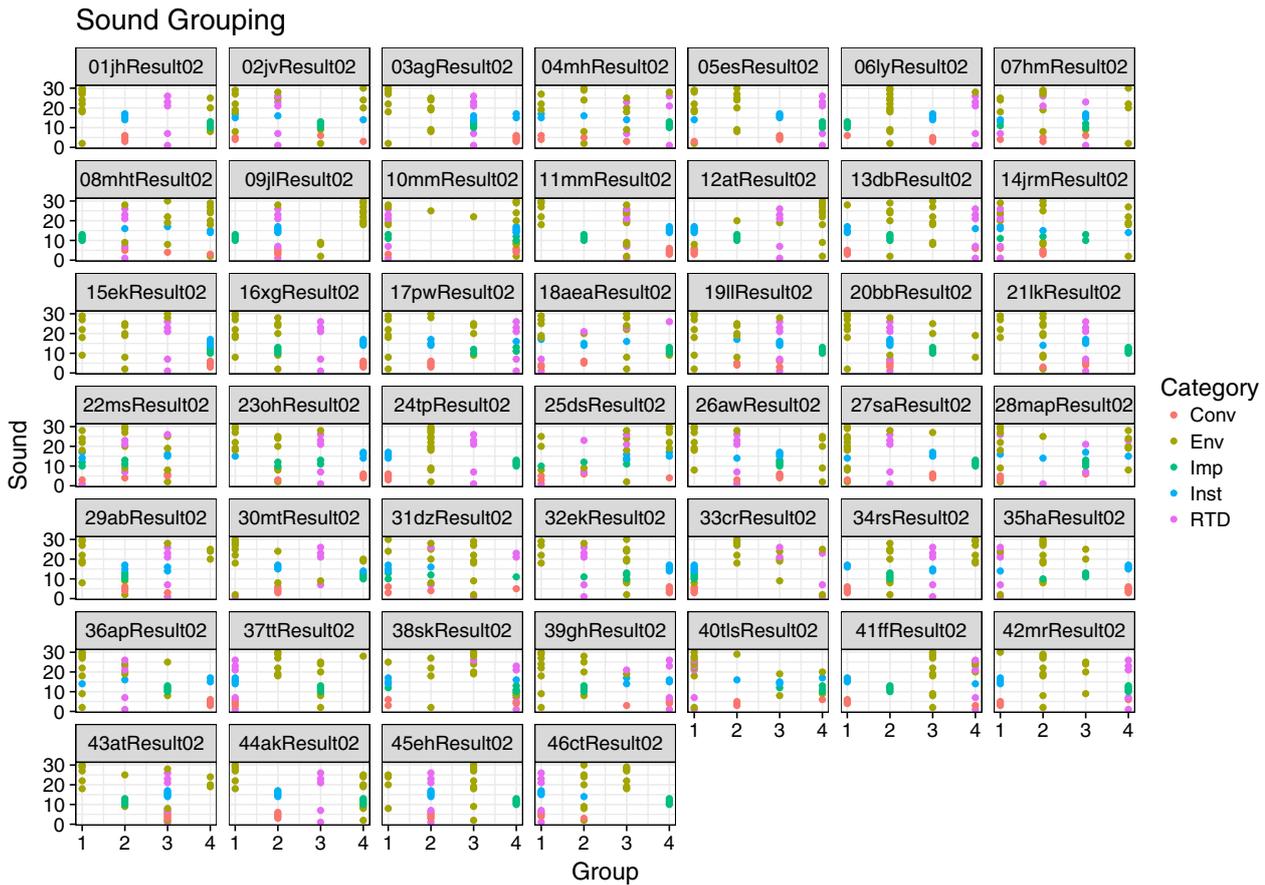


Figure 3. Sound grouping: the distribution of the sounds by participant and box in Task 2.

Note. Abbreviations: The sound design abbreviations are as before. Boxes are numbered 1–4 from left to right; anonymised participants are indicated above each individual display as the first two numbers. ‘Sound’ refers to items 1–30.

participant by participant, were 155.4, 146.9 and 58.7 with the range 48–276. Even the lowest value corresponds to more than 10% of the maximum possible. For the maximum distances, mean, median, standard deviation and range were, respectively, 398, 410, 33.3 and 334–450 (from 74.2 to 100% of the maximum possible). This shows that every participant behaved differently in Task 2/3 than in Task 1, as we hoped.

A first-order adjacency matrix was constructed for each of the tasks, based on the frequencies with which each sound was placed next to another. Figure 4a shows the matrix for Task 1: the diagonal is occupied by 0, since no item can appear twice. The matrix is symmetrical with respect to the diagonal (because these are undirected frequencies, AB equivalent to BA, providing one adjacency shared by A and B). In Task 1, all sounds are aligned in a single sequence, left to right, so this construction is straightforward. However, in Task 2/3 we provided no suggestion that the four boxes were themselves ordered. Figure 3 shows the disposition of the design groups by box for all participants: several members of a design group

were often co-boxed, with no regard for the position of that box. So DSP-manipulated sounds (magenta) commonly appear in the same box, but that box may be any of the four. The fact that the box ordering giving the lowest transposition distance between Task 1 and 3, was blue–green–orange–turquoise, the screen order, in only 9/46 cases, confirms that that the boxes were not treated as part of an ordering process. Consequently, for the adjacency analysis we assume no relation between the four suborders in Task 2/3, and each suborder is treated as a partial realisation of the overall ordering. This means that there are three less adjacency values obtained per participant from Task 2/3 than Task 1 (because no adjacencies between the boxes are assumed). Thus the adjacency matrix for Task 2/3 (Figure 4b) is obtained simply by adding together those from the occupied boxes from each participant.

Given this evidence on the interpretation of Task 2/3 ordering, there still remain two ways of interpreting orderings cumulated across participants: as being directed (a left to right adjacency of *x* and *y* is distinct

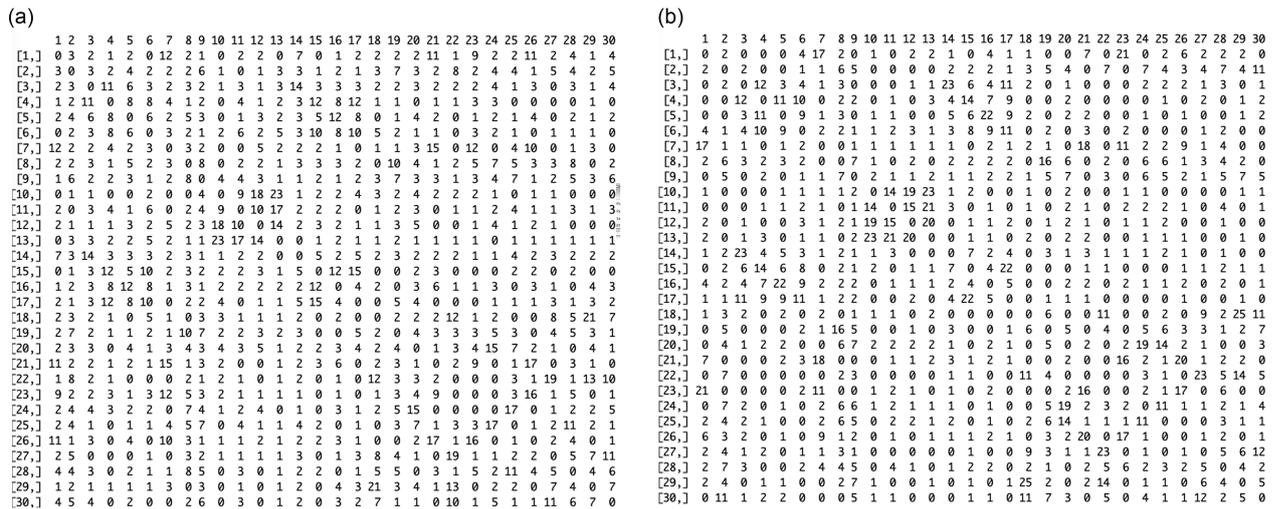


Figure 4 (a) Undirected adjacency matrix for Task 1. (b) Undirected adjacency matrix for Task 2/3
Note. Numbers 1–30 are the test sounds. The sum of the adjacencies in Task 1 is 2,668 (corresponding correctly to 29 adjacent pairings × 46 participants × 2 since the matrix is symmetric, undirected). There are 2,394 adjacencies in Task 2/3 because of the omission of counts between boxes, and the fact that one participant only used three of the four possible boxes. The easiest way to interrogate these data is probably to choose a column, and read downwards from the diagonal of zeros, bearing in mind that the matrices are symmetrical about the diagonal.

from its reversal), or undirected (where the two cases are not differentiated).

The proliferation of 0 values outside the Figure 4 diagonal shows that some adjacency pairings are unexploited: 172 (6.4% of counts) for Task 1, and 314 (13.1%) for Task 2/3. These two adjacency matrices are distinct, in accord with the observations comparing the individual participant sequences from Task 1 and Task 2/3. However, as commented already, one would expect correlation between the matrices. The Mantel test (a form of permutation analysis involving repeated correlations) confirms this, showing a strong correlation of 0.86 ($p = 0.001$, two-sided). This means that about 26% of the variance in the adjacency data is not shared in the two orderings, revealing the flexibility we are focused on. The distinction between the matrices remains even when a random 10% of participants’ data are removed from the matrix for Task 1 (since there are approximately 10% less counts in the Task 2/3 matrix). The Mantel test does not take account of intra- or inter-personal differences: it is rather like taking two global averages. So the strong correlation is unsurprising, and it remains clear that each individual performed differently in the two tasks.

4.2. Restricted Markov chain predictions of orderings

Next we consider the adjacency matrices as directed (i.e., we distinguish L to R sequences 1–2–3 and 3–2–1). This is treating the sequences as first-order

Markov chains restricted by the fact that no item can recur. Put simply, in such a Markov chain, the probability of an item appearing in position $n+1$ is entirely predicted by the occupancy of position n . This is termed the transition probability between the two positions. This allows us to determine the directed adjacency matrices using the R ‘Markov’ package, and then to construct the corresponding Markov chain models. The values in these directed adjacency matrices (not shown) are now transition probabilities, but they again allow the Mantel test. This confirms the difference between the two matrices, which remains true when 10% of the Task 1 participants are removed at random. We can again conclude that the adjacency matrices of Task 1 and Task 2/3 are distinct. Figure 5 illustrates this by comparing the Markov ordering predictions for the two tasks.

The all-participants’ Markov chain model also allows prediction of individual participants’ orderings in Task 1 based on their choice of the extreme left sound. These predictions are much better than chance; for example, showing 4–10 correct predictions (for which the chance probability is only approximately 10^{-6}). Various tests indicate that the data are legitimately described as Markovian, and there is no indication that Markov orders higher than one are required for their best modelling.

Thus whether directed or undirected, the Task 1 and Task 2/3 orderings are statistically distinct, and unlikely to arise by chance.

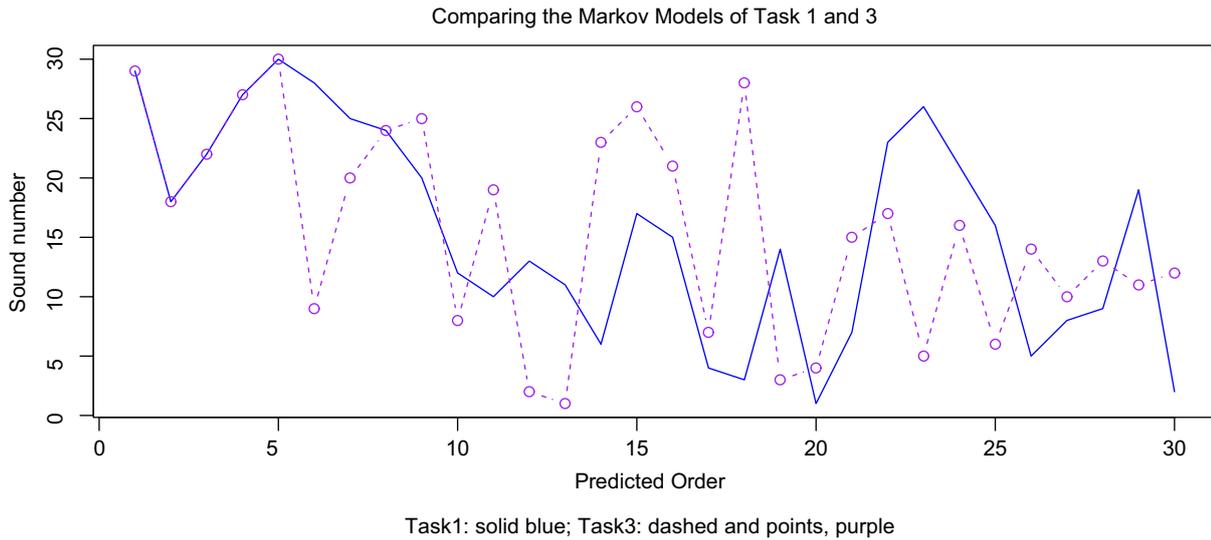


Figure 5. Markov model predicted orders for Tasks 1 and 2/3.

Note. The Markov models of Tasks 1 and 2/3 are used to predict the overall most likely ordering patterns across all participants, in each case given a starting sound of 29, and sequentially predicting the next sound using the transition probabilities. Predictions at each point are restricted to sounds which have not already been allocated.

4.3. Relationships between sounds' design group and their orderings: participant strategies

If our five design groups were perceptually influential, then members of a group would be adjacent more often than they are to members of another group, even though our purpose was to partially disrupt these relationships. A simple analysis of this issue used the data from Task 2, where categorisation is explicitly required. Table 2 shows a co-occurrence matrix representing these results. Because one group only contained four members, Table 2 shows the frequency that >25% of the sounds of one group were co-boxed with >25% of another. The diagonal values show that members of each design group commonly co-located with other members of that group, and the main 'confusion' (off-diagonal entries) was between convolved and instrument sounds, which is unsurprising since convolutions were done on the instruments.

In response to enquiries on ordering strategy(ies) used in any task, participants often mentioned using pitch, and quite often loudness (even though the sounds were RMS-equalised). We did not ask about their interpretation of our instructions on 'cohesively' and 'grouping ... by affinity', so as to minimise directional influences of our questions. A few people identified sonic categories (e.g., 'music', referring primarily to instruments), and a few mentioned their degree of liking. A very few explicitly used 'categorisation' in the box Task 2 more than in ordering Task 1, while a few considered ordering as a series of nearest neighbours in features. Most participants had no interest in attaching labels to the boxes after

they had occupied them (suggesting they might not have explicitly identified design features). Twenty-four respondents wanted additional boxes in Task 2/3, 11 wanted fewer (some having difficulty recollecting numerous sonic features), while the remaining 11 were indifferent to this. Participants were thus not invested in shared ordering strategies, rather, they were diverse, as we anticipated. A slight majority, preferring additional boxes, implicitly recognised the categorisation pressure of Task 2.

4.4. Network analysis of orderings

The co-occurrence matrix of Table 2 emphasises that sounds from different design groups co-relate in diverse ways. This suggests that a desired compositional usage of timbral items (partially recreated by our two tasks) is to form relationship networks that can be used diversely. Thus it is interesting to consider the network features (Betzel and Bassett 2017) implied by our data. Network models identify the intensity of the relationship between pairs of sounds, rather than providing a more abstract 'distance' measure. For example, some social network analyses might use numerous features that link any individual person (in this case, the 'agents' that are the subjects of any generic network model are individual people), such as frequency of contact, topics of contact, shared activities or disciplines, age, gender and financial status. In the present case, the information we are using that links our agents (being individual sounds) is solely their frequency of contact (adjacency in the orderings determined by the participants), but as we point out

Table 2. Co-occurrence matrix from Task 2 (categorisation)

	Env	Inst	Conv	Imp	RTD
Env	44	7	8	15	10
Inst	7	51	40	10	15
Conv	8	40	49	7	13
Imp	15	10	7	40	12
RTD	10	15	13	12	47

Note. Each number shows the frequency that >25% of the sounds of the design group specified on the vertical axis were assigned to the same box as were >25% of the group specified on the horizontal axis. Design group abbreviations are as before. Note that the column and row sums are heterogeneous.

later, other features could be relevant for different purposes (such as acoustic features, as distinct from perceptual relationships). We present both descriptive networks, which display the observed structures, and then generative network models, which suggest the processes by which the networks arise and allow extrapolation to generate the full set of networks consistent with a particular model.

Given the task natures, we only consider undirected, but in this section, valued networks. So an L to R link (adjacency) between sounds 1 and 2 is taken as equivalent to R to L, but its intensity is considered: the frequency of occurrence in either direction. Because of the sound group size unevenness and the networks observed, we sometimes subdivide the Environmental into Natural and Machine/Human categories (defined in Table 1). Figure 6 shows the descriptive networks of the two tasks. Note that each link intensity is represented by the thickness of its line, but 2D layout is algorithmically chosen to optimise clarity, and has little bearing on the interpretation of the network.

In Figure 6 the Impulse response and DSP groups are relatively discrete in the networks, while Instruments and Convolution are interspersed, as are Instrument, Machine and Natural Environment in Task 2/3. This is partly due to sound 28, a ‘waterphone’ instrument that produces complex rapidly changing timbres like the sound of a stream of water. This sound could be classified as ‘Machine’ (but there is a sense in which all musical instruments are machines). As emphasised, the most objective feature of the representation is the edge width, proportional to the link strength. For example, note that some of the strong edges within the DSP in Task 1 are weakened in Task 2/3, as are some within the Impulse response group. There are also some stronger edges for Task 2/3 than 1, such as 13–25 and 5–30. Overall, these network representations confirm and display the distinctiveness of Tasks 1 and 2/3.

The fundamental features of each node in a network include not only the intensity of its connection to other nodes, but also its ‘degree’ of interconnectedness (the number of distinct connections it makes). Many important network features depend on further consideration of its linkage structure, such as detecting ‘communities’: do user perceptions indicate sound communities corresponding to or different from the design groups? A community is a group of nodes that are more connected internally than to the rest of the network. These further features suggest mechanisms that may be responsible for network generation. For example, if items A and B are strongly linked, and so are B and C, then this is quite likely to result in the formation of a link between A and C, as we know from our own experiences of social networking (this is termed the triangular relation in models). Such analyses identify community distinctions between the tasks, and are not simply congruent with the design groups. Figure 7 shows communities detected by the so-called label propagation algorithm (note the different colouring system from Figure 6). The label propagation algorithm (see Betzel and Bassett 2017) assigns randomised labels to all nodes to commence, and then by progressively (empirically) changing the labels of connected nodes (e.g., applying the same label to the members of a linkage triangle) and assessing the outcome, it maximises the number of nodes which are directly connected to nodes with the same label, and in this case gave stable results.

Figure 7 suggests that in both tasks communities are formed by the impulse responses (sounds 10–13), DSP (sounds 1, 7, 21, 23, 26), and the mixture of convolutions (3, 4, 5, 6) with all but one of the instruments (14–17). The remaining (mainly Natural and Machine environmental) sounds form two communities in Task 1, and are interspersed as one in Task 2/3. Figure 7 is based on label propagation, but other aspects of the connections and weights can be used for community detection (Betzel and Bassett 2017). Among these, detection by ‘greedy optimization of modularity’ gives the same patterns as Figure 7. ‘Community betweenness’ is an alternative criterion for community discrimination, maximising the difference between intra-community and inter-community edge frequencies. For Task 1, it indicates 10 communities, nine of which contain a single item, while for Task 2/3 it indicates nine, seven of which have single members (four shared with Task 1), while one contains four, and the final large community contains the remaining 19. Thus diverse community analyses again confirm the distinctness of networks from the two tasks (for a detailed comparative review of community detection methods see Yang, Algesheimer and Tessone 2016).

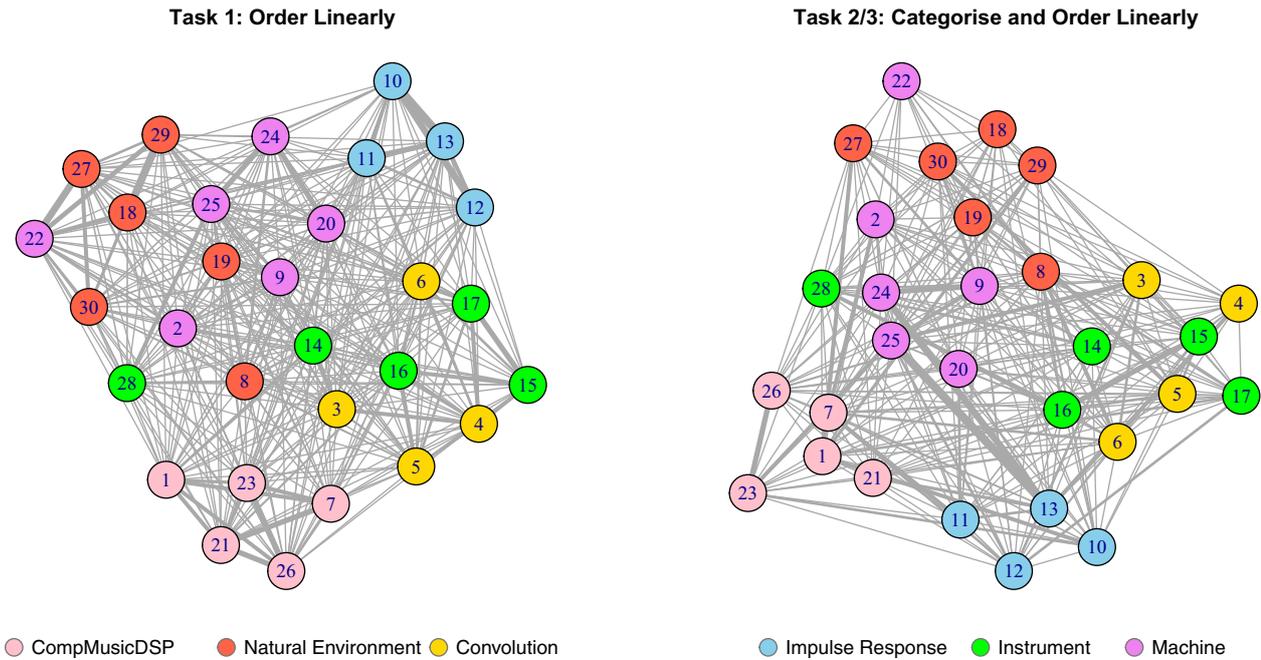


Figure 6. Descriptive networks for Tasks 1 and 2/3.

Note. Each node represents a numbered sound. The configurations use the Fruchterman–Reingold force-directed algorithm to maximise legibility. The graph shows the interspersing of the design groups, and the frequencies of node links are represented by edge widths.

Task1: Communities based on label propagation

Task2/3: Communities based on label propagation

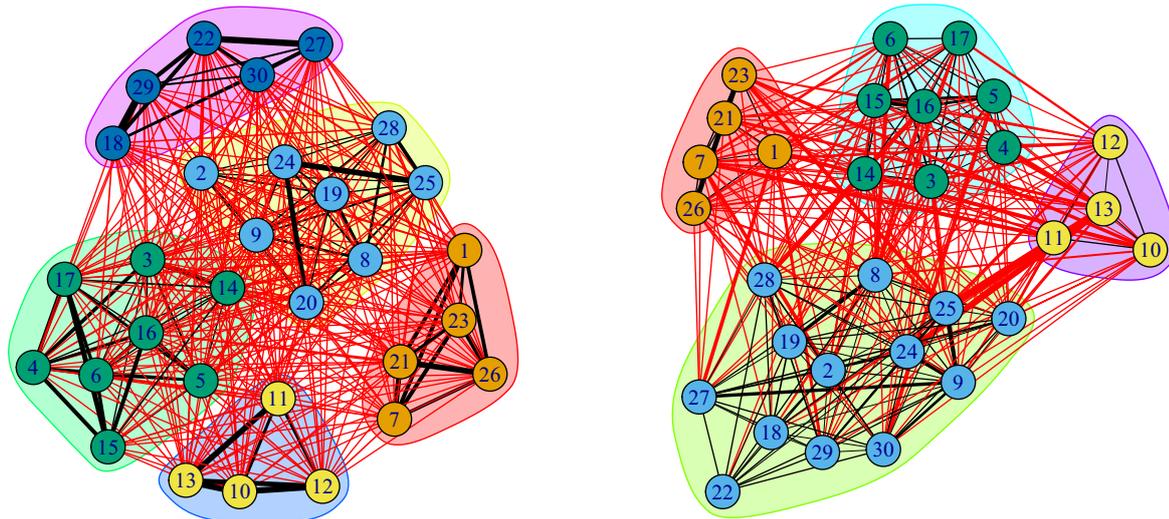


Figure 7. Communities based on label propagation for Tasks 1 and 2/3.

Note. Sounds are numbered as before, but the colouring systems of both background and vertices solely reflect the identified communities. Within community edges are black, between are red.

Table 3 compares some important network statistics, supporting their distinctiveness.

Besides differentiating the networks, the mean degree parameters (the mean number of other nodes to which a node is connected) indicate both networks are profuse (19.5 and 24.3 out of a maximum

attainable of 29): whereas the evolutionary popular music networks (Serrà et al. 2012) only have comparatively sparse degree medians of four for pitch classes, and eight for timbre. While timbre contrasts may be ‘rarely the basis for musical discourse’ (Serrà et al. 2012) in popular music, this is clearly not true of

Table 3. Network statistics for Tasks 1 and 2/3

Network:	Edge density	Transitivity	Diameter	Mean degree	Centrality	Mean distance
Task 1	0.84	0.84	3	24.3	0.16	1.16
Task 2/3	0.67	0.68	4	19.5	0.15	1.32

Note. Edge density describes the proportion of possible edges that are occupied, and transitivity is closely related (there are 435 possible undirected edges in these networks). Other measures are also closely interrelated. Diameter represents the maximum distance (number of edges) required to get from any node to any other node. Mean degree summarises the number of nodes connected to any one other node. Centrality is a global betweenness measure, while mean distance represents that between pairs of nodes on average. Note that in principle, part of these differences between Tasks 1 and 2/3 might flow from the fact that 10% less edge counts are provided by the latter task (as discussed earlier): again, comparisons based on randomly removing 10% of participants from Task 1 suggest any such effect is minimal.

our sounds and our experiment, nor likely to be true generally for EAM.

4.5. Generative network models: modelling the two different networks

Such models seek to explain network generation, rather than simply describe networks; for example, as mentioned earlier, the formation of ‘triangular’ node sets can be a generative path. Exponential random graph models (ERGM) are a widely used class of generative network models (Betzel and Bassett 2017). Given clear criteria of connection, describing a network is unambiguous. In contrast, generative models have to be assessed for relative ‘goodness of fit’ (GOF: the degree to which they reflect the properties of the data); and GOF may assess a selection of features, not just link frequencies. Ideally, a generative model would be assessed not only by comparing its modelled predictions (i.e., the features that it explicitly sought to explain) with those observed, but also by comparing unmodelled predictions with those observed. In these ways models can be compared such that the best available can be selected, and some of its limitations understood.

For valued nets, such as those studied here, the R ERGM-count package (from within the statnet package) does not yet provide a GOF test which goes beyond the predicted variables. In contrast, with simpler non-valued nets one can also test whether the generative model produces nets whose statistics coincide with the unmodelled parameters, such as some of those in Table 3. Thus, the models of Table 4 (valued nets) achieved good fit solely to the modelled variables, and were optimised by means of residual deviance and information criteria (representing precision and efficiency of the model, where efficiency is a measure of predictive power per predictor variable used, with a preference for simpler models with less predictors given equal precision). Some further technical details are given in the Table 4 notes.

Table 4 models the distinctness of the two networks, particularly in terms of the coefficient on

nodesqrtcovar, the index of the individuality of the behaviour of individual sounds. The quantitative interpretation of the estimated values is as follows: the odds determining the likely weight between any pair of sounds in Net 1 model is $\exp(-2.13)$, which is 0.12. That determining the weights between any pair of Impulse responses is $\exp(2.41)$, that is, 11.1 times higher. The corresponding figures for Net 2/3 are 0.10 and 18.5. While the general pattern of coefficients is similar for the two nets, nodesqrtcovar confirms the substantially lesser diversity of the weights of edges in Net 2/3, in accord with our hypothesis of transforming conceived relationships between sounds in the two task conditions.

4.6. Modelling the temporal transformation of the timbral networks

The R statnet term package allows modelling a *temporally changing* network on the basis of two (or more) distinct successive samples of it, which is essentially what we can take Net 1 and Net 2/3 to be. Term for valued networks is not readily feasible presently, but we determined models of the edge structure of unvalued Nets 1 and 2/3 to be of the form $\text{net} \sim \text{edges} + \text{triangles}$ (as noted earlier ‘triangles’ is a term representing a simple form of triadic closure). Nodematch terms (concerning whether or not an edge links members of a single sound design group) were not effective in these unvalued models (cf. Table 4 for the valued version). A term model includes both an edge formation and a dissolution model (since with time, both processes can occur), and Table 5 summarises our optimised model. Note that the deviance and information criteria values of this table cannot be compared with those of Table 3. This model met GOF criteria both for modelled and unmodelled (e.g., degree) variables.

As noted, Net 1 has 364 edges, Net 2/3 293 (in each case out of a possible maximum of 435), and these numbers give rise to the degrees of freedom of the Formation and Dissolution models. Taking these two nets as one step apart (with Net 1 first,

Table 4. Exponential random graph models for the valued networks of Tasks 1 and 2/3

Formula for both models: $\text{net}(1 \text{ or } 2/3) \sim \text{sum} + \text{nodesqrtcovar}(\text{center} = \text{TRUE}) + \text{transitiveweights}(\text{'min'}, \text{'max'}, \text{'min'}) + \text{nodematch}(\text{'type'}, \text{diff} = \text{TRUE}, \text{form} = \text{'sum'})$. Both models show Monte Carlo Maximum Likelihood Estimate results.

	Estimate	SE	z value	Pr(> z)
Task 1				
sum	-2.13	0.06	-34.03	< 1e-04 ***
nodesqrtcovar.centered	-7.89	1.61	-4.90	< 1e-04 ***
transitiveweights.min.max.min	-0.24	0.05	-4.39	< 1e-04 ***
nodematch.sum.type.Mach	0.87	0.14	6.37	< 1e-04 ***
nodematch.sum.type.Conv	1.08	0.19	5.70	< 1e-04 ***
nodematch.sum.type.Natural	0.89	0.14	6.58	< 1e-04 ***
nodematch.sum.type.Imp	2.41	0.17	14.37	< 1e-04 ***
nodematch.sum.type.Inst	0.59	0.18	3.38	0.00074 ***
nodematch.sum.type.RTD	2.02	0.13	15.16	< 1e-04 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Null Deviance: 0 on 435 degrees of freedom; Residual Deviance: -9036 on 426 degrees of freedom. AIC: -9018 BIC: -8981

	Estimate	SE	z value	Pr(> z)
Task 2/3				
sum	-2.32	0.05	-40.56	< 1e-04 ***
nodesqrtcovar.centered	-17.71	2.06	-8.58	< 1e-04 ***
transitiveweights.min.max.min	-0.28	0.04	-6.19	< 1e-04 ***
nodematch.sum.type.Mach	1.07	0.13	8.39	< 1e-04 ***
nodematch.sum.type.Conv	1.30	0.18	7.10	< 1e-04 ***
nodematch.sum.type.Natural	1.22	0.13	9.66	< 1e-04 ***
nodematch.sum.type.Imp	2.92	0.19	15.60	< 1e-04 ***
nodematch.sum.type.Inst	0.67	0.17	4.05	< 1e-04 ***
nodematch.sum.type.RTD	2.29	0.13	17.05	< 1e-04 ***

Null Deviance: 0 on 435 degrees of freedom; Residual Deviance: -10081 on 426 degrees of freedom. AIC: -10063 BIC: -10026
Note. The formula indicates that the model predicts the basic network edge weights by means of the specified variables. Estimate: the determined coefficient of the specified predictor in the model formula; SE, standard error (of estimate): SE that are low in relation to the Estimate indicate a reliable estimate, and the final Pr(probability) value indicates how likely it is the result could occur by chance rather than as modelled; AIC, Aikake's Information Criterion; BIC, Bayesian Information Criterion – in all cases lower values are better, and these values are used in model selection, but the two task models cannot be compared in this respect as explained below. All parameters are highly significant as indicated by the triple asterisks.

All 'sum' parameters refer to the sum of the relevant edge weights (the modeled response parameter for these valued networks, while *nodematch...* refers to the design group of the sound, with the same abbreviations at the end of the name as before, in this case considering six groups. Here, *nodesqrtcovar.centered* is an index of the distinctiveness of the behaviour of the individual nodes; while *transitive weights* is a triadic closure measure for valued nets, analogous to the concept of edge 'triangles' in non-valued nets.

as it was produced by the participants) in the possible transformation of the temporal network, one can simulate a series of steps from the model of Table 5. Starting with Net 1, and applying the generative model of Table 5, each and every single transformation step creates and dissolves multiple edges. The product of one step can be then used sequentially in a further transformation. Over multiple such simulations, one can assess the cumulative network adoption of the possible edges: this showed that all 435 possible edges always occurred at least once in the first nine networks, often earlier. When the two networks were treated as if in the reverse order (which could readily occur), with the categorisation/

ordering task (2/3) first so that the number of edges rises between the two time slices, simulating with the different resultant tergm model always showed full coverage of all 435 edges even within six time steps. Thus tergm confirms the potential of the modelled generative dynamic network process representing participants' possible developing concepts of inter-sound linkages to rapidly survey all the possible linkages. This again supports our proposal that listeners (including composers) can readily conceive diverse and changeable relationships between timbres. We wished to avoid categorisation as a first step in our experiment (since it would likely drastically influence a subsequent unilinear ordering

Table 5. A separable temporal exponential random graph model of unvalued Nets 1 and 2/3 Formation model. Formula: net ~ edges + triangles Monte Carlo MLE Results

	Estimate	SE	z value	Pr(> z)
edges	1.59	3.65	0.436	0.66
triangle	-0.06	0.15	-0.41	0.69

Null Deviance: 98.43 on 71 degrees of freedom. Residual Deviance: 98.13 on 69 degrees of freedom. AIC: 102.1; BIC: 106.7

	Estimate	SE	z value	Pr(> z)
edges	-0.32	0.38	-0.84	0.40
triangle	0.12	0.04	3.12	0.002 **

Null Deviance: 504.6 on 364 degrees of freedom. Residual Deviance: 435.3 on 362 degrees of freedom. AIC: 439.3; BIC: 447.1. The probabilities show that this model was less strong than those of Table 5.

task). On the other hand, doing the categorisation step first, contrary to our experiment, may well be a composerly approach (Dean 2017).

5. GENERAL DISCUSSION

The data confirm the distinctness of the two task orderings. Furthermore, the generative network analyses indicate that sequential transformation of the networks can cumulatively explore every possible link between sounds very efficiently, that is within nine or less time steps. This suggests the flexibility with which our participants can conceive diverse inter-timbre relationships even with uncommon sounds. An EAM composer can reasonably expect that listeners can adjust to the compositional relationships they foster.

Is this conceptual flexibility special to music? One aspect that might suggest this is the fact the orderings we required do create temporal sequencing when a participant listens to their sounds in their chosen order, temporal sequencing being of the essence of music. But we would expect that conceptual flexibility would be found in most contexts. Artistic areas of endeavour again suggest this; for example, abstraction in painting and 3D-arts is widely appreciated, but requires the formation of conceptual links between novel objects (which may also be temporal and kinetic in the case of video art). An experimental approach such as ours might therefore be of general use, and we should perhaps not be surprised by conceptual flexibility with sound orderings. Nevertheless, its demonstration has importance for composers and for the potential for affective meaning generation

and reception in music according to the way that different timbral relationships are arranged temporally.

A critique of sound distance mappings, but also of relationship mappings such as defined here, is that they compress cognitive relationships. This is probably of greater concern when viewing the relationships as fixed than when seeking to consider their flexibility. Also of interest are the mentioned roles of pitch and loudness perceptions in the organisation. These are likely to be significant, but arguably, virtually inseparable from ‘timbral’ assessments. Indeed, pitch is not usually necessarily salient in many of our sounds, and can be construed as a compression of timbral features, while loudness is a perception dependent strongly on the spectral distribution of physical energy. The spectral energies of our sounds varied across the sound, as is normal. Thus the only way we could limit the influence of loudness was to standardise RMS: had we chosen perceptual equalisation of loudness of samples per se, this would have distorted the relative impact of the timbral (spectral) components.

Another aspect of our sounds may be relevant: a listener may automatically seek to identify a sound source, and sometimes be successful. Acousmatic EAM has taught us to not always expect detectable agency, but a biological tendency to assess the origin of a sound has behavioural value. Our participants’ description of their methods often implied less holistic approaches (e.g., focusing on loudness), but the use of a distinction between ‘music’ and other sounds (mentioned by some) does indicate seeking the origin of a sound. Similarly, the co-boxing of instruments and their digital transforms would be consistent with a consideration of sound sources.

Also relevant are possible sequential effects within the experiment. We provided the tasks in the fixed order 1, 2, 3 so as to change and increase task demand after Task 1. It would be interesting to perform repetitions of orderings and categorisation. For Task 1 we predict variation in orderings (i.e., conceptual flexibility). While random orderings would not be of interest, nor would test–retest reliability be pertinent. Test–retest reliability tests how secure a verdict is, usually when multiple evaluators need to be harmonised. These tests are used when there is assumed to be a ‘true’ verdict: this does not hold here.

Because the experiment concerned conceptual flexibility, some possible analytical approaches have been excluded. For example, one could assess the acoustic features of the individual sounds as predictors of relationships. But here we postulate changing relationships between pairs of sounds, and hence a single set of acoustic features may not predict this. Instead, we chose to consider design groups and their

Computing the transposition distance between two (arbitrary) orderings

Task 1 ordering:

1 2 3 ... 28 29 30

Task 2/3 suborders in the four boxes:

1 21 10 27 28 22 2 5 11 29 23 7 15 30 3 8 9 16 17 20 14
24 4 6 18 25 26 19 12 13

One of the permuted Task 2/3 orders (boxes 1, 3, 2 4, reading L to R):

1 21 10 27 28 22 3 8 9 16 17 20 14 24 2 5 11 29 23 7 15
30 4 6 18 25 26 19 12 13

Examples of transposition distances for individual sounds:

Sound 1: Task 1 position 1, Task 2/3 position 1, distance 0

Sound 11: 1 position 11, 2/3 position 17, distance 6

Sound 27: 1 position 27, 2/3 position 4, distance 23

influences on network structure (Table 4, nodematch parameters). One approach to accommodating acoustic features would attach them to each node, and construct a model considering both node values and edge weights. One could also include as a feature of each node the second- and higher-order links, but the Markov analysis did not suggest orders above 1, so this seemed unlikely to be informative.

Approaching timbres as conceptual networks is not apparent in the psychoacoustic literature. One partial exception is the demonstration of ‘analogies’ between MDS-derived relationships between pairs of sounds, at least implying relationship networks (McAdams and Cunible 1992). However, in the computational acoustic feature networks between pieces of pop (Serrà et al. 2012), progressive homogenisation is observed, so here acoustic diversity and association does not seem a key expressive tool. Indeed, in contrast to the present work, pop-music feature networks seem sparse: each item is connected to 0 or 1 other. Serrà et al (2012) argue that pop seeks conventionality to be accessible.

Whereas network analysis of sounds in a piece or compositional process has apparently not been undertaken previously, constructing pieces on this basis is well known, as elaborated by Trevor Wishart (1985, 1994), and discussed more recently (Cancino 2013). Data such as here could be a compositional tool for choosing novel networks while making EAM and sound art (where environmental sounds are common).

What are the benefits for a composer of conceiving timbral relationships as flexible within a network? Most obvious perhaps is the possibility of surveying sounds that might follow a current sound by pursuing the edges of a network according to their intensity (or in a converse manner to achieve novelty), and within a community (or outside it in order to achieve novelty). Non-linearity of progression, an important idea in writing at least since hypertext, and music for a

comparable period become immediately accessible (reviewed Smith and Dean 1997). Indeed, the tasks undertaken by our participants are similar to those a digital music composer might apply, though a composer might well use Tasks 2/3 preceding or instead of 1. Creating a soundscape or EAM piece does often involve creating a sequential montage (though with overlaps), and individual sound objects within this are always moved around in an experimental-aesthetic editing process that takes advantage of the composer’s conceptual flexibility, and their ability to link this to ideas of transmissible expressive potential. The results here support the view that a composer’s organisation of the sounds is cognitively approachable to untrained listeners.

Here we have considered our results as indicating a cognitive action by listeners, hence the word ‘conceptual’. Our results on conceptual flexibility suggest that the use of ordering tasks such as here, or other tasks that force several different appraisals of the relationship between a group of timbres, might influence behaviour in subsequent paired distance rating tasks. It would be reasonable to predict consequent changed distance perceptions both between sounds used in an appraisal task such as ours, and between other unappraised sounds with some acoustic relation to them. Such a study remains for future attention.

Acknowledgement

Most of this work was presented at the Society for Music Perception and Cognition, July 2019, New York. We appreciate the excellent research assistance of Farrah Sa’adullah. The project (H10664) was approved by the Human Research Ethics Committee of Western Sydney University, and participants provided written informed consent. Valuable presentation comments from our reviewers were much appreciated.

REFERENCES

- Bailes, F. and Dean, R. T. 2012. Comparative Time Series Analysis of Perceptual Responses to Electroacoustic Music. *Music Perception* **29**(4): 359–75.
- Betzel, R. F. and Bassett, D. S. 2017. Generative Models for Network Neuroscience: Prospects and Promise. *Journal of the Royal Society Interface* **14**(136): 20170623.
- Bregman, A. S. 1994. *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Caclin, A., McAdams, S., Smith, B. K. and Winsberg, S. 2005. Acoustic Correlates of Timbre Space Dimensions: A Confirmatory Study Using Synthetic Tones. *The Journal of the Acoustical Society of America* **118**(1): 471–82.
- Cancino, J. P. 2013. The Agile Musical Mind: Mapping the Musician’s Act of Creation In T. Veale, K. Feysaerts and

- C. Forceville (eds.) *Creativity and the Agile Mind*. Berlin: de Gruyter. 335–54.
- Dean, R. T. 2017. *Creating Music: Composition. The Routledge Companion to Music Cognition*. London: Routledge.
- Dean, R. T. (ed.) 2009. *The Oxford Handbook of Computer Music*. New York: Oxford University Press.
- Dean, R. T., Bailes, F. and Dunsmuir, W. T. M. 2014a. Shared and Distinct Mechanisms of Individual and Expertise-Group Perception of Expressed Arousal in Four Works. *Journal of Mathematics and Music* 8(3): 207–23.
- Dean, R. T., Bailes, F. and Dunsmuir, W. T. M. 2014b. Time Series Analysis of Real-Time Music Perception: Approaches to the Assessment of Individual and Expertise Differences in Perception of Expressed Affect. *Journal of Mathematics and Music* 8(3): 183–205.
- Grey, J. M. and Gordon, J. W. 1978. Perceptual Effects of Spectral Modifications on Musical Timbres. *The Journal of the Acoustical Society of America* 63(5): 1493–500.
- Johnson, E. K. 2016. Constructing a Proto-Lexicon: An Integrative View of Infant Language Development. *Annual Review of Linguistics* 2: 391–412.
- McAdams, S. 1993. Recognition of Sound Sources and Events. In S. McAdams and E. Bigand (eds.) *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Clarendon Press. 146–98.
- McAdams, S. 2004. Problem-Solving Strategies in Music Composition: A Case Study. *Music Perception: An Interdisciplinary Journal* 21(3): 391–429.
- McAdams, S. and Cunible, J.-C. 1992. Perception of Timbral Analogies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 336(1278): 383–9.
- McAdams, S. and Siedenbueg, K. 2019. Perception and Cognition of Musical Timbre. In P. J. Rentfrow and D. J. Levitin (eds.) *Foundations in Music Psychology: Theory and Research*. Cambridge, MA: MIT Press. 71–120.
- McAndrew, S. and Everett, M. 2015. Music as Collective Invention: A Social Network Analysis of Composers. *Cultural Sociology* 9(1): 56–80.
- Müllensiefen, D., Gingras, B., Musil, J. and Stewart, L. 2014. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PloS one* 9(2): e89642.
- Olsen, K. N., Dean, R. T. and Leung, Y. 2016. What Constitutes a Phrase in Sound-Based Music? A Mixed-Methods Investigation of Perception and Acoustics. *PloS one* 11(12): e0167643.
- Pennington, J., Socher, R. and Manning, C. D. 2014. GloVe: Global Vectors for Word Embeddings. *Proceedings of EMNLP*.
- Rushton, J. 2001. Klangfarbenmelodie. In S. Sadie and J. Tyrrell (eds.) *Grove Dictionary of Music and Musicians*, 2nd edn. London: Macmillan. 652.
- Serrà, J., Corral, Á., Boguñá, M., Haro, M. and Arcos, J. L. 2012. Measuring the Evolution of Contemporary Western Popular Music. *Scientific Reports* 2: 521.
- Smith, H. and Dean, R. T. 1997. *Improvisation, Hypermedia and the Arts since 1945*. London: Routledge.
- Wishart, T. 1985. *On Sonic Art*. York: Imagineering Press.
- Wishart, T. 1994. *Audible Design. A Plain and Easy Introduction to Practical Sound Composition*. York: Orpheus the Pantomime.
- Xenakis, I. 1971. *Formalized Music*. Bloomington: Indiana University Press.
- Yang, Z., Algesheimer, R. and Tessone, C. J. 2016. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports* 6: 30750.
- Zacharakis, A., Terrell, M. J., Simpson, A. J., Pasiadis, K. and Reiss, J. D. 2017. Rearrangement of Timbre Space Due to Background Noise: Behavioural Evidence and Acoustic Correlates. *Acta Acustica United with Acustica* 103(2): 288–98.