From music to dance: the inheritance of semantic inferences

Pritty Patel-Grosz • Jonah Katz • Patrick Georg Grosz • Tejaswinee Kelkar • Alexander Refsum Jensenius

Abstract This paper looks at short musical segments and motion-capture animations of body movements that were generated spontaneously in response to those musical segments. Building on recent research on music semantics, we ask whether abstract meaning inferences that listeners draw on the basis of the musical segments are also inherited by the corresponding body movements. We present an experiment in which participants rate how well the emotion terms Angry, Bored, Calm and Excited are expressed by the auditory stimuli and visual stimuli. The experimental findings indicate a correlation between the sounds and animations with regards to the inferences that participants draw.

Keywords Music semantics · Dance semantics · Iconic semantics · Formal semantics · Experimental semantics

P. Patel-Grosz, University of Oslo, pritty.patel-grosz@iln.uio.no
J. Katz, West Virginia University, jokatz@mail.wvu.edu
P. Grosz, University of Oslo, p.g.grosz@iln.uio.no
T. Kelkar, University of Oslo, tejaswinee.kelkar@imv.uio.no
A. Jensenius, University of Oslo, a.r.jensenius@imv.uio.no

In Gabriela Bîlbîie, Berthold Crysmann & Gerhard Schaden (eds.), Empirical Issues in Syntax and Semantics 14, 0–00. Paris: CSSP. http://www.cssp.cnrs.fr/eiss14/

1 Overview

Music can give rise to abstract semantic inferences about music-external situations and/or emotions. We ask whether dance, defined as music-accompanying body movement for the scope of this paper,[1] also gives rise to similar abstract semantic inferences. Focusing on emotional meaning, we experimentally test whether inferences from a given musical sequence are inherited by body movement produced in response to this musical sequence. Our results indicate that such an inheritance of semantic inferences

[1]This glosses over dance that does not accompany music, see, e.g., Patel-Grosz et al. (2018; to appear).

does occur. This finding is consistent with a view where abstract meaning can be communicated in different modalities, here music and dance.

2 Theoretical underpinnings – from music to dance semantics

Recent research in formal semantics argues that music can give rise to inferences about music-external objects (so-called *virtual sources* or *denoted objects*), which allow listeners to infer descriptive or narrative meaning, Schlenker (2017; 2019a; 2021). A typical example is found in Saint-Saëns's Carnival of the Animals, where a low-pitched melody is mapped onto a large object, namely an elephant (Schlenker 2017: 11-12). By contrast, a high-pitched version of the same melody may instead give rise to the inference that there is a small object in the narrative – for example, a mouse. Such inferences are iconic in that the denotation of the meaning-bearing object – in this case the music – operates on its form. To illustrate, we can apply Greenberg's (2021) formalism to the above example and posit the iconic semantics in (1) for object-denoting pitch.

(1) For a piece of music $\boldsymbol{M}$, a constant k in a narrative situation s, $[\![\boldsymbol{M}]\!]$ is satisfied by s only if:
size-category($\iota x.x$ is an object in s) = k / pitch($\boldsymbol{M}$)

For Greenberg, an iconic semantics is defined such that the form of the sign, symbolized by the bold-typed $\boldsymbol{M}$ in (1), also occurs in its denotation. When we interpret a piece of music M with regards to a narrative situation s, we can draw an inference that the pitch of M is inversely mapped onto the size of a salient object in s. The higher the pitch, the smaller the object. This inverse mapping of pitch and size can be implemented by dividing a contextually given constant k by the pitch of M, which yields an abstract numerical size category. Assuming size categories rather than exact sizes is needed to account for the abstractness of such inferences. A similar difference between a high pitch and a low pitch (e.g., 880 Hz vs. 110 Hz) may be mapped onto an elephant (the large object) vs. a mouse (the small object), but it may just as well be mapped onto a hawk (the large object) vs. a sparrow (the small object), or onto a landscape (the large object) vs. a house (the small object). If we take our constant k to be 880, then a pitch of 880

places the denoted object into size category 1 (≈ small), whereas a pitch of 110 places the denoted object into size category 8 (≈ big).

When all inferences of this type are met for a situation *s* in a given narrative, then we can say that ⟦M⟧ is true in *s* (or ⟦M⟧ is satisfied by *s*). This means that a low-pitched melody is true of a narrative situation in which we are dealing with a large object, and false of a narrative situation in which we are dealing with a small object (relative to a baseline of what counts as large or small in the narrative). Such inferences are by their very nature abstract, i.e., it does not matter whether the object is an elephant, a landscape, or, even more abstractly, a magnificent idea.

Crucially, the properties of music that give rise to such iconic inferences (pitch, loudness, speed, silence, dissonance, change of key; see Schlenker 2019b:433-436) have counterparts in music-accompanying movement, for example dance. An observation from choreomusicology suggests that musical pitch corresponds to the direction of gestures in space in body movement, in that, for example, high pitch is correlated with upward movement whereas low pitch is correlated with downward movement (Mason 2012: 10); see Kelkar & Jensenius (2018) for critical discussion. Alternatively, Gadir (2014: 55), in a study of electronic dance music, observes that the musical pitch inversely correlates with the size of the body parts that dancers use to move to the music, i.e., lower pitch is correlated with bigger movements (of the legs, hips, etc), whereas higher pitch is correlated with smaller movements (of the arms, hands, etc). We can refer to this difference in terms of the amplitude of the body movement. We can thus posit the lexical entry in (2) for a given body movement *D*, which differs from (1) in that the size of the denoted object is now calculated by multiplying (rather than dividing) the constant *k* with the amplitude of the dance movements.

(2) For a music-accompanying movement ***D***, a constant *k* in a narrative situation *s*, ⟦D⟧ is satisfied by *s* only if:
size-category(*ιx.x* is an object in *s*) = *k* * amplitude(***D***)

The fact that (1) and (2) are non-isomorphic is unsurprising, since musical inferences are typically attributed to natural associations between sound qualities and their sources (Schlenker's 2019b:433-436 *reasons* for musical inferences). Bigger objects tend to create sounds at a lower pitch than

smaller objects. For body movements, the same association naturally does not hold: bigger objects will often create bigger movements than smaller objects (in terms of the absolute amplitude of the movement, not relativized to the given object's baseline), simply by virtue of their size. This may in turn have implications for the perception of body movement perceived to communicate meaning: if such a body movement is bigger, onlookers plausibly infer that the denoted object is also bigger, and so forth.

We take such analogies between music and body movement as our point of departure and present an experimental study that addresses the following questions: (i.) do abstract body movements (e.g. dancing or moving spontaneously to a piece of music) give rise to semantic inferences? (ii.) are there parallels between the inferences that we draw from hearing music, and the inferences that we draw from seeing abstract body movement? (iii.) if we perceive body movement *D* that was initially performed as an interpretation of a short musical sequence *M*, is there a correspondence between our inferences from *D* and our inferences from *M*?

To investigate these questions, we used materials from an experiment by Kelkar & Jensenius (2018).[2] In their study, participants were asked to trace short musical segments with their hands while standing centrally in a motion capture lab and being filmed by eight motion capture cameras; this gave rise to abstract music-accompanying body movement (i.e., dance), consisting in particular of upper-body movements. More specifically, we could classify the movement as a music-*responsive* body movement, in that it was caused by the music. Our own experiment is set up to test the hypothesis that body movements *D* which are performed in response to a musical sequence *M* 'inherit' properties of *M*, thus giving rise to the same or similar semantic inferences. We can illustrate this question for our toy example in (1)-(2): assume that a musical sequence *M* triggers the inference that the denoted object is big; the question is then whether a body movement *D* that was evoked by *M* would also trigger the inference that the object is big. Specifically, would a lower pitch of *M* trigger more expansive body movements in *D*?

Two qualifications are in place before we proceed with the discussion of our experiment. The first concerns the distinction between perception and

[2]See Kelkar (2019) for more detailed discussion of the stimuli.

production of music and dance; the second concerns the nature of inferences that we investigate.

In our experiment, we probe participants' perception of musical sequences and body movements. While both are produced intentionally (in the case of our stimuli), it is not necessarily the case that their producers consider the inferences that listeners and onlookers may draw. To use our previous example of musical pitch, a composer may intend for a low pitch melody to symbolize a large object, but may also lack any such intention; listeners would draw the exact same inference in either case – namely that the music describes a large object. The same reasoning applies to body movements.

As for the nature of the inferences, we aimed to test whether participants converge on a given meaning for a given stimulus, and whether there is a correspondence between the meaning that was assigned to a musical sequence and the body movement that ensued from it. The meanings that we tested in our experiment are the meanings associated with the emotion terms *Angry*, *Bored*, *Calm* and *Excited*. We used emotion terms as opposed to concrete properties such as size (cf. the toy example in (1)-(2)), to avoid participants directly interpreting properties of the music or movement; furthermore, there is a precedent of probing emotive meanings in music and movement in the findings of Sievers et al. (2013). It is worth emphasizing that, by design, this task does not directly probe for the descriptive/referential musical inferences proposed in Schlenker (2019b); these are more physical in nature, amounting to a description of the denoted object as big or small, more or less energetic, closer or further away, etc. As a consequence of this methodological choice, our conclusions apply primarily to emotional meaning, and it is an open question whether they carry over to referential semantics.[3]

3 Experimental design

3.1 Stimuli

Owing to the nature of materials from Kelkar & Jensenius (2018), we depart from toy inferences of the type in (1) and (2). We use six combinations of

[3]On the difference between emotional meaning and referential meaning, see, e.g., Meyer (1956), Patel (2008), and Juslin (2013). See Schlenker (2017:28-33, 2019a:86-95) on how they may be connected.

short musical sequences (between 1.45 seconds and 5.0 seconds in duration) and video animations of motion capture renderings of movements carried out to accompany those sequences by the study participants of Kelkar & Jensenius (2018). Our musical sequences thus correspond to the original experimental stimuli of Kelkar & Jensenius, one of which is illustrated in Figure 1.

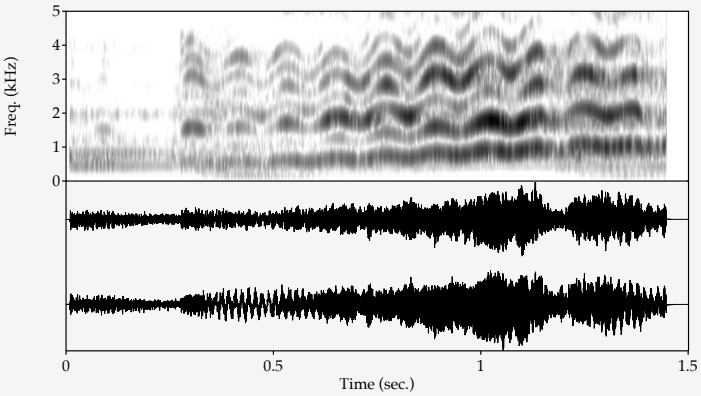

Figure 1 Waveform (stereo) and spectrogram from experimental item b, an operatic soprano voice singing an ascending major-scale melisma on an [e]-like vowel over a much quieter piano accompaniment.

Our visual stimuli were selected from the Kelkar & Jensenius's experimental results, in order to create music-video pairings. The type of motion capture-based animations used in our experiment is illustrated in Figure 2, which is a movement sequence that was produced in response to the musical stimulus in Figure 1.[4]

The six combinations of auditory and visual stimuli that we used were selected pseudo-randomly from the set of 32 combinations available to us. While we tried to avoid stimuli that made a particular emotion especially

[4]The complete set of visual stimuli can be found at the following link: `https://www.youtube.com/playlist?list=PL0dCnZzwa9N4aVfOTZce-WHVMPg5dYbdI` The auditory stimuli correspond to the auditory stimuli 02, 13, 19, 27, 29, and 30 in Kelkar & Jensenius (2018) (direct download link for ZIP file: `https://www.mdpi.com/2076-3417/8/1/135/s1?version=1516685118`), renumbered to 01, 02, 03, 04, 05, and 06 for our experiment.

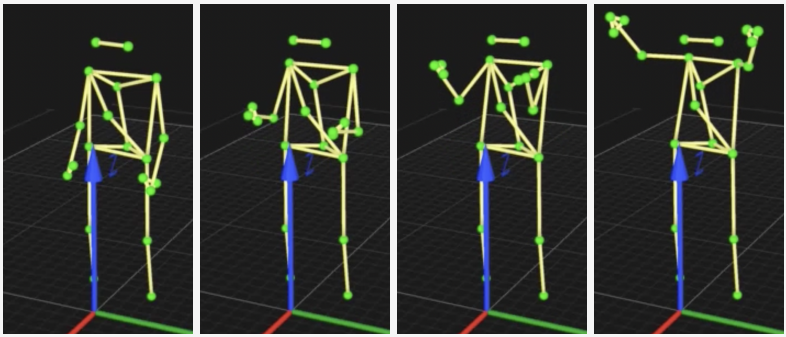

Figure 2 Stills of animations created from motion-capture data from experimental item 2.

salient, we did not have any evidence or prior expectation with regards to the emotion association of our six combinations; they were thus not counterbalanced with respect to emotion association.

All participants listened to the six sound files and separately watched the six silent animations; participants did not watch combinations of animations and sound files. Stimuli were organized in two blocks, one with only audio stimuli and one with only silent animations. The order of the two blocks was counterbalanced across subjects and stimuli were presented in random order within each block. Subjects heard or saw each stimulus four times, once with each of the emotion descriptors described below.

We used a within-subject design to maximize the power of the experiment. In this design, each subject watches each video stimulus and hears each audio stimulus multiple times, once with each emotional descriptor. This allows us to gather more data from fewer subjects, and to isolate the audio vs. video modality as an experimental manipulation without also changing the identity of subjects between conditions. The main potential drawback to such designs is the possibility of 'contamination', where completing one condition influences the behavior of a subject on following conditions. To adjust for such effects, we counterbalanced the order of conditions across subjects and incorporates the effect of task order into the statistical analysis reported below.

3.2 Task

Participants were prompted to rate on a slider scale from 0 to 100 how well the sound / animation expressed one of the following emotions: *Angry*, *Bored*, *Calm* and *Excited* – with 1 trial for each emotion (2x6x4 trials in total). Participants were able to play the sounds and animations as many times as they desired. The experiment took about 15 minutes to complete. The emotion terms are based on the four quadrants of Russell's (1980) circumplex model of emotion, where *Angry* is [Valence: negative, Arousal: positive], *Excited* is [Valence: positive, Arousal: positive], *Calm* is [Valence: positive, Arousal: negative], and *Bored* is [Valence: negative, Arousal: negative].

3.3 Participant recruitment

Participants were recruited via announcements on social media and various online fora devoted to music, dance, and linguistics. The experiment was carried out online in the PCIbex environment (Zehr & Schwarz 2018). Before the experiment, participants filled out a questionnaire on their demographic, linguistic, and musical background. Both native and non-native speakers of English participated in the study; only participants who reported being native speakers of English are analyzed here, since emotion words were provided in English, and cross-linguistic variation cannot be excluded. The results reported on below are for the 22 native English speakers who completed the experiment.[5]

3.4 Hypotheses

Our experiment tests several hypotheses related to (i-iii) in Section 2. In particular, we examine: (a.) whether participants draw consistent inferences about particular stimuli, i.e., if stimuli with high ratings for *Angry* received low ratings for *Calm*, and so forth; (b.) whether some of the information that auditory stimuli convey can be recovered from movement stimuli that were created as a response to those sounds. Positive answers to these questions would support the idea that music and music-accompanying movement encode descriptive information in comparable ways, i.e., that participants draw the same types of inferences about musical and movement stimuli.

[5]The data set is available at `https://doi.org/10.17605/osf.io/abgz4`

		Auditory	Visual
Angry	Mean	34	46
	SD	28	33
Bored	Mean	28	33
	SD	30	29
Calm	Mean	26	29
	SD	29	26
Excited	Mean	52	49
	SD	36	31

Table 1 Mean slider ratings and standard deviations for auditory and visual stimuli on the four descriptors in the study. Data pooled across all stimulus items and subjects.

4 Results

The first question we examined is whether listeners respond to auditory and visual stimuli in a broadly comparable way, bearing on (i) and (ii) above, repeated here in simplified form:

(i.) do abstract body movements give rise to semantic inferences?
(ii.) are there parallels between inferences from hearing music and inferences from seeing abstract body movement?

Table 1 shows the mean responses and standard deviations to each of the four emotion descriptors for auditory and visual stimuli. Overall rating levels are similar for the two modalities, as are the relative patterns amongst descriptors. Participants exhibit a tendency to assign higher scores for high-arousal descriptors (*Angry, Excited*) than low-arousal ones (*Bored, Calm*); this may be an artifact of the stimuli selection, since stimuli were not counterbalanced with regards to emotion association (see section 3). There are no gross differences between the two modalities here, suggesting that participants are as likely to infer emotional content from movement as they are from music.

Next, we ask if individual auditory and visual stimuli are subject to consistent inferences from participants. Figure 3 shows two attempts to validate the response space.

The left plot in Figure 3 tests a form of split-half reliability, where what

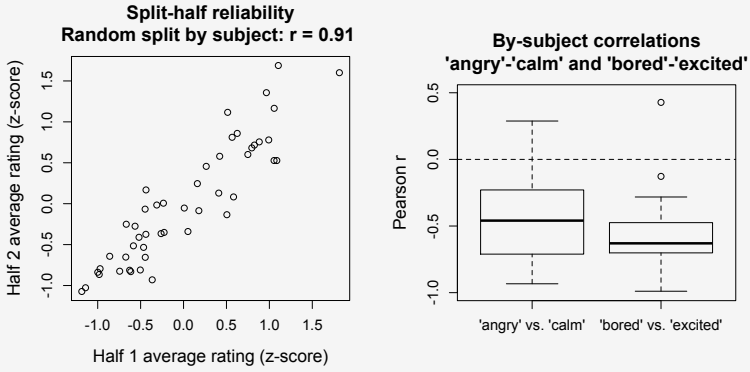

Figure 3 Left: correlation between two randomly-selected halves of the participant pool on the scores assigned to each combination of stimulus and descriptor. Right: Correlations between 'opposite' descriptors, computed across all stimuli within each participant.

is being split into random halves is the participant pool. The question is whether, for each stimulus, when a randomly-selected half of participants infer high levels of some descriptor from that stimulus, do the other half do the same? The answer is an emphatic yes (r = 0.91), showing that participants broadly agree on how much the terms *Angry*, *Calm*, *Excited*, and *Bored* are associated with particular stimuli. The right plot in Figure 3 summarizes within-subject correlations between 'opposite' descriptors. Our assumed theory (Russell 1980) situates the four descriptors in terms of a two-dimensional space of valence and arousal. If this is valid, we expect strong negative correlations between descriptors differing in both valence and arousal. As shown in the right plot, correlations are almost uniformly negative, some of them quite strongly so. This indicates, e.g., that when a participant infers high *Angry* content from a particular stimulus, they are likely to infer low *Calm* content from that stimulus.

Finally, we ask whether the inferences participants draw from visual stimuli tend to resemble the inferences they draw from the auditory stimuli that inspired the motion in the video animation, corresponding to (iii) in section 2. That is, do participants implicitly recover information from motion about the sound that the motion was intended to accompany? Figure 4 shows audiovisual correlations, treating each combination of stimulus, de-

scriptor, and participant as a separate observation. The observed correlation suggests that motion can be used to encode and decode some information from an auditory stimulus. That said, the effect here is not particularly large, suggesting that about 5% of the variance associated with visual scores can be accounted for by taking into account the audio scores of the corresponding items.

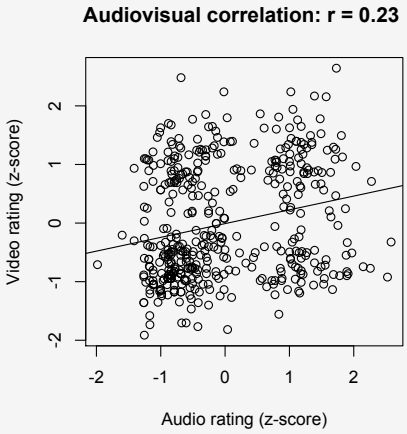

Figure 4 Correlation between slider scores for auditory stimuli and the visual stimuli that were created in response to them. Line shows general linear model.

While the scatterplot and simple correlation analysis above give us a useful summary of the data, they also ignore several aspects of the design that must be taken into account in a statistical analysis. In particular, our study has several *crossed random variables*, variables that are sampled from some larger population of interest. Here, we have asked a number of participants to rate a number of specific stimuli with regard to four particular linguistic terms used to describe emotions. Participant, stimulus item, and linguistic descriptor are all random variables, and fully analyzing the results requires a model that takes into account differences between participants, items, and descriptors while attempting to generalize across these variables.

To examine whether audiovisual correlations are robust across stimuli, descriptors, and subjects, we fit a linear mixed-effects regression model using the lme4 package in R (Bates et al. 2015). This type of model (also re-

ferred to as a 'multi-level' or 'hierarchical' model) is specifically designed for analyzing studies with multiple random variables in a crossed or nested design. It allows us to examine the *fixed effects* of interest, those that are systematically manipulated in the design of the experiment, while explicitly modeling variation associated with random variables. The dependent variable in our model was the slider-score for visual stimuli, using the slider score for the corresponding auditory stimulus as a predictor. The model also included fixed effects of the order in which the two tasks were performed (auditory first vs. visual first), as well as its interaction with auditory scores. The model included random intercepts for item, subject, and descriptor. We tested random effects for model improvement using the likelihood-ratio test; the by-item random slope of auditory score significantly improved fit and was retained. All slider scores were centred around the midpoint of the scale, to aid interpretation of fixed effects. The significance of fixed effects was gauged by dropping parameters and using the likelihood-ratio test. A summary table of the final model is shown below.

Random effects	Var.	SD			
Sub: intercept	23.7	4.9			
Item: intercept	16.0	4.0			
Item: audio response	0.07	0.26			
Descriptor: intercept	99.2	27.1			
Fixed effects	β	SE	t	χ^2	p
Intercept	-0.12	5.96	-0.02		
Audio response	0.29	0.12	2.32	4.73	0.03
Order: video first	-13.68	3.63	-3.77	11.64	<0.001
AudResp x VidFirst	-0.19	0.08	-2.32	5.33	0.021

Table 2 Summary of linear mixed-effects regression model of responses to video stimuli.

In the auditory-first order, auditory score was a significant (positive) predictor of visual score, as shown by the second fixed effect. On average, for every slider-point higher that a subject rated a sound for a particular affect word, they rated the corresponding video file 0.29 slider-points higher.

Visual scores were about 14 points lower on a 100-point scale when the visual condition was completed first than when the auditory condition was, as shown by the third fixed effect. And the correlation between visual and auditory scores was substantially lower when the visual condition was completed first, as shown by the interaction between auditory rating and task order. It appears, then, that participants draw inferences from animations of movements that mirror inferences from the auditory stimuli that inspired those movements, but they do so much more reliably when the auditory stimuli are presented first than when the visual stimuli are.

The random effect of descriptor significantly improved model fit, with the low-arousal descriptors assigned negative intercepts and the high-arousal ones assigned positive intercepts. This matches the observation from 1 that the high-arousal words are assigned higher scores in general than the low-arousal ones. The by-item random slope of audio score also significantly improved model fit. This means that some particular stimuli generated tighter correlations across audio and visual modalities than other stimuli did.

While the model here finds a significant positive effect of audio score on visual score, showing that some information is carried over between the two modalities, the scatterplot and statistical model should make it clear that this is not the only, the biggest, or the most important factor affecting scoring. In particular, while the general effect is robust enough to be unlikely to arise by chance, we've also seen two factors here that significantly affect the size of the correlation: particular stimulus items and task order. As a followup, we examine each of these in turn.

To further examine differences by stimulus item, slider ratings for the 4 descriptors were forced onto a 2-dimensional Euclidean valence-arousal space. This was done by averaging the 'opposite' descriptors onto a linear scale, then rotating the resultant two coordinates 45 degrees so as to weight the high-arousal descriptors for one axis and the high-valence descriptors for the other. This is a fairly naïve procedure and we do not claim absolute validity for the results, but they do allow us to inspect separation between the stimuli and correspondence between auditory and visual stimuli. Results are shown separately for each auditory stimulus and the correspond-

ing visual stimulus in Figure 5.[6]

Comparing the plots vertically, we see that several of the stimuli occupy overlapping ranges in the affective space (e.g., auditory b, e, and f). Nonetheless, each stimulus is located almost entirely in 1 or 2 quadrants (possibly with the exception of video a), and different stimuli differ in which quadrants they mainly occupy. Turning our attention to the horizontal comparisons in Figure 5, we observe that some stimuli have closer audiovisual correspondences than others, as indicated by the random slopes in our statistical model. In particular, both the random slopes returned by the model and the visualizations above suggest that stimuli c and f are judged quite similarly across modalities, while b is quite different (the other three items are intermediate). So one reason why the main effect of audiovisual correlation in our study is not extremely large is that it is not fully robust across items: some items have high correlations, the model judges that correlations are generally positive, but some items have smaller correlations or none at all. There are no obvious or straightforward properties that separate the stimuli with high audiovisual correspondence from those with less correspondence. For instance, stimuli b and d, which both display clear audiovisual mismatches, are not situated in similar regions of the two-dimensional space, nor are they especially dissimilar in their positions from other stimuli that display tighter correspondence. So at this point, we cannot draw any clear conclusions about what makes particular stimuli transmit affective information more effectively than others.[7] One possibility worth following up on, however, is the duration of stimuli: items c and f, which have the highest audio and video correlations, are also the two longest stimuli. It is possible that as visual and auditory stimuli grow longer, subjects are more certain about their affective content and therefore able to 'decode' such content more reliably across modalities.

Figure 6 examines the effect of task order on audiovisual correlations. The statistical model showed that subjects who completed the auditory rat-

[6]See footnote 4 for link to the materials; for improved readability, Figure 5 uses the letters *a*, *b*, *c*, *d*, *e*, and *f*, instead of numerals; the letters map onto the numerical order of the original stimuli (*02*, *13*, *19*, *27*, *29*, and *30*).

[7]One open issue beyond the scope of this paper concerns a formal background theory of how movement systematically reacts to sound and music, and how musical properties are preserved in music-responsive movement.

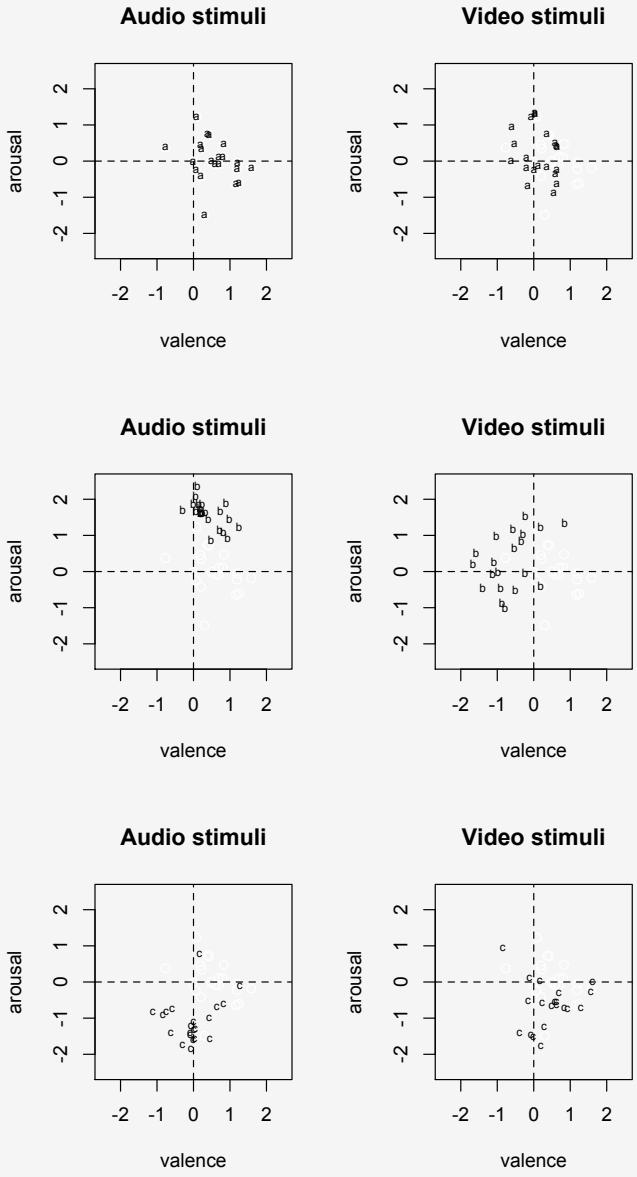

Figure 5 Ratings for each stimulus in a two-dimensional valence-arousal space. Each point represents one participant rating one stimulus on all 4 descriptors.

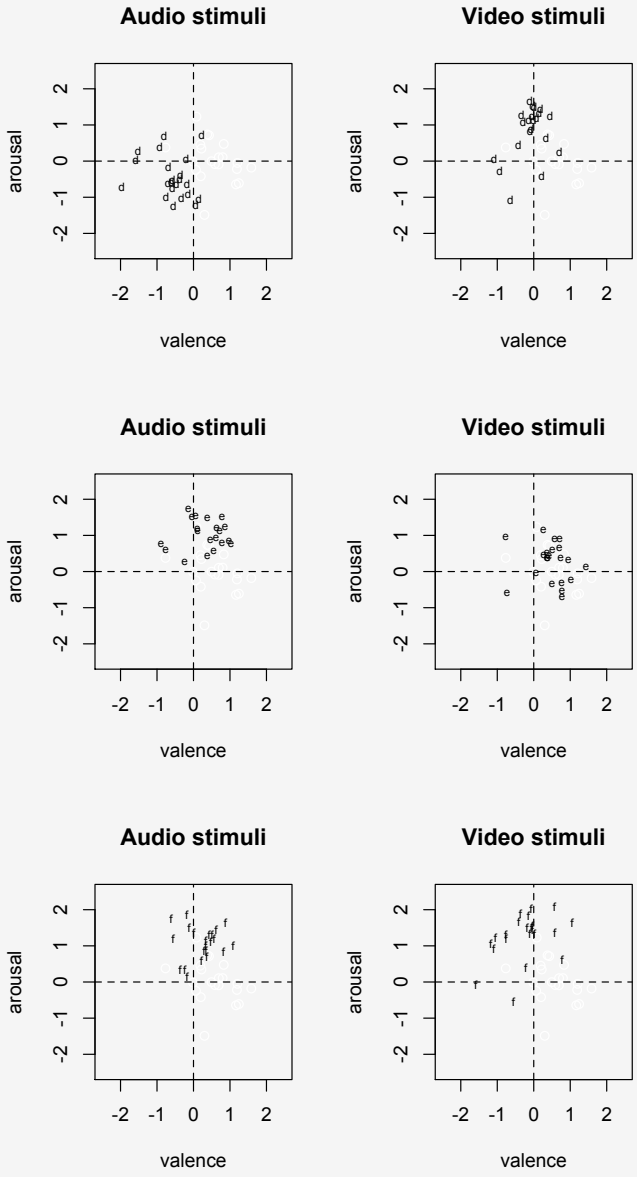

Figure 5 Ratings for each stimulus in a two-dimensional valence-arousal space. Each point represents one participant rating one stimulus on all 4 descriptors.

ing task first had higher correlations with their ratings on the visual task. Figure 6 clearly reflects this difference. It also shows that correlations between matched auditory and visual stimuli are somewhat more variable for those who rated the visual stimuli first.

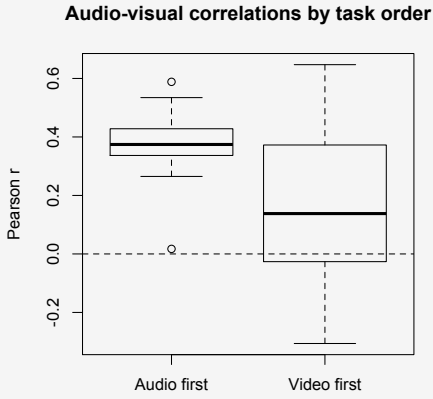

Figure 6 Distribution of within-subject correlations between matched auditory and visual stimuli, separated by task order.

5 Discussion

Our results suggest that inferences from music and inferences from body movement are coherent, consistent, and mutually informative. This is in line with a view where (i.) body movement gives rise to similar inferences to what we find in music, (ii.) there are parallels between the inferences from music and the inferences from body movement, and (iii.) listeners can recover information about inferences from music just from viewing body movement based on the music. While the effect of cross-domain inference was detected in the study and is statistically significant, it is not a particularly large effect, and does not generalize equally across all stimuli, nor across task orders.

The finding that correlations are more robust when the auditory condition occurs before the visual condition was not expected. We had anticipated that there might be some effect of order, but had no particular hypothesis about what that would be. A post-hoc hypothesis that might explain this finding involves the fact that, according to Schlenker's (2017;

2019a; 2021) theory, musical stimuli give rise to inferences on the physical movement of virtual sources / denoted objects (among other things). The inferred semantics of the auditory stimuli, when presented first, could thus activate various kinds of movement schemata; that would facilitate further processing of actual visual representations of movement. Because the motion-capture animations are straightforward representations of people moving, the effect of order could reflect such a facilitation in the auditory-first condition. Conversely, there is no reason to think that viewing movements activates auditory and/or musical schemata, so the auditory condition would not benefit from this facilitation after viewing movements. This fundamental asymmetry, if replicated in future work, could thus be seen as support for Schlenker's hypothesis that musical stimuli are interpreted in terms of physical, spatial movements.

The study also found significant variation across the six stimuli used here in how closely auditory and visual scores tracked one another. While there was no obvious generalization about the acoustic, visual, or perceived affective properties of stimuli that yielded closer multi-modal correspondence, this finding suggests that investigating such variation could be an interesting avenue for further research. Developing a detailed theory of how low-level auditory or visual cues affect valence and arousal may yield insights into cases where affective information is easier or harder to transmit across modalities.

Acknowledgments For extremely valuable written comments and in depth discussion, of the stimuli and experimental design in particular, we thank Nadine Bade, Chuck Bradley, Emmanuel Chemla, Guillaume Dezecache, Masha Esipova, Carlo Geraci, Kinjal Joshi, Nadya Modyanova, Leo Migotti, Timo Roettger, Philippe Schlenker, Henrik Torgersen and Sarah Zobel. For insightful discussion, we thank audiences at the Musical Meaning workshop at NYU, the 15th International Symposium of Cognition, Logic and Language (University of Latvia, Riga, November 26-27, 2021). We are enormously grateful to Naomi Francis, in particular for support with the copy-editing of this paper. Finally, we thank three CSSP 2021 reviewers for helpful and constructive feedback. This research was partially supported by funding from the Faculty of Humanities career development grant at the University of Oslo [PI: Patel-Grosz], and by the Research Council of Norway through its Centres of Excellence scheme (project number 262762) and by NordForsk's Nordic

Sound and Music Computing Network, NordicSMC (project number 86892).

References

Bates, Douglas & Maechler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Gadir, Tami. 2014. *Musical meaning and social significance: techno triggers for dancing*. University of Edinburgh dissertation.

Greenberg, Gabriel. 2021. The iconic-symbolic spectrum. Ms., UCLA, Feb. 2021.

Juslin, Patrik N. 2013. What does music express? basic emotions and beyond. *Frontiers in Psychology* 4. 596. https://doi.org/10.3389/fpsyg.2013.00596.

Kelkar, Tejaswinee. 2019. *Computational analysis of melodic contour and body movement*. University of Oslo dissertation.

Kelkar, Tejaswinee & Jensenius, Alexander R. 2018. Analyzing free-hand soundtracings of melodic phrases. *Applied Sciences* 8. 135. https://doi.org/10.3390/app8010135.

Mason, Paul H. 2012. Music, dance and the total art work: choreomusicology in theory and practice. *Research in Dance Education* 13. 5–24.

Meyer, Leonard B. 1956. *Emotion and meaning in music*. Chicago : The University of Chicago Press,

Patel, Aniruddh D. 2008. *Music, language, and the brain*. New York, NY : Oxford University Press,

Patel-Grosz, Pritty & Grosz, Patrick Georg & Kelkar, Tejaswinee & Jensenius, Alexander Refsum. 2018. Coreference and disjoint reference in the semantics of narrative dance. In *Proceedings of sinn und bedeutung 22*, 199–216.

Patel-Grosz, Pritty & Grosz, Patrick Georg & Kelkar, Tejaswinee & Jensenius, Alexander Refsum. to appear. Steps towards a semantics of dance. *Journal of Semantics*. eprint: https://ling.auf.net/lingbuzz/005634.

Russell, James A. 1980. A circumplex model of affect. *Journal of Personality and Social Psyschology* 39. 1161–1178. https://doi.org/10.1037/h0077714.

Schlenker, Philippe. 2017. Outline of music semantics. *Music Perception* 35. 3–37. https://doi.org/10.1525/mp.2017.35.1.3.

Schlenker, Philippe. 2019a. Prolegomena to music semantics. *Review of Philosophy & Psychology* 10. 35–111. https://doi.org/10.1007/s13164-018-0384-5.

Schlenker, Philippe. 2019b. What is super semantics? *Philosophical Perspectives* 32. 365–452. https://doi.org/10.1111/phpe.12122.

Schlenker, Philippe. 2021. Musical meaning within super semantics. *Linguistics & Philosophy*. https://doi.org/10.1007/s10988-021-09329-8.

Sievers, Beau & Polansky, Larry & Casey, Michael & Wheatley, Thalia. 2013. Music and movement share a dynamic structure that supports universal expressions of emotion. In *Proceedings of the national academy of sciences 110*, 70–75. https://doi.org/10.1073/pnas.1209023110.

Zehr, Jeremy & Schwarz, Florian. 2018. *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832.