**The Melodicon: Representing Melodic Relationships Using Network Science Methods**

**Abstract**

Semantic memory is represented as networks of words and their defining linguistic features. In semantic networks, related words are represented “closer” to each other than unrelated words, reflected in higher relatedness judgements and faster reaction times. While much research has investigated cognition underlying improvisation, the nature of improvisers’ knowledge structure has not been characterized. The current study examined whether network science can model the relationships between melodic sequences in improvised music. Using melodic sequences extracted from a large corpus of transcribed jazz improvisations, we sought evidence for melodic networks by asking participants to judge the relatedness of sequence pairs. We found that as distance increased, participants judged melodic sequences as less related. Moreover, the relationship between distance and reaction time was quadratic: participants slowed in RT up to distance four, then were quicker, a parallel finding to research in language. This study provides preliminary evidence for the existence of a melodic network, akin to semantic networks in language.

**The Melodicon: Representing Melodic Relationships Using Network Science Methods**

**Introduction**

**Method**

**Participants**

62 participants (30 men, 32 women; mean age 25.9 (*SD* = 8.38)) were recruited using the online study recruitment platform Prolific, filtering to individuals who lived in the US and Europe with high proficiency in English. 30 of these participants were specifically recruited as musicians, using the filtering criterion of having played an instrument for 5 years or more. 2 participants were excluded from analysis due to low-quality data; our final sample consisted of 27 musicians (defined as individuals who currently play an instrument) and 31 non-musicians (individuals who do not currently play an instrument). Participants received $9 USD/hour via Prolific’s payment system.

**Materials**

***Network calculation.*** Our network was constructed from the Weimar Jazz Database (WJD; Pfleiderer, 2017). The WJD corpus is a large collection of 456 annotated improvisation recordings and transcriptions by expert jazz musicians (e.g. Charlie Parker, John Coltrane, Miles Davis).

Constructing a network of musical sequences from this dataset involved several steps. First, we extracted 116,408 unique 5-note sequences from the set of 456 recordings. The frequency of sequences (i.e., how often each unique sequence occurred across the whole dataset) was highly skewed, with most (76%) occurring just once, but a handful (<0.2%) occurring over 30 times. Since more frequent sequences should be more familiar to participants, potentially affecting distance ratings, we removed extreme cases by excluding all sequences with a frequency below 1 or above 20 (20 was chosen as an upper bound since, following exclusion of sequences with a frequency of 1, 20 was the highest frequency within 3 standard deviations [5.38] of the mean [3.97]). This produced a reduced set of 27,380 unique sequences.

An initial network was formed by considering sequences as nodes, and the continuations between sequences as edges. We binarized this network (i.e., set all edge weights to 1), before using it to compute a distance matrix using the Brain Connectivity Toolbox in Matlab (Rubinov & Sporns, 2010). A distance matrix quantifies the distances between all nodes (i.e., sequences) in terms of the shortest number of steps between them in the network (Boccaletti et al., 2006). From a distance matrix it is then possible to select pairs of nodes separated by a given number of steps.

We selected sets of 40 pairs for each of the following 7 distances: 1, 2, 3, 4, 6, 10, and 20 steps. These distances were based on the distances tested in the original semantic distance paper by Kenett et al. (2017). The pairs in these 7 sets were selected pseudo-randomly, following two constraints: all nodes had to be unique (i.e., no sequence could occur twice across all sets of pairs), and no two sets could differ significantly in frequency (i.e., the 80 nodes in each set had to have equivalent distributions in terms of their frequency across the entire WJD). To ensure the sets did not differ in frequency (Fig. 1), we ran a one-way ANOVA in *R* v3.6.1. There was no statistically significant difference in frequency across the 7 distance conditions, *F*(6, 70) = 0.72, *p* = .634. This produced the 280 pairs of sequences used as stimuli in the present study.

Chart, box and whisker chart

Description automatically generated

Figure 1. Boxplot of node frequency across the 7 distance conditions. Node frequency did not significantly differ across distances.

***Stimuli.*** The 5-note sequence pairs calculated from the network were converted into MIDI format played with a piano sample via MATLAB, then written to WAV files using Winamp v5.8. Each pair was presented using one audio file at a tempo of 3 notes/second with an arrhythmic 2-second pause between the sequences (making each stimulus 5 seconds long). Stimuli were presented for distances 1, 2, 3, 4, 6, 10, and 20.

***Melodic relatedness task.*** We adapted the Semantic Distance Task (SDT; Kenett, Levi, Anaki, & Faust, 2017) for use with the melodic pairs. The task consisted of 7 conditions (1-, 2-, 3-, 4-, 6-, 10-, and 20-step distances), each containing 40 pairs. After listening to each pair, participants made a binary (yes/no) relatedness judgement as a response to the prompt “Are these sequences related?”. We considered the pairs for distances 1-4 as a-priori related, as these pairs contained overlapping note content (e.g. pairs with a distance of 1 only differed by one note with a 4-note overlap, pairs with a distance of 4 differed by four notes with a 1-note overlap).

***Survey.*** We created a survey to collect information about music listening habits and musical background, including items specifically related to jazz listening and experience improvising (Appendix 1). Several of these items, primarily related to musical training and education, were collected for exploratory analysis prior to conducting future studies that target professional musicians, but we report effects of music listening and musicianship here.

**Procedure**

Participants registered for the study using Prolific and were administered the tasks online following collection of informed consent via Qualtrics. Participants were instructed to complete the study in a quiet location, wearing headphones, and on a laptop or desktop computer that was connected to the Internet (the study would not run on a tablet or mobile device). The melodic relatedness task was administered using Pavlovia. Participants were given instructions, permitted to adjust their volume to a comfortable level, and provided 5 practice examples of the task. In the melodic relatedness task, each trial began with an 80 ms fixation cross appearing in the center of the screen. Next, the stimulus pair played for 5 s. Following the presentation of the stimulus pair, the participant decided whether the melodies were related to each other by pressing the “s” key to indicate “yes”, or the “k” key to indicate “no”. Once the participant pressed the key, the next trial was immediately initiated. Stimulus pairs were randomly presented in blocks of 50, with the opportunity to take a break between each block. This task took between 40 – 50 minutes to complete. After completion of the melodic relatedness task, participants were redirected to Qualtrics to complete the musical background survey. After completing the survey, participants were redirected back to Prolific to indicate study completion and eligibility for payment.

**Results**

***Melodic similarity analysis.*** Visual inspection of the mean responses (Fig. 2) revealed that, contrary to our initial predictions, on average participants judged the distance 20 pairs as related in 64.4% of trials, with all other distances being judged as expected. We believe that this finding warrants further investigation, i.e. a follow-up study that presents stimuli between 10- and 20- steps apart. We excluded trials with distance-20 stimuli from our main response and reaction time analyses and conducted a brief melodic similarity analysis to determine whether distance-20 stimulus pairs shared melodic content with lower distances.

Chart

Description automatically generated

***Figure 2.*** Plot of mean responses (y-axis) against distance (x-axis) for the melodic relatedness task, including distance 20 trials.

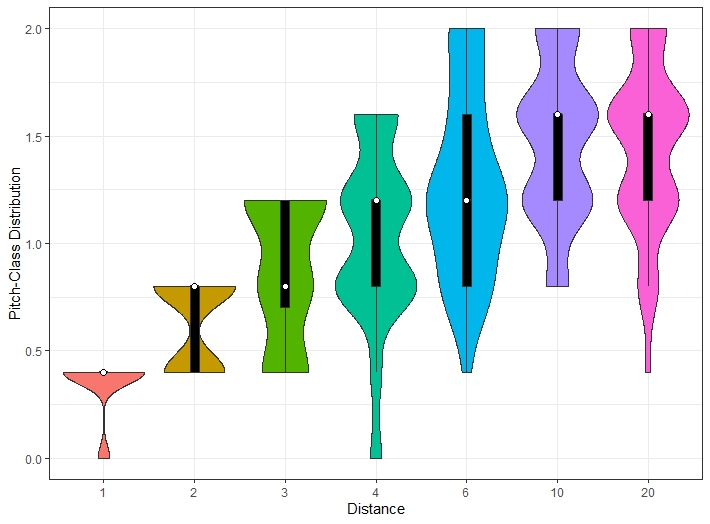
Pitch- and interval-class distributions, as well as melodic contour vectors, were calculated for each stimulus pair using the MATLAB MIDI Toolbox v 1.1 (Toiviainen & Eerola, 2016). The taxicab norm was used to compare these distributions between items in the pair. We then conducted a one-way ANOVA in *R* v3.5.1 to determine whether there were differences in pitch-class distribution across the 7 distance conditions. As expected, pitch-class distribution was significantly different across the 7 distances, *F*(6, 273) = 60.94, *p* < .001 (see Fig. 3a). To determine whether the distance-20 stimuli shared similar pitch-class distribution with lower distances, we conducted Tukey’s post-hoc tests, specifically looking for distances that were *not* significantly different in pitch-class distribution. All combinations of distances were significantly different in pitch-class distribution except distances 3 and 4 (*p* = .44), distances 6 and 10 (*p* = .17), distances 6 and 20 (*p* = .36), and distances 10 and 20 (*p* = .99). As it does not seem that distance-20 stimuli shared similar pitch-class distributions with distances 1-4, this does not seem like a likely explanation for relatedness judgements being higher.

A one-way ANOVA was then conducted to determine whether interval-class distributions significantly differed across the 7 distances. As expected, interval-class distribution was significantly different across the 7 distances, *F*(6, 273) = 30.45, *p* < .001 (see Fig. 3b). To determine whether the distance-20 stimuli shared similar interval-class distribution with lower distances, we conducted Tukey’s post-hoc tests, specifically looking for distances that were *not* significantly different in interval-class distribution. As a good deal of interval distributions were not significantly different across different distance combinations, we present a table of these here:

|  |  |  |  |
| --- | --- | --- | --- |
| *Distance 1* | *Distance 2* | | *p* |
| 2 | 3 | .054 | |
| 3 | 4 | .211 | |
| 3 | 6 | .211 | |
| 4 | 6 | 1.00 | |
| 4 | 10 | .980 | |
| 4 | 20 | .921 | |
| 6 | 10 | .980 | |
| 6 | 20 | .921 | |
| 10 | 20 | .999 | |

While it is clear that the distance-20 stimuli share an interval distribution with several other distances, it is a) not with distances 1-3, and b) not exclusive to combinations involving the distance-20 stimuli. Therefore, this also does not seem to be a likely candidate to explain the high relatedness judgements at distance 20.

As a final avenue for exploring how the distance-20 pairs might be similar to lower-distance pairs, we examined how melodic contour varied across distance. A one-way ANOVA was conducted to determine whether melodic contour significantly differed across the 7 distances. Contour varied significantly by distance, *F*(6, 273) = 2.52, *p* = .02 (see fig. 3c). However, the only significant difference in contour was between distances 1 and 4, *p* = .004. As contour similarity was not exclusive to distance pairings including 20, this also does not likely provide an explanation as to why the distance 20 stimuli were rated as similar. We report other possible explanations as well as future directions to discern why this is the case in the Discussion; the distance-20 ratings are excluded for the remainder of the Results section.

Diagram

Description automatically generatedChart

Description automatically generated

*Fig 3a-c. Violin plots of pitch-class distribution, interval-class distribution, and melodic contour of the stimuli used in this experiment. This melodic similarity analysis did not reveal similarities between the distance-20 stimuli and stimuli from lower distances.*

***Responses.*** Responses to the melodic relatedness task, which were binary, were analyzed using logistic regression via the glm() function in R *v.3.5.1.* To determine the most predictive model, musician status and music listening parameters taken from the survey were serially added to an intercept-only model in the order shown using ANOVA model comparisons. Many survey questions about musical training and experience were only presented to participants who were musicians, so these variables were examined in a separate exploratory analysis that only included these participants.

A logistic regression analysis with five predictors (*distance, musician status, distance\*musician status, hours a week spent listening to music, hours a week spent listening to jazz*) tested whether participants judged pairs of melodic sequences as related at distances 1, 2, 3, 4, 6, and 10 (removing trials where they judged the distance 20 stimuli). Overall, this model provided a significantly better fit than an intercept-only model, χ2 (5, *N* = 15248) = 691.17, *p* < .001. The model correctly classified approximately 57.6% of trials.

Controlling for other variables in the model, a 1-unit increase in distance decreased the odds a participant would judge a pair as related by a factor of .86, *z* = -18.71, *p* < .001, 95% CI [.85, .88]. Holding other variables constant, each additional self-reported hour spent listening to music increased the odds a participant would judge a pair as related by a factor of 1.01, *z* = 8.32, *p* < .001, 95% CI [1.005, 1.008]. Holding other variables constant, each additional self-reported hour spent listening specifically to jazz decreased the odds a participant would judge a pair as related by a factor of .97, *z* = -4.89, *p* < .001, 95% CI [.96, .98]. Controlling for other variables in the model, whether or not the participant was a musician was not reliably associated with the relatedness judgement, OR = 1.07, *z* = 1.05, *p* = .29, 95% CI [.96, .98]. The interaction between musicianship and distance was not significant, *z* = 1.91, *p* = .056. Taken together, these results indicate that for distances prior to 10, relatedness judgements decrease with increases in distance. Additionally, music listening habits are a more important factor in these judgements than musicianship at lower distances.

***Musician responses.*** For participants who were musicians, information on their musical background was collected. As above, we analyzed these traits (*primary instrument proficiency, proficiency at improvising, hours currently spent playing music per week, hours spent playing jazz per week, percentage of playing time spent improvising*) for distances below 10. For distances 1-10 in the musician group, a logistic regression with these predictors, controlling for distance, tested how the musician group made relatedness judgements. This model provided a significantly better fit than one that just included distance, χ2 (5, *N* = 7109) = 128.89, *p* < .001. Holding other variables constant, each one-unit increase in self-rated proficiency on their primary instrument increased the odds a participant would judge a pair as related by a factor of 1.10, *z* = 5.31, *p* < .001, 95% CI [1.06, 1.14]. Holding other variables constant, each one-unit increase in self-rated proficiency in improvisation decreased the odds a participant would judge a pair as related by a factor of .93, *z* = -3.43, *p* < .001, 95% CI [.89, .97]. Holding other variables constant, each one-hour increase in hours spent playing music per week decreased the odds a participant would judge a pair as related by a factor of .95, *z* = -5.82, *p* < .001, 95% CI [.93, .97]. Holding other variables constant, each one-hour increase in hours spent improvising per week increased the odds a participant would judge a pair as related by a factor of 1.17, *z* = 8.65, *p* < .001, 95% CI [1.13, 1.21]. Percentage of total playing time dedicated to improvising was not significantly related to the relatedness judgement, OR = .99, *z* = -.69, *p* = .49.

***Reaction time.*** As in the language studies conducted by Kenett and colleagues (2017), prior to examining the reaction time data, trials were excluded from analysis if they were “incorrect” and did not align with whether the stimulus pair actually shared notes. For distances 1-4, trials with a response of “no” were excluded from the reaction time analysis, while for distances 6 and 10, “yes” trials were excluded.

Inspection of the reaction time data revealed a likely quadratic relationship between distance and reaction time (Figure 3). A quadratic regression was performed to quantify the relationship between distance and reaction time with the added predictors musician status, hours per week spent listening to music, and hours per week spent listening to jazz. A quadratic model including these predictors fit the data significantly better than a linear model, *F*(5, 8905) = 33.1, *p* < .001, PRE = .018. There was a significant effect of musicianship such that musicians took .06 seconds longer to make relatedness judgements over all distances, *t*(8910) = 4.31, *p* < .001, 95% CI [.034, .092]. There was a significant effect of music listening such that each additional reported hour of music listening increased overall reaction time by .001s, *t*(8910) = 4.321, *p* < .001, 95% CI [.001, .002]. In particular, there was a significant effect of jazz listening, such that each additional self-reported hour of listening to jazz increased overall reaction time by .02s, *t*(8910) = 7.364, *p* < .001, 95% CI [.015, .026].

***Chart

Description automatically generated***

***Figure 4.*** Plot of quadratic fit of reaction time (y-axis) against distance (x-axis) for the melodic relatedness task, grouped by musician status, excluding distance-20 trials.

**Discussion**

Appendix Items

\*\*\*Survey

\*\*\*Task repo

\*\*\*Link to github repo? Make an open science framework repo? Both?

**References**

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. In *Physics Reports* (Vol. 424, Issues 4–5, pp. 175–308). North-Holland. https://doi.org/10.1016/j.physrep.2005.10.009

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. <https://doi.org/10.1016/j.neuroimage.2009.10.003>

Toiviainen, P. & Eerola, T. (2016). *MIDI Toolbox 1.1*. Github, URL: https://github.com/miditoolbox/1.1