The famous melodies stimulus set

Amy M. Belfi[1] · Kaelyn Kacirek[2]

Abstract

Famous musical melodies, such as "Row, Row, Row Your Boat" and "Hot Cross Buns," are frequently used in psychological research. Such melodies have been used to assess the degree of cognitive impairments in various neurological disorders, and to investigate differences between "naming" vs. "knowing." Despite their utility as an experimental stimulus, there is currently no standardized, openly available set of famous musical melodies based on a United States population, as prior work on the topic has primarily relied on creating stimuli in an ad hoc manner. Therefore, the goal of the present work was to create a set of famous musical melodies. Here, we describe the development of the Famous Melodies Stimulus Set, a set of 107 melodies. We provide normative data for the melodies on five dimensions: familiarity, age of acquisition, emotional valence, emotional arousal, and naming ability. Participants ($N = 397$) rated the melodies on these five variables, validating that most melodies were highly familiar and reliably named. While familiarity ratings were skewed, all other rating scales covered a relatively broad range, allowing for researchers to select melodies for future work based on particular attributes.

Keywords Music · Naming · Familiarity · Valence · Arousal

Introduction

Famous musical melodies, such as "Row, Row, Row Your Boat" and "Rudolph the Red-Nosed Reindeer," are frequently used in psychological research. For example, famous melodies have been used to assess the degree of cognitive impairments in various neurological disorders, such as Alzheimer's disease. A sizeable body of work indicates that individuals with Alzheimer's disease have preserved ability to identify melodies, despite memory deficits for other categories of stimuli (Cuddy & Duffin, 2005; Cuddy et al., 2012; Cuddy, Sikka, & Vanstone, 2015; Hsieh, Hornberger, Piguet, & Hodges, 2011; Vanstone et al., 2012). In contrast to Alzheimer's disease, other work indicates that individuals with semantic dementia show a deficit in naming famous melodies (Hsieh et al., 2011; Johnson et al., 2011), although evidence of impaired melody recognition in semantic dementia has been mixed (Hailstone, Omar, & Warren, 2009; Omar,

Hailstone, Warren, Crutch, & Warren, 2010). Overall, famous musical melodies have proved to be a useful stimulus for identifying cognitive functions that may be selectively preserved (or impaired) in various dementias (for review see Omar, Hailstone, & Warren, 2012).

Other neuropsychological studies have used famous melodies to investigate lexical and conceptual retrieval in patients with focal brain damage. For example, prior work indicates that patients with left temporal polar damage are impaired at naming famous persons, famous landmarks, and famous melodies, suggesting that the left temporal pole is a "heteromodal" convergence region for semantically unique items (Belfi, Kasdan, & Tranel, 2019; Belfi & Tranel, 2014; Schneider, Heskje, Bruss, Tranel, & Belfi, 2018). Famous melodies have also been used to identify apperceptive agnosia for music in patients with temporal lobe lesions (Ayotte, Peretz, Rousseau, Bard, & Bojanowski, 2000; Baird, Walker, Biggs, & Robinson, 2014). Such melodies have been used to investigate memory and language deficits (or a lack thereof) in other populations, including individuals with primary progressive aphasia (Macoir et al., 2016), amnesia due to herpes simplex encephalitis (Finke, Esfahani, & Ploner, 2012), and cochlear implant users (Olszewski, Gfeller, Froman, Stordahl, & Tomblin,

✉ Amy M. Belfi
   amybelfi@mst.edu

1   Department of Psychological Science, Missouri University of
    Science and Technology, 135 H-SS, 500 W. 14th St.,
    Rolla, MO 65409, USA

2   Department of Engineering Management and Systems Engineering,
    Missouri University of Science and Technology, Rolla, MO, USA

2005; Stordahl, 2002; Volkova, Trehub, Schellenberg, Papsin, & Gordon, 2014).

While famous melodies have clear utility for neuropsychological research, they are also used more generally to investigate perception and cognition in healthy populations. Famous melodies have been used to investigate differences between the perception and imagination of music (Herholz, Halpern, & Zatorre, 2012), and the distinction between recognition and naming, or 'tip-of-the-tongue' phenomena (Kostic & Cleary, 2009). Famous melodies are also a useful stimulus for investigating the timing of melody recognition (Bailes, 2010). For example, prior work indicates that individuals can recognize musical melodies within 500 ms or a few notes (Büdenbender & Kreutz, 2016; Dalla Bella, Peretz, & Aronoff, 2003; Filipic, Tillmann, & Bigand, 2010; Huijgen et al., 2015; Tillmann, Albouy, Caclin, & Bigand, 2014). Familiar melodies are also a frequently used stimulus for evoking autobiographical memories (Belfi, Karlan, & Tranel, 2016; Ford, Addis, & Giovanello, 2011; Janata, Tomic, & Rakowski, 2007).

While familiar melodies are used in a wide variety of psychological studies, there is currently no set of famous melodies that have been validated in a United States sample of participants. Various "famous" or "familiar" melody tests have been previously reported in the literature, but these stimulus sets were either created idiosyncratically to match the knowledge and musical preferences of individual case studies (Steinke, Cuddy, & Jakobson, 2001), contain only a small number of melodies (Liégeois-Chauvel, Peretz, Babaï, Laguitton, & Chauvel, 1998), and/or do not have associated normative data (Hsieh et al., 2011; Kostic & Cleary, 2009). Furthermore, several previously used sets of famous melodies were developed outside the United States. Although there are likely some melodies familiar to all Western musical listeners, such stimulus sets may not be entirely appropriate for a US-based participant group (for example, a melody set published in French; Peretz, Babai, Lussier, Hebert, & Gagnon, 1995).

While there are no standardized sets of famous musical melodies, there are similar stimulus sets focusing on other features of music besides familiarity. For example, there are multiple sets of melodies chosen or composed to represent different emotions (e.g., a "happy" or a "sad" melody, or melodies with a positive or negative valence). There are numerous sets of such emotional melodies, consisting of either sung or instrumental music (Eschrich, Münte, & Altenmüller, 2008; Koelsch et al., 2013; Lepping, Atchley, & Savage, 2016; Livingstone & Russo, 2018; Rainsford, Palmer, & Paine, 2018; Vieillard et al., 2008). There are also sets of musical stimuli used to study musical expectancy, which often consist of series of chord progressions (Koelsch, Gunter, Wittfoth, & Sammler, 2005). Outside

of the realm of music, normed stimulus sets are quite common – for example, there are stimulus sets of visual stimuli representing different emotions (Kurdi, Loano, & Banaji, 2017; Lang, Bradley, & Cuthbert, 2008), everyday objects (Snodgrass & Vanderwart, 1980; Tranel, Logan, Frank, & Damasio, 1997) or unique objects (Tranel, Enekwechi, & Manzel, 2005); sets of words in English (Deyne, Navarro, Perfors, Brysbaert, & Storms, 2018; Scott, Keitel, Becirspahic, Yao, & Sereno, 2018) and other languages (Chedid et al., 2018); and sets of voices (Darcy & Fontaine, 2019; Zäske, Skuk, Golle, & Schweinberger, 2019).

In sum, while prior research has often relied on famous musical melodies to investigate various cognitive and perceptual processes, and while normed stimulus sets are frequently used in other sensory domains, a set of standardized famous melodies has not yet been developed. The goal of the present work was to design such a set of famous musical melodies. Some prior work using famous melodies has taken the approach of using 'naturalistic' excerpts – for example, work investigating music as an autobiographical memory cue has used excerpts of popular music from the Billboard charts (e.g., Belfi et al., 2016; Janata et al., 2007), and work investigating memory for emotional music has used film scores (Eschrich et al., 2008). In contrast to this approach, here we sought to create stimuli that isolated the melodies themselves. Therefore, melodies in the present stimulus set consist of a single line melody with no harmonic accompaniment and no lyrical content. In contrast to other stimulus sets of unfamiliar (or merely recognizable) melodies, the present set was designed to contain musical melodies that were highly likely to be both recognized and named. As the primary goal was to create a set of highly familiar melodies, participants rated melodies on their familiarity and provided a written response with the melody name. Ratings of age of acquisition, emotional valence, and emotional arousal were also collected. Below we first describe the development of the stimulus set and selection of melodies, followed by the collection of normative data and characterization of the melodies.

Stimulus development

First, we sought to obtain a broad array of famous melodies. Our initial stimulus set consisted of 52 melodies used in our previous work (Belfi & Tranel, 2014). These stimuli consist of familiar melodies such as "Happy Birthday" and "Row, Row, Row Your Boat" (see Belfi, Kasdan, & Tranel (2019) for the full list of the original 52 melodies). While this initial stimulus set

contained a large number of famous melodies, we sought to exhaust any additional possible melodies that may have been missed in our prior study. To this end, we first conducted an initial survey to identify any additional melodies that could be added to the stimulus set.

Methods

Participants

This initial survey was conducted using Amazon's Mechanical Turk (AMT). We restricted participation to workers in the United States who had completed at least 1000 previous AMT tasks and obtained approval ratings of at least 95%. Other inclusion criteria were that participants must be native English speakers and older than 18 years old. A total of 108 participants completed the task. Data from eight participants were excluded because they failed to provide appropriate answers to the questionnaire (for example, writing nonsense words), leaving a total of 100 participants. Of these 100 participants, 61 were men and 38 were women (one participant preferred not to answer). Participants ranged in age from 20 to 68 years old (M = 35.56, SD = 11.48). The task took approximately 5 min and participants were compensated at a rate of $0.75 for completing the survey.

Procedure

Participants were given the previously published list of 52 melodies (from Belfi & Tranel, 2014) and were asked to suggest five melodies that were not present on the list but that would fit the criteria of being famous melodies that are highly recognizable to a US audience. Participants saw five empty text boxes where they were to type the names of the five additional melodies.

Selection of melodies for final stimulus set

Our method of soliciting additions to the melody list could have potentially resulted in a total of 500 possible unique responses (100 participants each listing five melodies). We first removed responses that were clearly not melodies (e.g., blanks, responses such as "nice," nonwords, punctuation marks, etc.). Next, we identified melodies that were already included on the stimulus set (e.g., several participants wrote "Somewhere Over the Rainbow" when "Over the Rainbow" was already included in the set). Other melodies listed were unique names, but nonunique melodies (e.g., the "ABC Song" is the same melody as "Twinkle, Twinkle Little Star"). Once these redundant melodies were eliminated, we

next categorized the remaining melodies into eight categories or "genres" of melodies: children's, patriotic, movie/TV, Christmas, religious, classical, pop, or "other" for melodies that did not fall into one of the above categories. Our goal was to create a final stimulus set with a roughly distributed set of melodies across these categories. We also wanted to avoid choosing too many melodies from the same musical artist (for example, "Beat It," "Thriller" and "Billie Jean" were all commonly named melodies by Michael Jackson, as were "I Want to Hold Your Hand," "Yellow Submarine" and other songs by The Beatles), so we limited the number of melodies to three per artist. Melodies that were named by more than one participant were more likely to be selected than melodies only named once; however, several melodies were selected that were only named once. Our goal was to select melodies that were roughly equally distributed among categories (for example, we did not want to select an overabundance of children's melodies).

This selection process resulted in a total of 107 melodies comprising the final stimulus set (see the Appendix for the final list of melodies and their categorization). The 107 melodies in the final stimulus set contained the following number of melodies in each category: children's ($n = 20$), patriotic ($n = 11$), movie/TV ($n = 16$), Christmas ($n = 17$), religious ($n = 6$), classical ($n = 8$), pop ($n = 18$), and other ($n = 11$).

Melody construction

Once the final list of melodies was determined, the next step was to create the audio files to use in the normative data collection phase. Melodies were constructed following our previously published procedures (Belfi & Tranel, 2014). Briefly, the software MuseScore was used (musescore.org) to create each melody, which were single-line melodies with no harmonic accompaniment in a MIDI piano timbre. Each melody consisted of roughly one to two musical phrases and were an average of 13.31 s long (5.35 SD, range, 6–37 s). This relatively large range in stimulus length is common in work on musical melodies (Larrouy-Maestri, Harrison, & Müllensiefen, 2019), as our goal was to maintain a similar amount of musical information per melody, as opposed to absolute time.

Normative data collection

We next sought to collect normative data on variables typically used for other stimulus sets, of melodies, images, and other categories of stimuli. To this end, we collected ratings on two emotional categories that have

been frequently used in prior research: valence and arousal. As it has been hypothesized that melodies have a later age of acquisition than other categories of stimuli (Belfi et al., 2019), we also collected ratings on age of acquisition. Given that the focus of this stimulus set is highly familiar melodies, we collected ratings on familiarity. Finally, we sought to identify the "nameability" of each melody; therefore, we also asked participants to name each melody. To summarize, we collected normative ratings on the following variables: valence, arousal, familiarity, age of acquisition, and naming.

Methods

Participants

Our goal was to create a normed stimulus set of famous melodies, similar to other normed stimulus sets of emotional music (Livingstone & Russo, 2018; Vieillard et al., 2008). We sought to collect normative data from a wide range of participants, so that the stimuli would be suitable for use in a variety of research contexts. We collected data from two separate populations: undergraduate students at Missouri S&T who completed the study for research credits, and participants from Amazon's Mechanical Turk (AMT) who completed the study for monetary compensation. A secondary goal of the present work is to compare the results obtained from the undergraduate vs. the AMT groups, to identify possible differences between these two populations.

In determining our target number of participants, we first looked at prior work developing musical stimulus sets. For example, the Ryerson Audio-Visual Database of Emotional Speech and Song (RADVESS; Livingstone & Russo, 2018) is a stimulus set of voice recordings expressing various emotions in speech and song. For the RADVESS database, each stimulus was rated ten times (i.e., by ten individual participants). Another frequently used set of emotional musical stimuli obtained ratings from 20 participants for each stimulus (Vieillard et al., 2008). Given our prior work indicating fairly large individual differences in ratings of musical stimuli (Belfi, 2019), we sought to be more exhaustive in the number of participants for the present stimulus set. Therefore, to be as thorough as possible, our target number was to have each stimulus in the Famous Melodies Stimulus Set rated by at least 20 to 30 participants from each population (AMT and undergraduate), for a total of at least 50 individual ratings per stimulus.

Undergraduate participants

A total of 206 undergraduate participants completed the task. The task was administered online to participants in the Psychology Department Subject Pool at Missouri S&T, who completed the task for course credit. Participants were excluded if they failed to pass a "foil" question, which asked them to provide a specific answer (e.g., participants heard a voice saying "Please type the word 'banana' on the following screen"). A total of 22 participants failed to pass this foil and were therefore excluded, resulting in a total $N = 184$ participants. This rate of exclusion (10.6%) is similar to that found in prior research using online participants (Meade & Craig, 2012). Our final set of undergraduate participants consisted of 132 men, 50 women, and 2 nonbinary individuals. The average age was 19.74 (SD = 1.61) and participants had, on average, 3.17 years of formal musical training (SD = 3.55). To alleviate participant fatigue and to prevent participants from making similar ratings across scales, each participant rated all melodies on a single variable (participants were randomly assigned to variables). This resulted in the following numbers of participants for each variable: valence ($n = 34$), arousal ($n = 38$), familiarity ($n = 43$), age of acquisition ($n = 34$), and naming ($n = 35$).

Amazon Mechanical Turk participants

A total of 191 AMT participants completed the task. We restricted participation to workers in the United States who had completed at least 1000 previous AMT tasks and obtained approval ratings of at least 95%. The task took approximately 30 min to complete and participants were compensated $3 for their time. A total of 37 participants failed the foil question, leaving a total of $N = 154$ participants. This rate of exclusion (19%) is slightly higher than that from the undergraduate sample, but not entirely unusual for online studies using AMT workers. The final set of AMT participants consisted of 89 men, 64 women, and one genderqueer individual. The average age was 38.48 (SD = 11.90) and participants had, on average, 2.33 years of formal musical training (SD = 3.62). As with the undergraduate group, to alleviate participant fatigue and to prevent participants from making similar ratings across scales, each participant rated all melodies on a single variable (participants were randomly assigned to variables). This resulted in the following numbers of participants for each variable: valence ($n = 24$), arousal ($n = 31$), familiarity ($n = 44$), age of acquisition ($n = 28$), and naming ($n = 27$).

The two groups (AMT and undergraduate) significantly differed in age and years of musical training. The AMT sample was significantly older than the undergraduate sample, $t(335) = 21.07$, $p < 0.001$, 95% CI:[16.98, 20.48], and the undergraduate sample had significantly more years of musical

training, t(336) = – 2.03, p = 0.04, 95% CI:[– 1.65, – 0.02]. The proportions of men and women in the two samples were also significantly different, X^2 = 7.63, p = 0.005. There was a greater proportion of women in the AMT sample (41.8%) than the undergraduate sample (27.6%).

Procedure

The experimental task was created using JSPsych (de Leeuw, 2015) and implemented using PsiTurk (Gureckis et al., 2016). Upon providing informed consent, participants were randomly assigned to one of five ratings: valence, arousal, familiarity, age of acquisition, or naming. Participants completed only one of the five ratings across all stimuli, to minimize participant fatigue and allow for independence between ratings. The specific rating scales are as follows. Valence, arousal, familiarity, and age of acquisition were rated on Likert scales. For *valence*, participants were asked "How negative or positive is this melody?", and responded on the following scale: "very negative, "somewhat negative," "neither negative nor positive," "somewhat positive," "very positive." For *arousal,* participants were asked "How relaxing or stimulating is this melody?" and responded on the following scale: "very relaxing," "somewhat relaxing," "neither relaxing nor stimulating," "somewhat stimulating," "very stimulating." For *familiarity,* participants were asked "How familiar is this melody?" and responded on the following scale: "Not at all familiar," "slightly familiar," "somewhat familiar," "moderately familiar," "very familiar." For *age of acquisition,* participants were asked "Estimate the age at which you first learned this melody" and responded on the following scale: "Never," "age 0–2," "age 3–4," "age 5–6," "age 7–8," "age 9–10," "age 11–12," "age 13+". For *naming*, participants were asked "What is the name of this melody?" and responded by typing the name into a blank text box. After rating all 107 melodies, participants completed a brief demographics questionnaire, which included age, gender, and years of musical training.

Data quantification

All rating scales were converted to numerical values for analysis, starting with a 0 for the item at the leftmost end of the scale. For age of acquisition, the response "never" was reverse-coded as an 8 (to denote a "later" age of acquisition). We conducted all analyses a second time, removing all "never" trials for age of acquisition, and this did not substantially change our results (the analyses described below include the "never" trials). For *naming* trials, the experimenters read through all typed responses and manually scored them as correct or incorrect, as in prior research (Belfi & Tranel, 2014). Briefly, if the participant's response matched the correct name of the melody, it was scored a 1 for correct. If the response did not match the name of the melody, it was scored a 0 for

incorrect. If a participant provided an alternate but correct name, these instances were also scored as correct (for example, "What Child is This" for "Greensleeves").

Results

Normative ratings

The primary goal of this project was to create a normed set of famous melody stimuli. To this end, we calculated average ratings on each of the five variables: valence, arousal, familiarity, age of acquisition, and naming. Collapsed across both groups, the melody with the lowest valence was Scarborough Fair (M = 1.36), while the melodies with the highest valence were Star Wars Theme and The Entertainer (M = 3.44). The melody with the lowest arousal was Silent Night (M = 0.72) while the melody with the highest arousal was The Entertainer (M = 3.50). The melody with the lowest familiarity was This Little Light of Mine (M = 0.70) while the melodies with the highest familiarity were Here Comes the Bride and Jingle Bells (M = 3.90). The melody with the lowest age of acquisition was Old MacDonald (M = 2.30) while the melody with the highest age of acquisition was Sweet Caroline (M = 7.45). Finally, the melodies with the lowest naming percentage were YMCA and ABC (the Jackson Five song, not the "ABC Song" which shares its melody with Twinkle Twinkle Little Star) (0% correct), while the melody with the highest naming percentage was Happy Birthday (88.7% correct). See Appendix for a full list of melodies and the mean ratings on each of the five rating scales.

We also sought to investigate relationships among these variables by calculating Pearson's correlations between each of the pairs of variables. There were significant correlations between the following variables: valence and arousal (r = 0.70, t(105) = 10.04, p < 0.001, 95% CI:[0.58, 0.78]), valence and familiarity (r = 0.48, t(105) = 5.65, p < 0.001, 95% CI:[0.32, 0.61]), valence and age of acquisition (r = – 0.52, t(105) = – 6.10, p < 0.001, 95% CI:[– 0.63, – 0.35]), valence and naming (r = 0.42, t(105) = 4.78, p < 0.001, 95% CI:[0.25, 0.56]), naming and familiarity (r = 0.78, t(105) = 12.99, p < 0.001, 95% CI:[0.69, 0.84]), naming and age of acquisition (r = – 0.77, t(105) = – 12.63, p < 0.001, 95% CI:[– 0.84, – 0.68]), and familiarity and age of acquisition (r = – 0.90, t(105) = – 21.86, p < 0.001, 95% CI:[– 0.93, – 0.86]). See Figure 1 for a graphical depiction of these correlations.

Cluster analysis

When creating the stimulus set, melodies were classified into one of the following categories: children's, patriotic, movie/ TV, Christmas, religious, classical, pop, or "other" for melodies that did not fall into one of the above categories. One question was whether these categories systematically differed
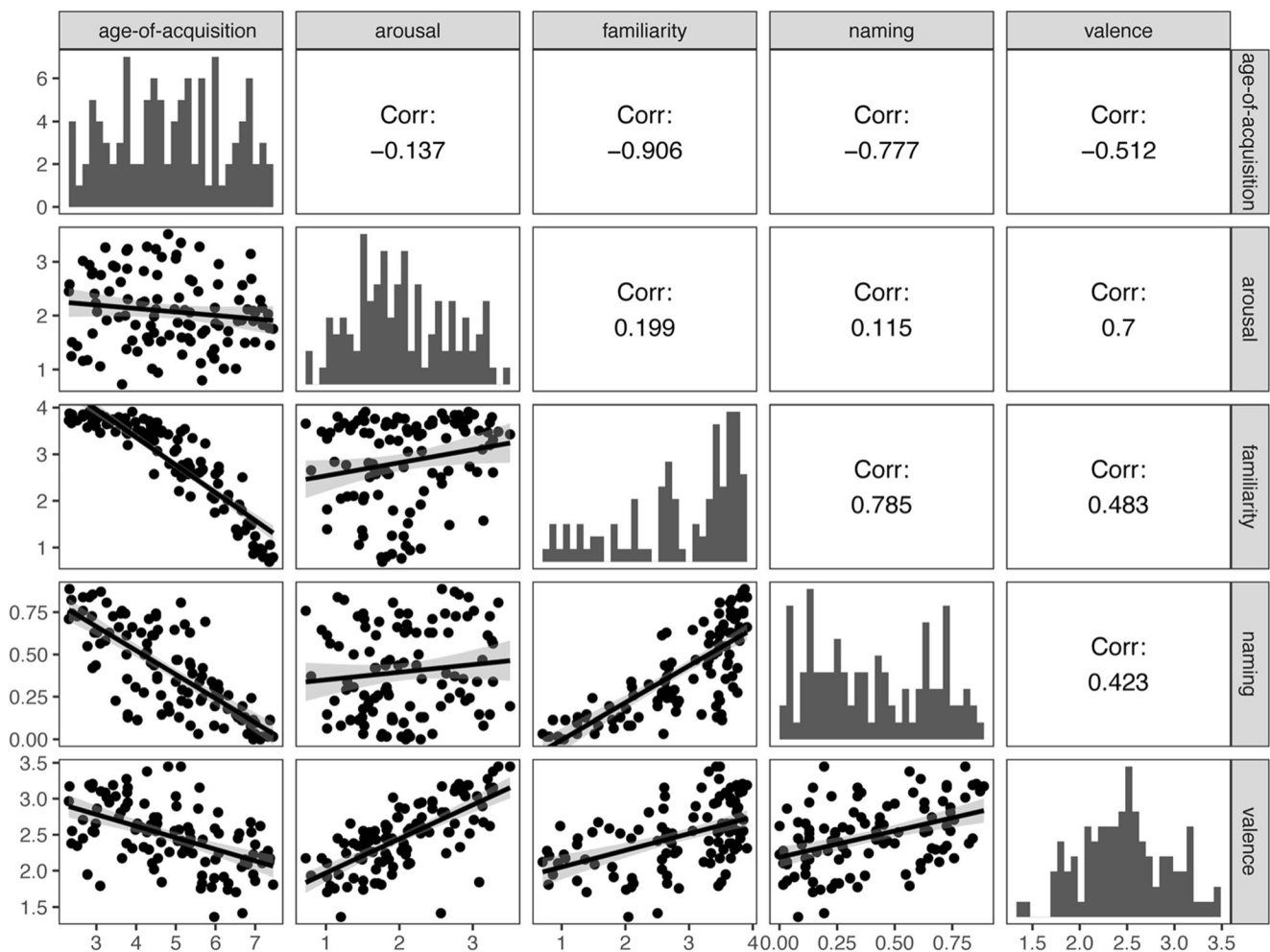
Fig. 1 Correlations between the five measured variables. *Lower panels* depict scatterplots with a linear regression line superimposed. *Gray band* surrounding regression line indicates the 95% confidence interval. *Diagonals* depict histograms indicating the frequency of responses on each variable. *Upper panels* indicate the *r* values for the correlations

in terms of their ratings on the five variables of interest. Our goal was to see whether or not we could 'group' melodies based on their normative ratings, and whether these groups would systematically map on to melody category. To this end, we conducted a k-means cluster analysis in an attempt to group melodies based on their normative ratings.

While we had an a priori prediction that our stimulus set would group into eight clusters (based on the eight pre-determined categories of children's, patriotic, movie/TV, Christmas, religious, classical, pop, or "other"), we first sought to identify the optimal number of clusters based on our data. We first standardized the ratings by converting them from the raw ratings to z-scores. Next, we conducted two analyses to identify the optimal number of clusters, using the NbClust function from the NbClust package (Charrad, Ghazzali, Boiteau, & Niknafs, 2014) and the fviz_nbclust function from the factoextra package in R. First, we identified the optimal number of clusters using the "elbow" method, which identifies the optimal number of clusters based on the total within-cluster sum of squares (WSS), a measure of

variation within each cluster. This method suggested that *two* was the optimal number of clusters, as two clusters had lower WSS than one cluster, but adding a third cluster did not substantially reduce the WSS (see Supplementary Figure 1A https://osf.io/wrqzm/). We also identified the optimal number of clusters using the silhouette method, which measures the similarity of (or, distance between) of objects within each cluster as compared to the similarity of items between clusters (Kaufman & Rousseeuw, 2005). A higher silhouette value indicates better clustering, such that within-cluster similarity is maximized while between-cluster similarity is minimized. This method also suggested two as the optimal number of clusters (see Supplementary Figure 1B https://osf.io/wrqzm/).

K-means clustering was performed using the kmeans function in R with a predetermined number of clusters set to two, based on our identification as two as the optimal number of clusters. There were 63 melodies in Cluster 1 and 44 melodies in Cluster 2. Cluster 1 consisted of melodies that were rated as more familiar, having a lower age of acquisition, and were

Table 1 Mean (SD) values for each variable for the two clusters

Cluster	Valence	Arousal	Familiarity	Age of acquisition	Naming
1	2.68 (0.40)	2.24 (0.72)	3.50 (0.32)	3.91 (0.92)	0.56 (0.20)
2	2.19 (0.35)	1.83 (0.51)	1.89 (0.70)	6.27 (0.73)	0.16 (0.11)

more frequently named than melodies in Cluster 2. Melodies in Cluster 1 also had higher average ratings of valence and arousal, although these emotional variables were not as clearly different between the two clusters (Table 1). See Fig. 2 for a graphical depiction of these clusters.

We also sought to investigate whether certain categories of melodies were more likely to fall into one cluster or another: for
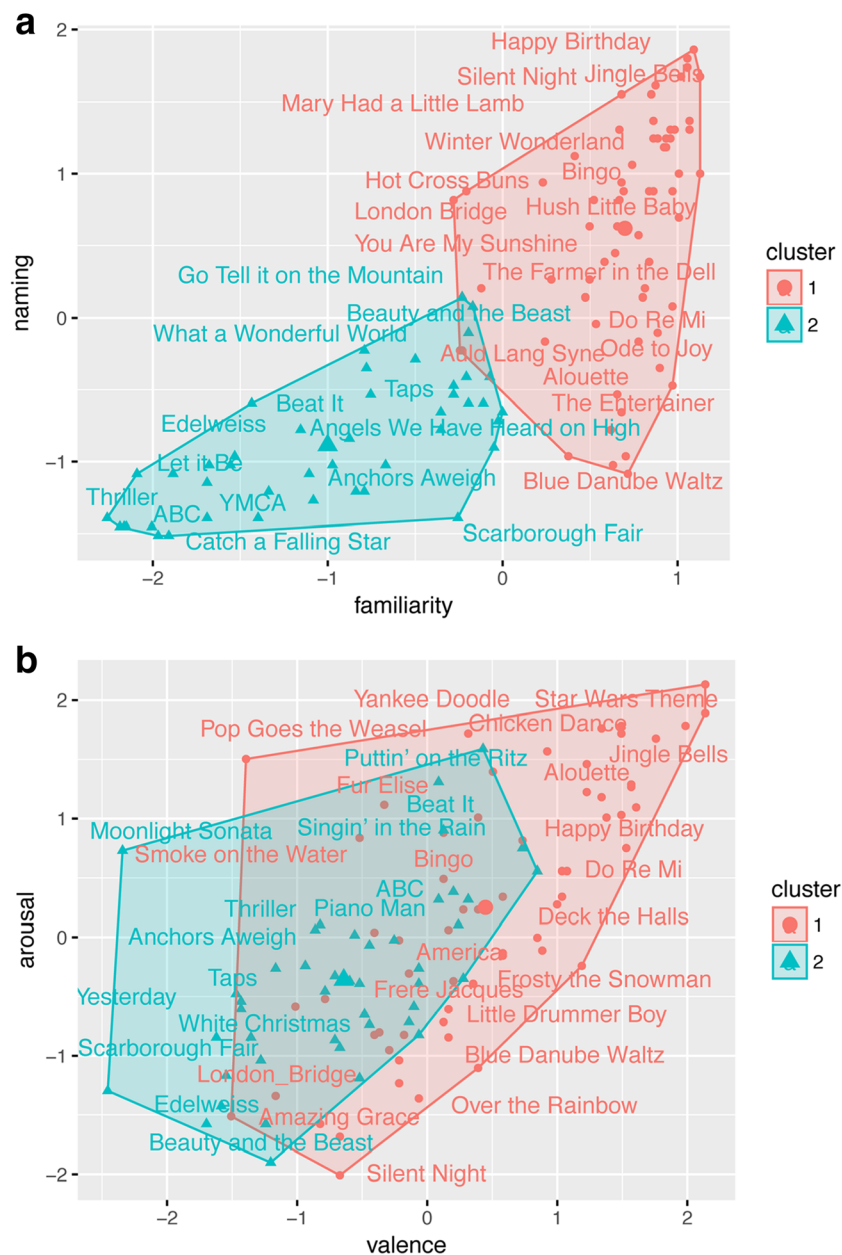

Fig. 2 Cluster visualization. Clusters 1 and 2 are mapped onto **a** naming and familiarity, and **b** valence and arousal. To preserve readability, not all melody names are included in the figure
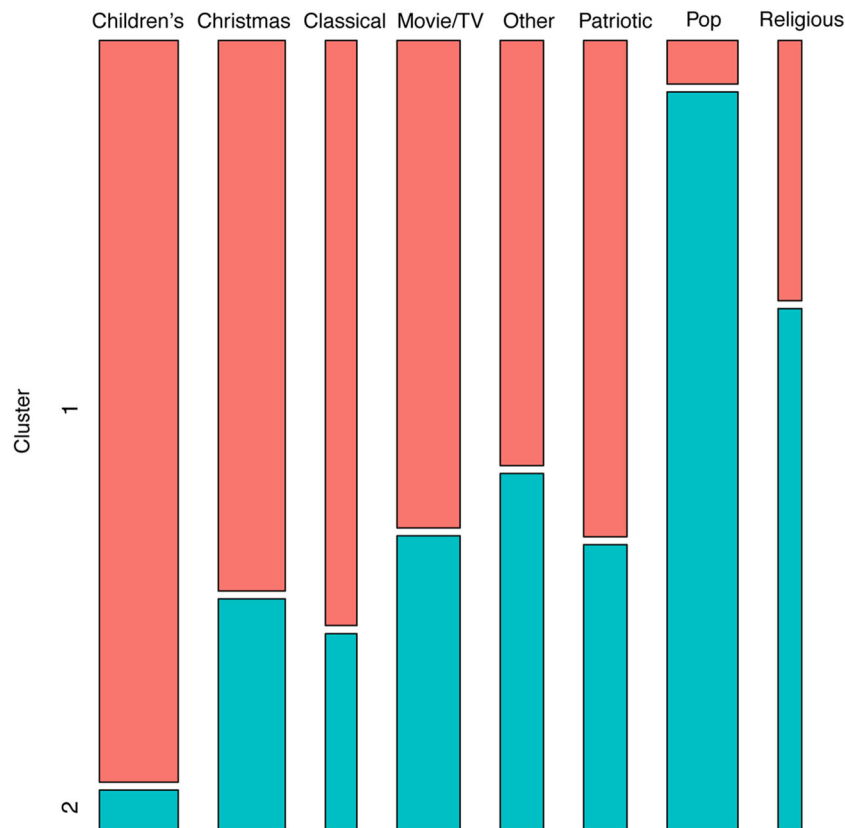
Fig. 3 Proportion of melodies in each category in Cluster 1 and 2. Cluster 1 is represented in *red*; Cluster 2 is represented in *blue*. All melody categories had a greater proportion of melodies in Cluster 1, except Pop and Religious melodies. *Bar width* represents total number of melodies in each category (e.g., there are more Children's melodies in the stimulus set [$n = 20$] than religious melodies [$n = 6$]). *Bar height* represents the proportion of melodies in each cluster

example, whether Christmas melodies were more likely to fall in Cluster 1 than Cluster 2. We conducted a Chi-square test to evaluate differences in the proportion of melodies in each category between clusters. This revealed a significant difference between clusters $X^2 = 35.63$, $p < 0.001$. Pairwise comparisons indicated that the only category that had significantly more melodies in Cluster 2 than 1 was "pop". That is, pop melodies were rated proportionally less familiar, had a higher age of acquisition, and were less frequently named (and therefore over-represented in Cluster 2). See Fig. 3 for a graphical depiction of the proportion of melodies in each cluster, and see Appendix for the full list of melodies and their cluster designation.

Interrater reliability

We sought to assess interrater reliability across the five variables by calculating the intraclass correlation coefficient using the "icc" function from the "irr" package in R. We calculated our ICCs using a two-way model using the "agreement" measure (McGraw & Wong, 1996). As in prior work in similar stimulus sets, we calculated both ICCs measured from single and average ratings (Livingstone & Russo, 2018), and have presented these in Table 2. These results indicate poor to moderate agreement across the five variables for single measures, but very high agreement for all five variables for average measures.

Table 2 Interrater reliability for each of the five rating scales

Rating	ICC type	ICC value	95% CI	
			Lower bound	Upper bound
Valence	Single	0.16	0.13	0.20
	Average	0.91	0.89	0.94
Arousal	Single	0.27	0.22	0.33
	Average	0.96	0.95	0.97
Familiarity	Single	0.38	0.32	0.45
	Average	0.98	0.97	0.99
Age of acquisition	Single	0.36	0.30	0.44
	Average	0.97	0.96	0.98
Naming	Single	0.28	0.23	0.34
	Average	0.96	0.95	0.97

Differences between AMT and undergraduate groups

Due to possible differences in ratings between the undergraduate and AMT groups, we provide these normative data for each group separately (see Supplementary Table 1 https://osf.io/wrqzm/). We first sought to examine whether the groups differed in their normative ratings on each stimulus. To this end, we conducted separate linear mixed-effects models for each variable (valence, arousal, familiarity, age of acquisition, and naming) using the *lme4* package in R (Bates, Maechler, Bolker, & Walker, 2015). For each model, we considered group (AMT vs. undergraduate) and stimulus as fixed-effects, and included random intercepts for participants. Therefore, we treated each stimulus as a condition, and sought to identify differences between groups for each condition. The *car* package in R was used to calculate *p* values for the regression coefficients (Fox & Weisberg, 2011).

Most stimuli did not significantly differ between groups, although this depended on the variable. The two groups tended to agree more on ratings of arousal (only five out of 107 melodies significantly differed between groups), and disagreed most on familiarity (13 out of 107 melodies significantly differed). A few melodies were consistently rated differently across rating scales: For example, the Harry Potter Theme ("Hedwig's Theme") was more arousing, more familiar, had an earlier age of acquisition, and was named significantly more frequently by the undergraduate than the AMT group. Overall, arousal and valence ratings tended to be more consistent across groups than familiarity, age of acquisition, or naming. See Supplementary Table 1 for the full list of stimuli with their normative ratings for each groups and indications of which ratings differed between groups.

Discussion

The goal of the present work was to create a set of famous musical melodies with normative data on several cognitive and affective dimensions. First, our stimulus set was designed by asking a large online sample of participants to provide names of famous and highly recognizable melodies. The resulting Famous Melodies Stimulus Set contains 107 melodies that were rated by 338 participants. Melodies were rated on their perceived emotional valence, emotional arousal, familiarity, age of acquisition; in addition to these ratings, participants provided the name of each melody, and a corresponding correct naming percentage was calculated for each. Overall, the familiarity of our stimulus set was validated: Most melodies were rated as highly familiar and were typically named correctly. This suggests that the Famous Melodies Stimulus Set contains melodies that are familiar and recognizable to a United States population.

Despite our intended goal of creating a stimulus set of highly recognizable melodies, not *all* melodies were highly familiar. Cluster analysis revealed a subset of melodies that were rated as less familiar and were less likely to be named. This suggests that the stimulus set contains two sub-groups of melodies: One subgroup contains melodies that are highly familiar, likely to be named, and have an early age of acquisition. The second subgroup contains melodies that are less familiar, less likely to be named, and have a later age of acquisition. It may seem surprising that *any* of the melodies were rated as highly *un*familiar, since melodies were included precisely because participants suggested them as highly familiar and recognizable. To investigate what might be contributing to this paradoxical unfamiliarity of some of the melodies, we examined whether certain categories of melodies were more likely to fall into the unfamiliar cluster. In doing so, we found that melodies in the "pop" category were overrepresented in this cluster of less familiar melodies.

To interpret why "pop" melodies were rated as unfamiliar, despite being some of the most frequently named melodies in our initial development of the stimulus set (e.g., "Beat It" by Michael Jackson was named by multiple participants), it is important to consider the nature of the melodies in the set. Our stimuli were not naturalistic music, but instead, the melodies were created as a single-line melody in a piano keyboard timbre. Most listeners are likely not accustomed to hearing popular music in this form. In contrast, children's tunes and Christmas songs are often sung a capella, and are arranged in many different ways and performed by many different artists, so listeners may be more familiar with the "pure" melody itself for those categories of melodies. Therefore, the abstracted nature of the "pure" melodies used here may make pop melodies more challenging to identify. One advantage of this variance in familiarity is that researchers can select stimuli with varying degrees of familiarity, based on the goals of their experiment. For example, prior work has typically compared performance between familiar and unfamiliar melodies (Baird & Samson, 2009; Dalla Bella et al., 2003). It also may be useful for researchers who wish to compare responses to the pure melody (from the present stimulus set) with the naturalistic music itself (for example, comparing the song "Beat It" with the melody version from the stimulus set).

While some melodies were rated as unfamiliar, the overall distribution of melodies in the familiarity rating space was skewed towards high familiarity. In contrast, responses on the other rating scales were more evenly distributed (see Fig. 1). This has the benefit of allowing researchers to pick stimuli for certain purposes. For example, researchers may want to select stimuli that are equally familiar, but vary in age of acquisition. Another

feature of the stimulus set is that researchers can select stimuli that are highly familiar but not "nameable." For example, "Battle Hymn of the Republic" was rated as highly familiar, but had a low percentage of correct naming. These types of melodies may be useful for research investigating differences between recognition (or familiarity) and identification (or naming) of melodies (Belfi & Tranel, 2014; Kostic & Cleary, 2009). While we identified two clusters of melodies, one of which contained highly familiar, and the other less familiar melodies, emotional variables did not vary between the two clusters. This suggests that emotional arousal and valence are distinct from melody familiarity. Therefore, researchers may also select pieces that are matched on valence and arousal, but differ in terms of familiarity or naming ability.

Although there were some melodies rated as highly familiar but with low rates of correct naming, we generally found strong correlations between these variables. That is, melodies that were more familiar were more likely to be named. Familiarity also had a strong inverse relationship with age of acquisition: Melodies that were highly familiar were more likely to be learned at an early age. These correlations were perhaps strikingly high, particularly since each participant rated all melodies on only one rating scale (so the ratings were made independently across rating dimensions). This strong relationship between familiarity and age of acquisition is consistent with prior research suggesting that an early age of acquisition facilitates lexical retrieval for item names, and that items learned later in life are more susceptible to loss (Bell, Davies, Hermann, & Walters, 2000; Tzortzis, Goldblum, Dang, Forette, & Boller, 2000).

Despite high correlations among some variables, such as age of acquisition and familiarity, we found mixed interrater reliability. For all variables, interrater reliability was low to moderate for 'single' measures. These values are similar to measures of interrater reliability for other musical stimulus sets (Livingstone & Russo, 2018). In contrast, we found high interrater reliability for 'average' measures. For both single and average measures, interrater reliability tended to be higher for familiarity and age of acquisition than emotional valence and arousal. The interrater reliability values found here are consistent with our prior work investigating emotional and aesthetic judgments of music (Belfi, 2019) and poetry (Belfi, Vessel, & Starr, 2017). In both cases, interrater reliability was relatively low for judgments of valence, arousal, and aesthetic appeal. It seems to be the case that aesthetic objects, such as music, tend to have lower interrater reliability, which may reflect the fact that individual differences play a large role in musical preference (North, 2010; Palmer & Griscom, 2013).

A secondary goal of the present work was to compare our results from an undergraduate sample to a sample of participants from AMT. We found that responses from AMT and undergraduate participants were quite similar, despite differences in demographic variables between the two samples. There were, however, a few notable differences between the two samples: For example, the Harry Potter theme (Hedwig's Theme) was consistently rated differently between the two groups. The undergraduate sample rated this melody as significantly more arousing, more familiar, and as having a lower age of acquisition than the AMT group. The undergraduate group also named this melody significantly more frequently (91%) than the AMT group (41%). This difference in this specific melody possibly reflects age differences between the undergraduate and AMT samples. With some notable exceptions such as this, there were few melodies that significantly differed between groups on multiple rating scales. We therefore feel that the Famous Melodies Stimulus Set is suitable for use in a range of experimental settings. See Supplementary Table 1 for the full list of stimuli with their normative ratings separated for each group.

The current stimulus set improves on prior work by developing a normed set of famous melodies with ratings on several dimensions. However, this work is not without limitations. First, while we attempted to collect a large number of famous melodies for use in our stimulus set, it was not a completely exhaustive set of all highly familiar melodies. Furthermore, the majority of the melodies here are taken from lyrical songs. Lyrical melodies greatly outnumber non-lyrical in our set (89 with lyrics, 18 without), to the point where comparing data between lyrical and non-lyrical melodies would be difficult. We suspect that our stimulus set does not contain many non-lyrical melodies because participants are often familiar with such melodies, but are unable to name them (such as highly familiar classical music). Since our stimulus selection procedure relied on participants naming melodies (and our stimulus set was created with the goal of containing melodies that could be named), this likely excluded many of these familiar non-lyrical melodies. Therefore, the Famous Melodies Stimulus Set would likely be less appropriate for investigating research questions aimed at comparing lyrical versus non-lyrical music. And although the Famous Melodies Stimulus Set contains normative ratings of emotional features (valence and arousal) these stimuli would also likely be less suitable for use as a purely emotion-inducing stimulus. While it is not unlikely that hearing these melodies could induce emotions, it may be that they induce such emotions via autobiographical memories, feelings of nostalgia, or other semantic associations. Additionally, the present stimulus set had

relatively low interrater reliability on the emotional ratings. Therefore, if a researcher were interested in studying the 'pure' effect of music on emotions, it would be more appropriate to use stimulus sets designed specifically to evoke particular emotions (e.g., Lepping et al., 2016).

Although there are certainly types of studies where other musical stimuli would be more suitable, the Famous Melodies Stimulus Set has a wide range of applications for the study of cognitive functions such as memory and language. As has been shown in prior work, these melodies could be used to investigate memory for melodies in dementia (Cuddy & Duffin, 2005; Hsieh et al., 2011) or conceptual and lexical retrieval in lesion patient populations (Belfi et al., 2019; Belfi & Tranel, 2014). These stimuli could be particularly useful in other neuropsychological studies, for example studies

examining singing behavior in persons with aphasia (Kasdan & Kiran, 2018). They could also be used to study feelings of familiarity for melodies (Filipic et al., 2010), or memory for musical lyrics. The melodies in the Famous Melodies Stimulus set all consist of single-line piano MIDI melodies, and could therefore provide an interesting set to compare to more 'naturalistic' musical excerpts.

To conclude, famous melodies are frequently used in research in both healthy and clinical populations. The Famous Melodies Stimulus Set has several distinct advantages over prior work: First, the present stimulus set has been validated on a large set of participants from the United States. As the goal of the present work was to create a set of stimuli that were highly familiar, it is important to create a regionally validated stimulus set. As prior research has shown, musical familiarity is strongly driven by cultural influences, which can affect music perception in many

Appendix

Table 3.

Melody	Category	Cluster	Valence	Arousal	Familiarity	Age	Naming
ABC	Pop	2	2.62	2.29	0.98	7.13	0.00
Alouette	Children	1	3.03	2.90	3.48	3.48	0.23
Amazing Grace	Religious	1	2.17	0.94	3.48	4.55	0.65
America (My Country 'tis of Thee)	Patriotic	1	2.41	1.87	3.31	5.05	0.47
America The Beautiful	Patriotic	1	2.64	1.81	3.61	4.37	0.45
Anchors Aweigh	Patriotic	2	2.05	1.91	1.79	6.66	0.11
Angels We Have Heard on High	Christmas	2	2.26	1.64	2.79	4.84	0.16
Auld Lang Syne	Other	1	2.40	1.52	3.46	4.63	0.26
Battle Hymn of the Republic	Patriotic	1	2.95	2.30	3.44	3.79	0.13
Beat It	Pop	2	2.53	2.68	1.48	6.90	0.24
Beauty and the Beast	Movie/TV	2	1.93	0.80	2.66	5.66	0.37
Bingo	Children	1	2.53	2.41	3.63	2.95	0.63
Blue Danube Waltz	Classical	1	2.66	1.33	3.52	4.03	0.11
Camp Town Races	Other	1	3.09	2.87	3.57	3.77	0.35
Can You Feel the Love Tonight	Movie/TV	2	2.16	1.49	1.37	6.74	0.13
Carol of the Bells	Christmas	1	2.66	2.75	3.68	4.53	0.37
Catch a Falling Star	Other	2	2.09	2.12	1.52	6.81	0.03
Chicken Dance	Children	1	3.38	3.28	3.31	4.27	0.56
Deck The Halls	Christmas	1	2.86	2.07	3.76	3.02	0.63
Do Re Mi	Movie/TV	1	2.74	2.30	3.07	4.53	0.35
Don't Stop Believing	Pop	2	2.17	1.45	1.06	7.37	0.11
Edelweiss	Movie/TV	2	1.71	1.01	1.39	6.52	0.15
Fly Me To The Moon	Pop	2	1.86	1.51	1.24	7.00	0.03
For He's a Jolly Good Fellow	Other	1	2.93	2.26	3.39	4.21	0.50

(continued)

Melody	Category	Cluster	Valence	Arousal	Familiarity	Age	Naming
Frere Jacques	Children	1	2.55	1.67	3.76	2.90	0.42
Frosty the Snowman	Christmas	1	3.02	1.91	3.64	3.26	0.81
Fur Elise	Classical	1	2.33	2.83	3.76	4.19	0.27
Go Tell it On The Mountain	Christmas	2	2.45	1.81	2.62	5.37	0.44
God Bless America	Patriotic	1	2.74	1.99	3.29	4.45	0.44
Greensleeves	Christmas	2	1.76	1.12	2.84	5.63	0.23
Hallelujah Chorus	Classical	1	2.90	3.13	3.10	5.02	0.47
Happy and You Know It	Children	1	3.16	2.77	3.66	2.89	0.73
Happy Birthday	Children	1	3.17	2.58	3.87	2.32	0.89
Harry Potter Theme	Movie/TV	1	2.12	1.72	3.23	5.74	0.69
Have Yourself a Merry Little Christmas	Christmas	2	2.24	1.28	2.77	5.16	0.29
Here Comes the Bride	Other	1	2.31	1.54	3.91	3.90	0.66
He's Got the Whole World in His Hands	Religious	2	2.59	2.14	1.75	5.98	0.19
Hey Jude	Pop	2	2.24	1.81	2.64	6.08	0.29
Home on the Range	Other	2	2.60	1.84	2.57	5.31	0.27
Hot Cross Buns	Children	1	2.02	1.68	2.57	4.45	0.61
Hush Little Baby	Children	1	1.79	1.06	3.46	3.10	0.56
In the Hall of the Mountain King	Classical	1	1.84	3.09	3.51	4.65	0.15
Itsy Bitsy Spider	Children	1	2.45	1.16	3.80	2.66	0.84
I've Been Working on the Railroad	Children	1	2.95	2.45	3.66	3.32	0.63
Jeopardy Theme	Movie/TV	1	2.29	2.10	3.54	5.23	0.68
Jesus Loves Me	Religious	2	2.45	1.52	2.09	5.37	0.08
Jingle Bells	Christmas	1	3.19	2.94	3.91	2.82	0.84
Joy to the World	Christmas	1	2.88	2.00	3.77	3.73	0.74
Lean on Me	Pop	2	2.21	1.83	0.80	7.19	0.02
Let it Be	Pop	2	2.36	2.06	1.24	7.00	0.10
Little Drummer Boy	Christmas	1	2.29	1.52	3.71	4.32	0.71
London Bridge	Children	1	1.95	1.17	3.57	2.77	0.55
Love Me Tender	Pop	2	1.91	1.01	1.82	6.21	0.06
Mary Had a Little Lamb	Children	1	2.34	1.43	3.72	2.52	0.73
Minuet in G	Classical	2	2.86	2.45	2.60	5.56	0.03
Moonlight Sonata	Classical	2	1.41	2.57	2.51	6.68	0.19
My Bonnie Lies Over the Ocean	Children	2	2.41	1.59	2.51	5.15	0.23
My Favorite Things	Movie/TV	1	2.38	2.06	2.72	5.35	0.45
O Christmas Tree	Christmas	1	2.74	1.97	3.79	3.56	0.66
Ode to Joy	Classical	1	2.57	1.83	3.69	3.98	0.31
Old MacDonald	Children	1	2.97	2.45	3.72	2.31	0.71
On Top of Old Smokey	Other	2	2.28	1.58	2.82	5.31	0.21
Over the Rainbow	Movie/TV	1	2.10	1.01	3.33	4.40	0.61
Piano Man	Pop	2	2.28	2.03	0.77	7.32	0.02
Pomp and Circumstance	Other	2	2.16	1.86	2.57	6.00	0.26
Pop Goes the Weasel	Children	1	2.71	3.01	3.85	2.66	0.76
Puttin' on the Ritz	Pop	2	2.52	2.96	2.74	6.08	0.24
Row, Row, Row Your Boat	Children	1	2.55	1.51	3.85	2.39	0.74
Rudolph the Red-Nosed Reindeer	Christmas	1	3.10	2.75	3.84	3.11	0.87
Scarborough Fair	Other	2	1.36	1.20	2.05	5.97	0.08
She'll Be Coming Round The Mountain	Children	1	3.28	3.20	3.47	3.76	0.61
Silent Night	Christmas	1	2.17	0.72	3.66	3.65	0.76
Singin' in the Rain	Movie/TV	2	2.81	2.58	2.37	6.06	0.32

(continued)

Melody	Category	Cluster	Valence	Arousal	Familiarity	Age	Naming
Smoke on the Water	Pop	1	2.24	2.64	3.34	5.42	0.39
Star Spangled Banner	Patriotic	1	2.81	2.62	3.75	3.61	0.74
Star Wars Theme	Movie/TV	1	3.45	3.35	3.48	5.13	0.81
Super Mario Brothers	Other	1	2.53	2.67	2.64	5.35	0.63
Supercalifragilisticexpialidocious	Movie/TV	1	3.16	3.28	2.61	5.60	0.34
Sweet Caroline	Pop	2	1.81	1.75	0.79	7.45	0.02
Swing Low, Sweet Chariot	Religious	2	1.74	1.51	2.01	6.31	0.18
Take Me Out to the Ballgame	Other	1	2.55	2.12	3.47	4.50	0.74
Taps	Patriotic	2	1.83	1.67	2.66	5.66	0.24
The Entertainer	Movie/TV	1	3.45	3.51	3.43	4.81	0.19
The Farmer in the Dell	Children	1	2.66	2.23	3.60	2.98	0.44
The Pink Panther	Movie/TV	1	2.62	3.23	3.68	4.50	0.73
This Land is Your Land	Patriotic	1	2.53	1.59	3.45	4.27	0.52
This Little Light of Mine	Religious	2	2.12	1.77	0.70	7.35	0.03
Thriller	Pop	2	2.10	2.14	0.94	7.18	0.02
Tomorrow	Movie/TV	2	2.43	1.68	2.68	5.68	0.42
Twelve Days of Christmas	Christmas	1	3.19	2.93	3.75	3.42	0.73
Twinkle, Twinkle Little Star	Children	1	2.38	1.25	3.67	2.37	0.82
US Air Force Theme	Patriotic	2	2.52	2.29	1.92	6.60	0.13
Washington Post March	Patriotic	2	2.67	3.14	1.57	6.89	0.08
We Will Rock You	Pop	2	1.95	1.90	0.86	6.95	0.11
We Wish You a Merry Christmas	Christmas	1	3.21	2.81	3.84	2.90	0.85
What A Wonderful World	Pop	2	1.78	1.29	2.09	5.90	0.34
When the Saints Go Marching In	Religious	1	2.60	2.23	3.49	4.13	0.63
Whistle While You Work	Movie/TV	2	2.57	2.33	2.21	5.08	0.13
White Christmas	Christmas	2	1.90	1.38	2.10	5.97	0.31
William Tell Overture	Classical	1	3.16	3.23	3.20	3.79	0.15
Winter Wonderland	Christmas	1	2.55	2.12	3.06	4.97	0.65
Yankee Doodle	Patriotic	1	3.09	3.26	3.79	3.23	0.58
Yesterday	Pop	2	1.83	1.71	2.13	6.32	0.26
YMCA	Pop	2	2.22	2.09	1.03	6.95	0.00
You Are My Sunshine	Children	1	2.38	1.38	3.63	3.76	0.50
You've Got a Friend In Me	Pop	2	2.45	1.90	1.25	6.56	0.13
Zip-a-dee-doo-dah	Movie/TV	1	3.03	3.06	2.62	5.00	0.34

ways (Hannon, Soley, & Ullal, 2012; Teo, Hargreaves, & Lee, 2016). Although prior work has used sets of famous musical stimuli, some of these were normed in other countries, making them less suitable for US-based researchers. In addition to having the distinct advantage of being normed on a US-based sample, these stimuli are also openly available. They are accompanied by normative data from a large number of subjects on several dimensions: emotional valence, emotional arousal, age of acquisition, familiarity, and naming. Unlike other musical stimulus sets, which have typically just focused on emotional features of music, the Famous Melodies Stimulus Set has both emotional ratings as well as ratings typically associated with everyday objects (such as tools, animals, or persons), which will allow for a broad usage of the Famous Melodies Stimulus set. Finally, each melody was rated by a large number of participants, and the ratings are unconfounded, since raters rated each melody on only one rating dimension. We hope that the Famous Melodies Stimulus Set will serve as a valuable resource for researchers studying all aspects

of music cognition and perception, as well as cognitive functioning more broadly.

References

Ayotte, J., Peretz, I., Rousseau, I., Bard, C., & Bojanowski, M. (2000). Patterns of music agnosia associated with middle cerebral artery infarcts. *Brain*, *123*, 1926–1938.

Bailes, F. (2010). Dynamic melody recognition: Distinctiveness and the role of musical expertise. *Memory & Cognition*, *38*(5), 641–650. https://doi.org/10.3758/MC.38.5.641

Baird, A., & Samson, S. (2009). Memory for music in Alzheimer's disease: unforgettable? *Neuropsychology Review*, *19*(1), 85–101.

Baird, A., Walker, D. G., Biggs, V., & Robinson, G. A. (2014). Selective preservation of the beat in apperceptive music agnosia : A case study. *Cortex*, *53*, 27–33. https://doi.org/10.1016/j.cortex.2014.01.005

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Belfi, A. M. (2019). Emotional Valence and Vividness of Imagery Predict Aesthetic Appeal in Music. *Psychomusicology: Music, Mind, and Brain*, *29*, 128–135.

Belfi, A. M., Karlan, B., & Tranel, D. (2016). Music evokes vivid autobiographical memories. *Memory*, *24*, 979–989.

Belfi, A. M., Kasdan, A., & Tranel, D. (2019). Anomia for musical entities. *Aphasiology*, *33*(4), 382–404.

Belfi, A. M., & Tranel, D. (2014). Impaired naming of famous musical melodies is associated with left temporal polar damage. *Neuropsychology*, *28*(3), 429–435. https://doi.org/10.1037/neu0000051

Belfi, A. M., Vessel, E. A., & Starr, G. G. (2017). Individual ratings of vividness predict aesthetic appeal in poetry. *Psychology of Aesthetics, Creativity, and the Arts*, *12*, 341–350. https://doi.org/10.1037/aca0000153

Bell, B. D., Davies, K. G., Hermann, B. P., & Walters, G. (2000). Confrontation naming after anterior temporal lobectomy is related to age of acquisition of the object names. *Neuropsychologia*, *38*(1), 83–92. https://doi.org/10.1016/S0028-3932(99)00047-0

Büdenbender, N., & Kreutz, G. (2016). Long-term representations of melodies in Western listeners: Influences of familiarity, musical expertise, tempo and structure. *Psychology of Music*, 030573561667140. https://doi.org/10.1177/0305735616671408

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6).

Chedid, G., Wilson, M. A., Bedetti, C., Rey, A. E., Vallet, G. T., & Brambati, S. M. (2018). Norms of conceptual familiarity for 3,596 French nouns and their contribution in lexical decision. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1106-8

Cuddy, L. L., & Duffin, J. (2005). Music, memory, and Alzheimer's disease: Is music recognition spared in dementia, and how can it be assessed? *Medical Hypotheses*, *64*(2), 229–235.

Cuddy, L. L., Duffin, J. M., Gill, S. S., Brown, C. L., Sikka, R., & Vanstone, A. D. (2012). Memory for melodies and lyrics in Alzheimer's disease. *Music Perception*, *29*(5), 479–491.

Cuddy, L. L., Sikka, R., & Vanstone, A. (2015). Preservation of musical memory and engagement in healthy aging and Alzheimer's disease. *Annals of the New York Academy of Sciences*, *1337*(1), 223–231. https://doi.org/10.1111/nyas.12617

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: a gating paradigm study. *Perception & Psychophysics*, *65*, 1019–1028. https://doi.org/10.3758/BF03194831

Darcy, I., & Fontaine, N. M. G. (2019). The Hoosier Vocal Emotions Corpus: A validated set of North American English pseudo-words for evaluating emotion processing. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01288-0

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Deyne, S. De, Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*, 1–20. https://doi.org/10.3758/s13428-018-1115-7

Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC Neuroscience*, *9*, 48. https://doi.org/10.1186/1471-2202-9-48

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review*, *17*, 335–341.

Finke, C., Esfahani, N. E., & Ploner, C. J. (2012). Preservation of musical memory in an amnesic professional cellist. *Current Biology*, *22*(15), R591–R592. https://doi.org/10.1016/j.cub.2012.05.041

Ford, J. H., Addis, D. R., & Giovanello, K. S. (2011). Differential neural activity during search of specific and general autobiographical memories elicited by musical cues. *Neuropsychologia*, *49*, 2514–2526.

Fox, J., & Weisberg, S. (2011). An R Companion to Applied Regression (Second). Thousand Oaks, CA: Sage.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., … Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842. https://doi.org/10.3758/s13428-015-0642-8

Hailstone, J. C., Omar, R., & Warren, J. D. (2009). Relatively preserved knowledge of music in semantic dementia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *80*(7), 808–809.

Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 543–548. https://doi.org/10.1037/a0027225

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382–1397. https://doi.org/10.1162/jocn_a_00216

Hsieh, S., Hornberger, M., Piguet, O., & Hodges, J. R. (2011). Neural basis of music knowledge: Evidence from the dementias. *Brain*, *134*, 2523–2534.

Huijgen, J., Dellacherie, D., Tillmann, B., Clément, S., Bigand, E., Dupont, S., & Samson, S. (2015). The feeling of familiarity for music in patients with a unilateral temporal lobe lesion: A gating study. *Neuropsychologia*, *77*, 313–320. https://doi.org/10.1016/j.neuropsychologia.2015.09.007

Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterization of music-evoked autobiographical memories. *Memory*, *15*, 845–860.

Johnson, J. K., Chang, C.-C., Brambati, S. M., Migliaccio, R., Gorno-Tempini, M. L., Miller, B. L., & Janata, P. (2011). Music

recognition in frontotemporal lobar degeneration and Alzheimer disease. *Cognitive and Behavioral Neurology*, 24(2), 74–84.

Kasdan, A., & Kiran, S. (2018). Please don't stop the music: Song completion in patients with aphasia. *Journal of Communication Disorders*, 75(June), 72–86. https://doi.org/10.1016/j.jcomdis.2018.06.005

Kaufman, L., & Rousseeuw, P. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley.

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: an ERP Study. *Journal of Cognitive Neuroscience*, 17(10), 1565–1577. https://doi.org/10.1162/089892905774597290

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage*, 81, 49–60. https://doi.org/10.1016/j.neuroimage.2013.05.008

Kostic, B., & Cleary, A. M. (2009). Song recognition without identification: When people cannot "Name That Tune" but can recognize it as familiar. *Journal of Experimental Psychology: General*, 138, 146–159. https://doi.org/10.1037/a0014584

Kurdi, B., Loano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set. *Behavior Research Methods*, 49(2), 457–470.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Technical Report A-8. Gainesville, Florida.

Larrouy-Maestri, P., Harrison, P. M. C., Müllensiefen D. (2019). The mistuning perception test: A new measurement instrument. Behavior Research Methods 51 (2):663-675

Lepping, R. J., Atchley, R. A., & Savage, C. R. (2016). Development of a validated emotionally provocative musical stimulus set for research. *Psychology of Music*, 44(5), 1012–1028. https://doi.org/10.1177/0305735615604509

Liégeois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, 121, 1853–1867.

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE*, 13(5). https://doi.org/10.1371/journal.pone.0196391

Macoir, J., Berubé-lalancette, S., Wilson, M. A., Hudon, C., Gravel, P., Potvin, O., … Laforce, R. (2016). When the Wedding March becomes sad: Semantic memory impairment for music in the semantic variant of primary progressive aphasia. *Neurocase*, 22(6), 486–495. https://doi.org/10.1080/13554794.2016.1257025

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlations coefficients. *Psychological Methods*, 1, 390–390. https://doi.org/10.1037/1082-989X.1.4.390

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. https://doi.org/10.1037/a0028085

North, A. C. (2010). Individual differences in musical taste. *American Journal of Psychology*, 123(2), 25–45. https://doi.org/10.5406/amerjpsyc.123.2.0199

Olszewski, C., Gfeller, K., Froman, R., Stordahl, J., & Tomblin, B. (2005). Familiar melody recognition by children and adults using cochlear implants and normal hearing children. *Cochlear Implants International*, 6(3), 123–140. https://doi.org/10.1002/cii.5

Omar, R., Hailstone, J. C., & Warren, J. D. (2012). Semantic memory for music in dementia. *Music Perception*, 29, 467–477.

Omar, R., Hailstone, J. C., Warren, J. E., Crutch, S. J., & Warren, J. D. (2010). The cognitive organization of music knowledge: A clinical analysis. *Brain*, 133, 1200–1213.

Palmer, S. E., & Griscom, W. S. (2013). Accounting for taste: individual differences in preference for harmony. *Psychonomic Bulletin &*

Review, 20(3), 453–461. https://doi.org/10.3758/s13423-012-0355-2

Peretz, I., Babai, M., Lussier, I., Hebert, S., & Gagnon, L. (1995). Corpus d'extraits musicaux: Indices relatifs a la familiarite , a l'age d'acquisition et aux evocations verbales. *Revue Canadienne de Psychologie*, 211–238.

Rainsford, M., Palmer, M. A., & Paine, G. (2018). The MUSOS (MUsic SOftware System) Toolkit: A computer-based, open source application for testing memory for melodies. *Behavior Research Methods*, 50, 684–702. https://doi.org/10.3758/s13428-017-0894-6

Schneider, B., Heskje, J., Bruss, J., Tranel, D., & Belfi, A. M. (2018). The left anterior temporal lobe is a convergence region mediating the relation between names and semantic knowledge for unique entities. *Cortex*, 109, 14–24. https://doi.org/10.1016/j.cortex.2018.08.026

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2018). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1099-3

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215.

Steinke, W., Cuddy, L., & Jakobson, L. (2001). Dissociations among functional systems governing melody recognition after right-hemisphere damage. *Cognitive Neuropsychology*, 18(5), 411–437.

Stordahl, J. (2002).Song recognition and appraisal: a comparison of children who use cochlear implants and normally hearing children. *Journal of Music Therapy*, 39(1), 2–19. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12015808

Teo, T., Hargreaves, D. J., & Lee, J. (2016). Musical Preference, Identification , and Familiarity : A Multicultural Comparison of Secondary Students from Singapore and the United Kingdom Linked references are available on JSTOR for this article :, 56(1), 18–32.

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia: Evidence from a gating paradigm. *Cortex*, 59, 84–94. https://doi.org/10.1016/j.cortex.2014.07.012

Tranel, D., Enekwechi, N., & Manzel, K. (2005). A test for measuring recognition and naming of landmarks. *Journal of Clinical and Experimental Neuropsychology*, 27(1), 102–126. https://doi.org/10.1080/138033990513663

Tranel, D., Logan, C. G., Frank, R. J., & Damasio, A. R. (1997). Explaining category-related effects in the retrieval of conceptual and lexical knowledge for concrete entities: Operationalization and analysis of factors. *Neuropsychologia*, 35, 1329–1339.

Tzortzis, C., Goldblum, M., Dang, M., Forette, F., & Boller, F. (2000). Absence of amusia and preserved naming of musical instruments in an aphasic composer. *Cortex*, 1(1987), 227–242.

Vanstone, A. D., Sikka, R., Tangness, L., Sham, R., Garcia, A., & Cuddy, L. L. (2012). Episodic and semantic memory for melodies in Alzheimer's disease. *Music Perception*, 21, 501–507.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, 22, 720–752.

Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., & Gordon, K. A. (2014). Children's identification of familiar songs from pitch and timing cues. *Frontiers in Psychology*, 5(AUG), 1–9. https://doi.org/10.3389/fpsyg.2014.00863

Zäske, R., Skuk, V. G., Golle, J., & Schweinberger, S. R. (2019). The Jena Speaker Set (JESS)—A database of voice stimuli from unfamiliar young and old adult speakers. *Behavior Research Methods* https://doi.org/10.3758/s13428-019-01296-0