Neural processing of poems and songs is based on melodic properties

Mathias Scharinger [a,b,c,*], Christine A. Knoop [a,d], Valentin Wagner [a,e], Winfried Menninghaus [a]

[a] Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany
[b] Research Group Phonetics, Institute of German Linguistics, Philipps-University Marburg, Pilgrimstein 16, Marburg 35032, Germany
[c] Center for Mind, Brain and Behavior, Universities of Marburg and Gießen, Germany
[d] Department of Music, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany
[e] Experimental Psychology Unit, Helmut Schmidt University / University of the Federal Armed Forces Hamburg, Germany

ARTICLE INFO

ABSTRACT

The neural processing of speech and music is still a matter of debate. A long tradition that assumes shared processing capacities for the two domains contrasts with views that assume domain-specific processing. We here contribute to this topic by investigating, in a functional magnetic imaging (fMRI) study, ecologically valid stimuli that are identical in wording and differ only in that one group is typically spoken (or silently read), whereas the other is sung: poems and their respective musical settings. We focus on the melodic properties of spoken poems and their sung musical counterparts by looking at proportions of significant autocorrelations (PSA) based on pitch values extracted from their recordings. Following earlier studies, we assumed a bias of poem-processing towards the left and a bias for song-processing on the right hemisphere. Furthermore, PSA values of poems and songs were expected to explain variance in left- vs. right-temporal brain areas, while continuous liking ratings obtained in the scanner should modulate activity in the reward network. Overall, poem processing compared to song processing relied on left temporal regions, including the superior temporal gyrus, whereas song processing compared to poem processing recruited more right temporal areas, including Heschl's gyrus and the superior temporal gyrus. PSA values co-varied with activation in bilateral temporal regions for poems, and in right-dominant fronto-temporal regions for songs. Continuous liking ratings were correlated with activity in the default mode network for both poems and songs. The pattern of results suggests that the neural processing of poems and their musical settings is based on their melodic properties, supported by bilateral temporal auditory areas and an additional right fronto-temporal network known to be implicated in the processing of melodies in songs. These findings take a middle ground in providing evidence for specific processing circuits for speech and music in the left and right hemisphere, but simultaneously for shared processing of melodic aspects of both poems and their musical settings in the right temporal cortex. Thus, we demonstrate the neurobiological plausibility of assuming the importance of melodic properties in spoken and sung aesthetic language alike, along with the involvement of the default mode network in the aesthetic appreciation of these properties.

1. Introduction

Music and language as distinct human cognitive abilities have a long tradition of being compared to one another (e.g., Besson and Schön 2001; Patel 2010; Steele 1775; Winn 1984). Theoretical discussions emerged in Greek antiquity, if not earlier (Winn, 1984). Recent research has highlighted shared principles, but also acknowledged processing dissimilarities and functional differences of the two domains (Mehr et al., 2019; Sammler, 2020; Zatorre and Baum, 2012). This paper contributes to this long-standing line of research by comparing the neural processing of spoken poems and their musical settings.

Several major attempts have been made to unravel the neurobiological foundations of music and language (Besson and Schön, 2001;

Brattico and Pearce, 2013; Brown et al., 2006; Janata, 2015; Koelsch et al., 2002; Koelsch and Siebel, 2005; Patel, 2010; Peretz et al., 2015; Sammler, 2020; Sammler and Elmer, 2020; Schön et al., 2005; Zatorre, 2013). In this research, a seminal finding in earlier brain imaging work (Bever and Chiarello, 1974), according to which language and music are predominantly processed in the left and right hemispheres has given way to a more fine-grained and detailed functional differentiation focusing on different sensory modalities as well as on different types of linguistic and musical stimuli.

Regarding the time course of processing, speech and music share identical early processing routines along the sensorineural architecture from the cochlea to the primary auditory cortex in the temporal lobes of the human brain. Later processing steps show increased functional spe-

cializations, with the left hemisphere usually supporting shorter units, and the right usually supporting longer units of processing (Clunies-Ross et al., 2015; Luo and Poeppel, 2012; Poeppel, 2003). This distinction is in line with the oscillatory dynamics of the brain: higher frequencies (in the gamma-range, i.e., 30–80 Hz) are biased towards the left and lower frequencies (in the delta-range, i.e., 0.3–4 Hz) toward the right hemisphere. Moreover, temporal processing is primarily supported by the left and spectral processing by the right hemisphere, relating to left-biased smaller integration windows and right-biased larger integration windows (Boemio et al., 2005; Britton et al., 2009; Hall et al., 2002; Obleser et al., 2008; Patterson et al., 2002; Schönwiesner et al., 2005; Zatorre and Belin, 2001).

The temporal-spectral dissociation is in line with the finding that the perception of lower-level prosodic information relies on a right-hemispheric network that is also recruited during music perception, including song (Gandour et al., 2004; Kreitewolf et al., 2014; Merrill et al., 2012; Sammler et al., 2015; Tong et al., 2005). This network encompasses primary and secondary auditory areas along the temporal gyrus, including Heschl's gyrus and planum temporale, and frontal areas, most prominently the inferior frontal gyrus. Importantly, the left-hemispheric counterpart is preferentially recruited when the focus is on linguistic rather than musical (Merrill et al., 2012) and/or emotional processing (Kreitewolf et al., 2014).

Given the relevance of prosodic contours in speech and melody alike, several previous studies have already considered melodic properties, i.e., ordered sequences of discretely perceived pitches, as a relevant functional link between language and music (Alkon, 1959; Chow and Brown, 2018; Menninghaus et al., 2018; Patel, 2005). Although pitch is more discrete in music than in speech (Burns, 1999; Sievers, 1912; Zatorre and Baum, 2012), pitch sequences in music and speech have been described as similarly melodious (Scharinger et al., 2022), and, on a higher level of processing, seem to be similarly describable as "melodic gestalts". Melodic affinities between speech and music have also been discussed in musicology (Christiansen, 2004; Wingfield, 1992), and linguistic theory has highlighted structural prosodic properties that can be used to describe aesthetic features in music and language. Roman Jakobson, one of the most influential structural linguists of the 20th century, identified parallelism as a central principle of poetic language (Jakobson, 1960). In a nutshell, Jakobson postulates that parallelism in poetry entails all kinds of phonemic and suprasegmental sound repetitions (e.g., rhyme or the appearance of identical phonemes in identical positions within a line of verse), as well as grammatical, morphosyntactic, and semantic parallelisms. Jakobson's assumptions regarding the crucial role of parallelism in poetry have been widely acknowledged, specifically with regard to meter and rhyme. Recent research has provided evidence that the topical comparison of poems and songs – which are designated by the same word in some languages (e.g., *mélos* in Greek, *carmen* in Latin) – can be operationalized by recourse to sound-based recurrence structures, most notably pitch autocorrelations, in the prosodic (melodic) trajectories of recited poems (Menninghaus et al., 2018). The authors have shown that melodic properties of spoken poems are based on the recurrence structure of syllable-based discrete pitches. These melodic properties were objectively quantified using an autocorrelation approach and subjectively rated on aesthetically evaluative scales, including melodiousness ratings. Pitch-autocorrelations, particularly across lags corresponding to the stanza, predicted melodiousness ratings as well as the likelihood that the poems had been selected by professional composers as the textual basis of a musical song. In light of this finding, measures of pitch autocorrelations are particularly relevant, as they seem to correspond to the cognitive processes that underly the perception of melodic contours.

Previous brain imaging studies on the neural processing of melodies can be broadly divided into two types. On the one hand, electrophysiological approaches to melody processing have focused on effects of melodic and harmonic expectancy violations (Calma-Roddin and Drury, 2020; Gordon et al., 2010; Koelsch, 2009; Koelsch et al., 2007; Koelsch et al., 2019; Omigie et al., 2019; Vuust et al., 2012). Next to identifying components of event-related potentials (ERPs) that index the resulting prediction error in these expectancy-based accounts of music processing (Huron, 2006; Jay, 1990; Koelsch et al., 2019; Rohrmeier and Koelsch, 2012; Schmuckler, 1989), some studies have also assigned the violation effect to specific brain areas, e.g., the right auditory cortex (Lappe et al., 2013; Rohrmeier and Koelsch, 2012).

Several fMRI-studies, on the other hand, have contrasted familiar with unfamiliar melodies, expected with unexpected melody endings, and sung with spoken versions of the same sentence material (Angulo-Perkins and Concha, 2019; Cheung et al., 2018; Farbood et al., 2015; Garcea et al., 2017; Koelsch et al., 2002; Kunert et al., 2015; Leipold et al., 2019; Levitin and Menon, 2003; Nan and Friederici, 2013; Steinbeis and Koelsch, 2008; Yu et al., 2017). These studies converge in stipulating shared prosodic (intonation-based) processing networks in the right hemisphere, a specialization for spoken (and not sung) input in the left hemisphere and a differentiation for linguistic and musical syntax in the left and right inferior frontal gyrus. The processing hierarchy of speech and music is comparable in that basic pitch properties of speech sounds or musical sounds are supported by the primary auditory cortex, especially the right Heschl's gyrus in case of musical pitch. In contrast, higher-order processing, e.g., combinations of sound or tone sequences, is subserved by the inferior frontal gyrus and parts of the (pre)motor system.

Up to now, however, no previous brain imaging study has compared speech and music processing on the basis of quantifying melodic properties by means of pitch autocorrelations, as suggested by Menninghaus et al. (2018). Furthermore, very few studies have used ecologically valid materials, e.g., complete songs or short recited texts or poems. Moreover, Menninghaus et al. (2018) have specifically focused on the predictive power of autocorrelations for subjective aesthetic evaluations, making the measure well-suited for a novel attempt to compare the neural bases of the perception of spoken language and music.

Extending this line of research, the present study is the first to directly compare the neural processing of complete, spoken poems, and of musical settings of these very poems (i.e., songs). We sought to determine which brain areas covary with pitch autocorrelations in poems and songs alike, and which brain areas possibly dissociate a musical vs. speech-based melodic processing.

We were also interested in the aesthetically evaluative dimension of speech and music processing and therefore additionally collected continuous liking ratings throughout the exposure to the poems and songs. This way, we intended to identify brain areas that are specifically involved in the processing of objective melodic properties on the one hand, and in subjective aesthetic evaluations of these properties on the other. Because we used ecologically valid stimuli, we were optimistic that a basic liking rating might be feasible even in the alienating environment of a noisy fMRI-scanner. A survey on previous uses of continuous liking ratings (Wagner et al., 2020) reported only negligible effects of such ratings on aesthetic evaluations. Therefore, continuous ratings should generally not have adverse effects on aesthetic processing.

Six months after performing this study, we invited all participants to an additional round of ratings of the same stimuli. The purpose of collecting these ratings was to obtain—beyond the general liking ratings collected in the main study—a more nuanced profile of distinct aesthetic qualities and rewards associated with the respective stimuli. In addition, these data put us in a position to analyze to what extent the noisy and stressful conditions of the scanner study might have had adverse effects on the aesthetically evaluative ratings.

We performed our study under the following hypotheses: (a) pitch-autocorrelations should scale with activity in the right-hemispheric prosodic fronto-temporal processing areas (see Sammler et al. 2015); (b) poems and songs should recruit areas along the superior and middle temporal gyrus in both the left and the right hemisphere; (c) liking ratings should correspond to activity within the reward network, including orbitofrontal cortex, cingulate cortex, the striatum and further subcortical

areas. Overall, we set out to show that a direct comparison of poems and their musical settings allows to further our understanding of the affinities and differences in the prosodic processing of language and music and of the neurobiological bases of (aesthetic) perception/evaluation.

2. Materials and methods

2.1. Stimuli

From an original set of 40 German poems, 20 of which were set to music, we selected 11 that had previously been used for a concert study (Scharinger et al., 2022). These 11 poems were written between 1805 and 1883, and their corresponding 11 musical settings, for male voice and piano, created by Romantic composers between 1817 and 1896. Because the composers of the 11 musical settings did not alter the texts of the underlying poems, differences in prosody and neural processing were not conflated by textual differences, but were likely to be exclusively reflective of the differences in prosody between the two conditions, i.e., the original poems and their respective musical settings. An overview of the 11 original poems and the 11 songs can be found in the Supplementary materials (see Table S1).

All spoken poems and their musical settings were recited/sung by a professional male singer (baritone). Acoustic signals were recorded at 48 kHz sampling rate and 24-bit amplitude resolution. The songs, originally composed for male voice and piano, were accompanied by a professional piano player. We recorded only the singing voice; this was achieved through a direction-sensitive microphone focused on the singer, allowing for a separation of voice and piano. When presented in the scanner, the piano was inaudible. As a result, the spoken and the sung acoustic material allowed for an unconflated comparison.

To these two conditions (original poem and song), we added two further conditions: a modified poem condition and a high-level baseline pitch-contour control condition. The modified poem condition was adopted from a previous study (Menninghaus et al., 2018) in which the same original poems were modified such that prototypical features of poetic prosody (e.g., rhyme and sustained binary meter) were experimentally removed, leading both to a reduced strength of pitch autocorrelations and lower melodiousness ratings. Rhymes were eliminated by replacing one of the (usually two) rhyming words with a synonymous content word. In addition, the original regular meter was disrupted (1) by adding or removing function words, (2) by replacing content words with synonyms differing in number of syllables, or (3) by only changing word order (see Table S2 in the Supplementary materials for illustration). Thus, the modified poem condition consisted of the 11 modified versions (with neither rhyme nor meter) of the original poems (see Menninghaus and Wallot 2021, for poems with similar modifications).

The pitch-contour control condition consisted of sine tones. Durations and fundamental frequencies of these sine tones were derived from the pitch contour of the original spoken poems. Pitch contours were extracted in PRAAT (Boersma and Weenink, 2019) and mean pitch values calculated for each spoken syllable (cf. Fig. 1A). These mean pitch values, transposed upwards by two octaves for better audibility, were mapped onto a semi-tone scale. Resulting frequency values served as basis for synthesizing sine tones the durations of which were obtained through discretizing the syllable durations. In this process, the raw duration values were mapped onto the nearest fractions of 1/16 s (i.e., 62.5 ms; approximating 1/16-notes in musical notation when performed at *adagio* tempo [60 beats per minute], as was the case for most songs). Note that this procedure is identical to the one of preprocessing the poems for autocorrelation analyses.

The total of 44 stimuli (11 original poems, 11 songs, 11 modified poems, 11 sine-tone sequences [controls]) was pseudo-randomly distributed over 4 runs, separately for each participant, to avoid order effects. Care was taken that the total stimulation durations in each run were comparable and could be covered by the same number of func-

Table 1
Averages of durations and scans for each stimulus condition.

Stimulus Condition	Duration [sec]	Scans [Number]
Poems	51.3	26
Musical settings	135	68
Modified poems	46.7	23
Pitch controls	52.6	27

tional scans. Table 1 shows the stimulus characteristics based on duration and number of scans.

2.2. Participants

42 participants were recruited for the study. After initial screening, one participant had to be excluded, leaving a total of 41 participants (mean age: 25 ± 4 years, range 18–32 years, 22 females, 19 males). All participants were native speakers of German, had normal or normal-to-corrected vision and no history of neurological or hearing impairments. They reached at least 90% on the abbreviated Edinburgh handedness scale (Oldfield, 1971). About half of the participants had learnt to play a musical instrument. At the same time, none of them was a professional musician.

27 of these 41 participants were successfully recruited for the additional round of ratings after the completion of the fMRI study (mean age: 25 ± 3.6 years, range 20–32, 15 females, 12 males). The time between the fMRI study and the offline rating was 6 months on average. All participants gave their written informed consent to both the brain imaging and the post-hoc rating study that had been approved by the Ethics Committee of the Medical Department of the Goethe-University in Frankfurt, Germany. The ethical approval attested that the studies were in accordance with the declarations of Helsinki.

2.3. Scanning parameters

The fMRI study was carried out at the Frankfurt-based Brain Imaging Center (BIC) on a Siemens 3T scanner (Magnetom Trio Syngo). The scanning protocol included a localizer scan at the start of the measurement, a field map for field inhomogeneity corrections, echo-planar images (EPIs) for registering functional data, and a T1-weighted anatomical scan for improved co-registration and normalization at the end of the measurement. The parameters for the scans are as follows: field map: 39 slices, 2 mm thickness, gap: 1 mm, FoV: 192 mm, TR: 700 ms, TE(1): 4.89 ms, TE(2): 7.35 ms, flip angle: 60°; EPIs: 39 slices, 2 mm thickness, gap: 1 mm, FoV: 192 mm, TR: 2500 ms, TE: 30 ms, flip angle: 90°; T1-weighted scan: MPRAGE, 176 slices, 1 mm thickness, gap: 1 mm, FoV: 256 mm, TR: 2000 ms, TE: 2.48 ms, flip angle: 8°.

The presentation of 11 stimuli in each run was covered by 330 EPIs, thus amounting to about 14 min. Together with scanning preparation, field map and anatomical scans, the total scanning time was about 60 min per participant.

2.4. Registration of liking ratings and of further aesthetically evaluative ratings

Liking ratings were collected through an MR-compatible response device, a Biopac TSD160A differential pressure transducer (Biopac Systems Inc., Goleta, CA, USA). The transducer consists of two pressure sensitive buttons that were connected to the participants' left and right index fingers. Participants were instructed to indicate differential degrees of liking with the left and right index fingers. We split the scale into a left button press (to be used for indicating stronger degrees of liking) and a right button press (to be used for indicating weaker degrees of liking, cf. Fig. 1C). Note that this splitting of the scale was necessary since there was no haptic feedback as to the scales' ends. For a single

scale, it would have been impossible to determine a mid-point. For this reason, we decided to use a neutral midpoint (no button press left or right) and determine the scales' poles on the basis of this midpoint. The assignment of stronger and weaker degrees of liking to button presses with the right or left index fingers was counter-balanced across participants. The transducer operated with pressure between –2.5 and +2.5 cm H_2O. Values were recorded with a sampling frequency of 1000 Hz.

In the additional data-collection six months after the scanner study, participants were asked to rate the 11 original poems, 11 modified poems, and 11 songs on 19 seven-point, i.e., non-continuous scales. The scales were selected on the basis of previous online surveys and covered a broad range of emotional and aesthetically evaluative dimensions which are applicable to both poems and songs. For instance, one scale asked for melodiousness with 1 corresponding to "not at all melodious" to 7 corresponding to "very melodious". The original German scales, along with their translations, can be found in the Supplementary materials (Table S3).

2.5. Delivery of auditory stimuli

During the main experiment, auditory stimuli (11 original poems, their 11 modified counterparts, 11 songs and 11 controls) were delivered over MR-compatible noise-cancelling headphones (OptoAcoustics OptoActive II). These headphones reduced the EPI-gradient noise by about 60 dB, allowing for a satisfying signal-to-noise ratio. Signal amplitudes were adjusted for each participant to be clearly audible over the EPI-gradient noise. In this process, care was taken not to reach levels that would have caused discomfort.

2.6. Procedure

Participants of the brain imaging experiment were screened, provided written informed consent, and were then placed in the scanner in supine position. They were instructed to listen to the auditory stimuli during the main experiment and to indicate their liking or disliking using the buttons of the differential pressure transducer. We asked them to provide continuous liking ratings (for all stimuli) that should reflect spontaneous liking judgments at each point in time during listening. For the listening periods, participants were instructed to either close their eyes or to focus on a fixation cross projected in the center of participants' visual field inside the scanner. After scanning, participants completed a demographic questionnaire.

The 27 participants who accepted our invitation to a follow-up study were asked to give 19 ratings on 7-point scales for each of the test stimuli of the brain imaging experiment (i.e., 11 original poems, 11 modified poems and 11 songs) after hearing them.

For all scales, participants were asked to provide ratings ranging from 1 (not at all) to 7 (highly). The scales appeared in random order, using the experimental stimulation software Presentation (Neurobehavioral Systems, Inc., Berkeley, CA, USA). Similarly, auditory stimuli, presented over high-quality headphones (beyerdynamic Inc., Heilbronn, Germany), occurred in random order.

2.7. Preprocessing

2.7.1. Imaging data

Preprocessing of fMRI data was carried out in SPM 12 (Wellcome Trust Centre for Neuroimaging, www.fil.ion.ucl.ac.uk/spm) within Matlab (Version 9.4, Mathworks, Nattick, MA, USA, 2018), using custom-made scripts.

Preprocessing steps followed a standard protocol (Ashburner and Good, 2003) and involved the application of slice time correction with the middle slice serving as reference in the correction process. Based on the slice-time corrected volumes, realignment and unwarping procedures were applied in order to minimize movement-induced variance. Field maps were used as basis for voxel-displacement maps (VDMs) in

order to correct field inhomogeneities. T1-weighted anatomical scans were co-registered to the mean functional EPI image of each participant. Subsequent segmentation used the SPM12 standard tissue probability maps involving light bias regularization. EPIs were resampled to a voxel size of $3 \times 3 \times 3$ mm. All normalized images were smoothed with an isotropic 8-mm FWHM kernel in order to minimize effects of anatomical differences.

2.7.2. Liking data

Raw liking data consisted of pressure values bound between –2.5 and +2.5 cm H_2O. In a first step, raw values for each stimulus were centered and then added to a composite measure. The addition depended on whether the positive direction of the liking scale involved a left or right button press of the participant. For instance, if the right button press indicated increased liking, the pressure values of the left button were subtracted from the pressure values of the right button. Reversely, if the left button press indicated increased liking, the pressure values of the right button were subtracted from the pressure values of the left button. In order to allow for a mapping onto a 7-point space, resulting composite pressure values were shifted into the positive number space by adding the absolute of the minimum pressure value (per individual participant and stimulus). The mapping onto a 7-point item space was achieved by the following formula:

$$R_L = \left(\frac{6}{\max(R_P)} \times R_p \right) + 1$$

with R_L denoting the rating in (7 point) Likert-space and R_P denoting the shifted composite pressure value of the ratings in (cm H_2O) pressure space.

2.8. Statistical analyses

Statistical inferences for fMRI data were made on the basis of Gaussian field theory (Friston, 2004) and followed a two-level approach implemented in SPM12. The first level analyses were performed on all conditions per participant. The design matrix consisted of three experimental conditions (original poems, modified poems, songs) and a high-level control condition (sine tone pitch tracks). Scans collected during silent time periods between the stimuli or at the beginning or end of each run (i.e., no stimulation) were modelled as events of no interest. Six movement parameters were included as further regressors of no interest. Additionally, the design matrix included the mean of each run (i.e., experimental run) as an additional regressor. An explicit mask was applied in the first level analysis that had been calculated from an average of EPI volumes across all participants. The average EPI volume was subsequently binarized with FSL (Jenkinson et al., 2012). The BOLD (blood oxygen level dependent) signal was approximated by a canonical hemodynamic response function. A high-pass filter with a cut-off of 128 s was used to account for slow drifts.

Three different first-level analyses were carried out. For calculating main effects, the design matrix consisted of the aforementioned conditions. In order to additionally model the correlations of liking ratings and pitch autocorrelation values with the BOLD signal, two separate first level analyses were calculated that included the parametric modulators *liking ratings* and *pitch autocorrelations (cf. Fig 1B)*.

The parametric modulator *liking ratings* was obtained by averaging the continuous liking ratings from the composite measure in 2500 ms bins (corresponding to the duration of a single EPI, cf. Fig. 1C).

The parametric modulator *pitch autocorrelations* was determined as follows: First, digitized audio data of original poem, modified poems and songs were subjected to fundamental (f0) frequency extraction using TANDEM-STRAIGHT (Kawahara and Morise, 2011) within Matlab (Version 9.4, Mathworks, Nattick, MA, USA, 2018). Fundamental frequency (f0) was determined between 32 and 650 Hz, covering the voice range of a baritone singer. The extraction of f0 was done with an internal sampling frequency of 6000 Hz. Subsequently, autocorrelations
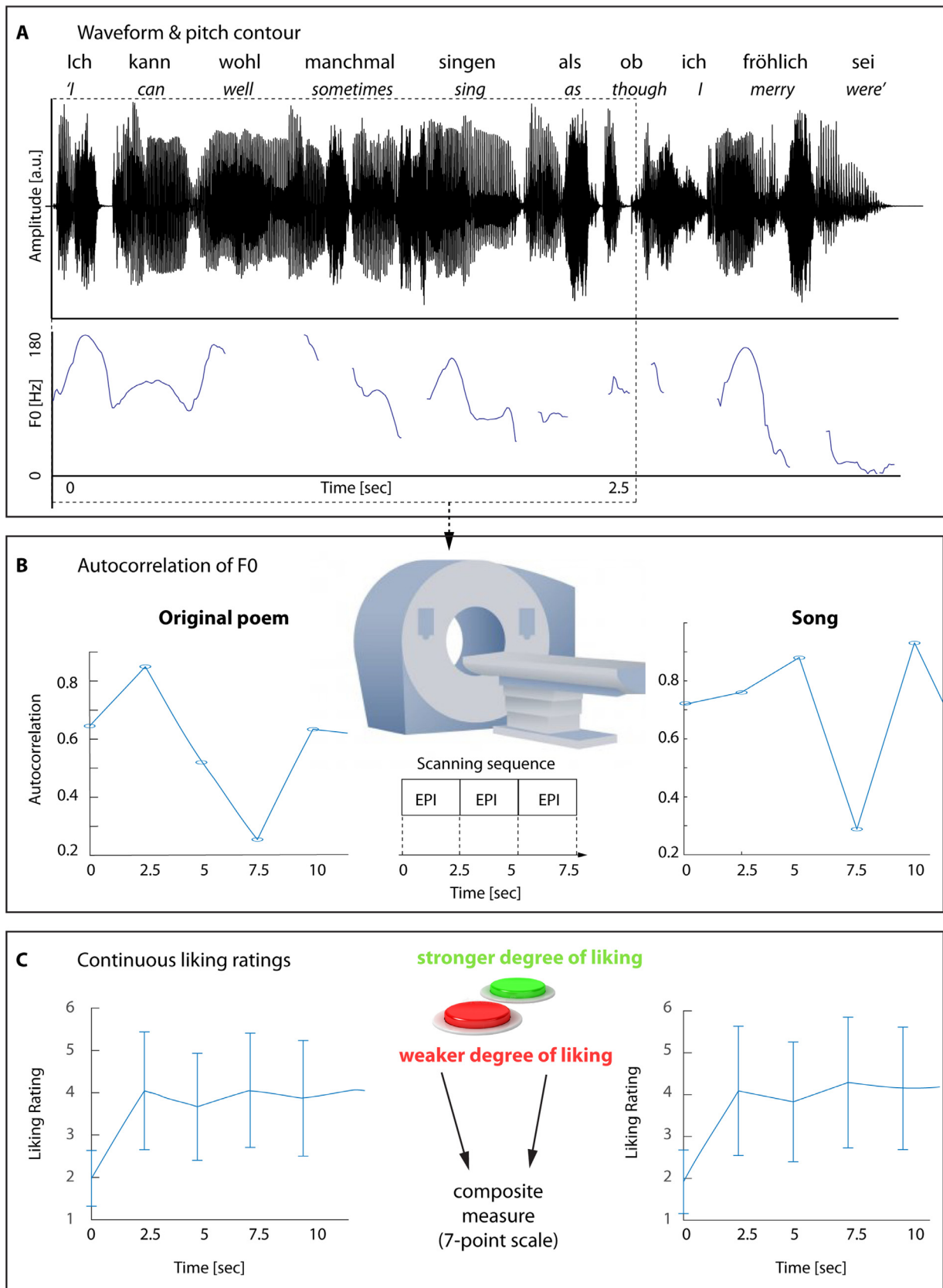
Fig. 1. Illustration of experimental material. (A) F0 contours were automatically extracted from recorded versions of original poems, modified poems and songs and submitted to autocorrelation analyses. (B) Autocorrelation analyses were carried out in consecutive, non-overlapping windows with a size corresponding to the duration of echo-planar images (EPIs). Note that the mean value of the first window was aligned with time 0 (or scan 1). (C) Illustration of continuous liking ratings after their transformation into a composite 7-point scale. As for autocorrelation values, mean values in 2500 ms bins were calculated, with the first window being aligned to time 0 (or scan 1). Whiskers indicate the standard error of the mean. All examples stem from the first 10 s of the original poem "Wehmut" by Joseph von Eichendorff and its corresponding song by Robert Schumann.

of f0 were calculated within bins of 2500 ms (corresponding to EPI-durations), with lags corresponding to 1/6000 Hz, i.e., 0.16 ms. Autocorrelation calculation within these bins used all lags up to the lag corresponding to 90% of bin duration (2250 ms). In order to only consider meaningful autocorrelation values, a bootstrapping procedure with 1000 repetitions based on the reshuffling of lags and an alpha-level of 0.001 was applied. The significance level was chosen (a) to "match" the statistical error size of the uncorrected significance levels of the fMRI-analyses (< 0.001), and (b) based on the number of repetitions for the bootstrapping. The quotient of meaningful to overall values corresponds to the proportion of significant autocorrelations (PSA). A detailed illustration of how specific parts of an exemplary song correspond to specific PSA values is provided in the Supplementary Materials (Fig. S1).

For each of the 2500 ms-bins, corresponding to individual EPIs, the PSA value was used as parametric modulator (cf. Fig 1B). Note that the time bins with 2500 ms duration correspond to the average duration of a line in the spoken poetry stimuli of this study. The PSA value for these time bins therefore describes the local recurrence pattern of pitch within a line.

The parametric modulators were also correlated with the post-hoc liking ratings and autocorrelations based on syllable pitch in order to allow for better comparison with the study by Menninghaus et al. (2018). Autocorrelations based on syllable pitch were calculated as described in Menninghaus et al. (2018). The algorithm used to arrive at stimulus-specific PSA values was identical to the one described above, except that the initial material consisted of time-bins corresponding to individual syllables of the acoustic material. To this end, all stimuli underwent a two-pass syllabic annotation process, with an initial semi-automatic syllable edge identification and a following manual inspection of syllable boundaries.

After model estimation on the first level of the fMRI analyses, main contrasts were specified comparing each experimental condition (original poem, modified poem, song) against the high-level baseline (pitch-derived sine tones). Furthermore, conditions were also directly compared with each other (original poem vs. modified poem, original poem vs. song, modified poem vs. song). All resulting contrasts were included in a second level analysis.

At the second level, all contrasts (for main effects) or beta values (for parametric modulator analyses) were compared against zero using one-sample *t*-tests. For the main contrasts, thresholds with family-wise error rate (FWE) corrected *p*-values (< 0.05) were applied. For main contrast comparisons, we used an uncorrected threshold at $p < 0.001$; for all remaining contrasts, an uncorrected threshold of $p < 0.005$ combined with a cluster extent of 17 voxels was used. This corresponds to a whole-brain alpha of $p < 0.05$, as determined from a Monte Carlo simulation describe in Slotnick et al. (2003) with an overall smoothing of 9 mm for all comparisons. Overall smoothing was estimated on the basis of the final statistical map within SPM12.

Anatomical locations obtained from second-level analyses were labeled according to the Automated Anatomical Labeling Atlas (AAL; Tzourio-Mazoyer et al., 2002). Planum temporale was labelled according to the Westbury Atlas (Westbury et al., 1999).

For illustration and comparison of individual conditions within specific regions obtained from second level analyses, beta values were extracted and averaged within regions of interest (ROIs). ROIs were defined as spheres with 5 mm radii around centers determined by mean MNI-values for all peak coordinates obtained from the whole-brain analyses. Beta values for these ROIs were derived from general linear model estimations including the same factors of no interest as described above. For all main contrasts, beta values were transformed into percentage signal change values according to the following formula, otherwise, (mean) beta values of the respective ROIs are reported:

$$PSC = \left(\frac{\beta_{Test} - \beta_{Control}}{mean(\beta_{Session})} \right) \times 100$$

with PSC = percentage signal change.

All subsequent and additional statistical analyses were carried out in R (Version 4.03, *R Foundation for Statistical Computing, Vienna, Austria*).

In order to quantify laterality effects of our whole-brain analyses, we additionally computed laterality indices, following the suggestions by Matsuo et al. (2012). Average laterality indices were calculated on the basis of SPM-t files obtained from the SPM analyses. We chose this basis to be able to consistently report laterality indices for both the contrast analysis and the parametric modulation analyses.

3. Results

3.1. Behavioral data

By-item means of continuous liking ratings were significantly correlated with by-item means of post-hoc liking ratings ($r = 0.69$, $t = 5.34$, df=31, $p < 0.001$), thereby supporting their validity as continuous measure. Continuous liking ratings differed between the three test conditions, as assessed by the main effect of *stimulus type* in a one-way ANOVA in R (F(2,30) = 40.98, $p < 0.001$). Planned comparisons using the multcomp package (Bretz et al., 2011) revealed that original poems received slightly higher ratings than modified poems ($t = 2.07$, $p = 0.1$, all *p*-values Bonferroni-corrected, see Fig. 2). Moreover, the original poems were also rated higher than their musical settings ($t = 8.67$, $p < 0.001$); the same holds for the modified poems ($t = 6.59$, $p < 0.001$).

At a first glance, the low ratings for the songs may seem surprising. Note, however, that the songs presented all belong to a very particular and relatively infrequent category of songs: namely, previously published high art-poems that have been set to music. This cross-over of high-art poems and songs is a unique art form much less known than poems (to which everyone has some exposure during school education) and more prototypical songs, be they popular or high art songs. Therefore, the relatively low liking ratings for the poem-based songs presented in our study most likely reflects an effect of the well-established principle of familiarity-driven aesthetic liking (Reber et al., 2004) and the mere-exposure effect (Zajonc, 1968). Furthermore, participants explicitly indicated that they preferred recited poems to poems set to music ($t = 5.45$, df=40, $p < 0.001$) and that they also preferred instrumental classical music to poems set to music ($t = 5.92$, df=40, $p < 0.001$). The familiarity argument is also corroborated by additional demographic data suggesting that 59% of the participants played an instrument for at least one year, while only 19% of the participants indicated that they were singing.

In order to reduce the dimensionality of the post-hoc ratings and in order to have a more complex perceptual correlate of melodiousness, we applied a factor analysis using the R-package "psych" (Revelle, 2020). The factor analysis was further motivated by the multi-faceted properties of the melodiousness percept: Previous studies suggest that melodiousness may not be represented by a single scale (Scharinger et al., 2022). Therefore, the factor analysis provided here is also important for enhancing comparability. The optimal number of factors for the 19 scales (including liking ratings) was calculated using the parallel analysis technique (Horn, 1965) that compares the observed eigenvalues of a correlation matrix with those from random data. This technique yielded an optimum of three factors. Subsequently, using this number, a principal factor analysis with oblique rotation and a factoring based on minimizing residuals was conducted (factoring reliability = 0.88). In our choice of factor names, we attempted to use nouns from the available adjectives that are representative of all subsumed scales.

We obtained the following factors (listing the rating scales with factor loadings greater than 0.3): Factor 1 (henceforth labeled "Beauty") with the scales liking, invigorating, calming, touching, moving, stirring, harmonious, intensive, lively, poetic, beautiful, rhythmic and sad; Factor 2 (henceforth labeled "Joy") with the scales invigorating, calming, joyful, cheerful, lively, negative [with negative loading], positive and sad [with negative loading]; Factor 3 (henceforth labeled "Melody") with the scales liking, invigorating, stirring, harmonious, in-
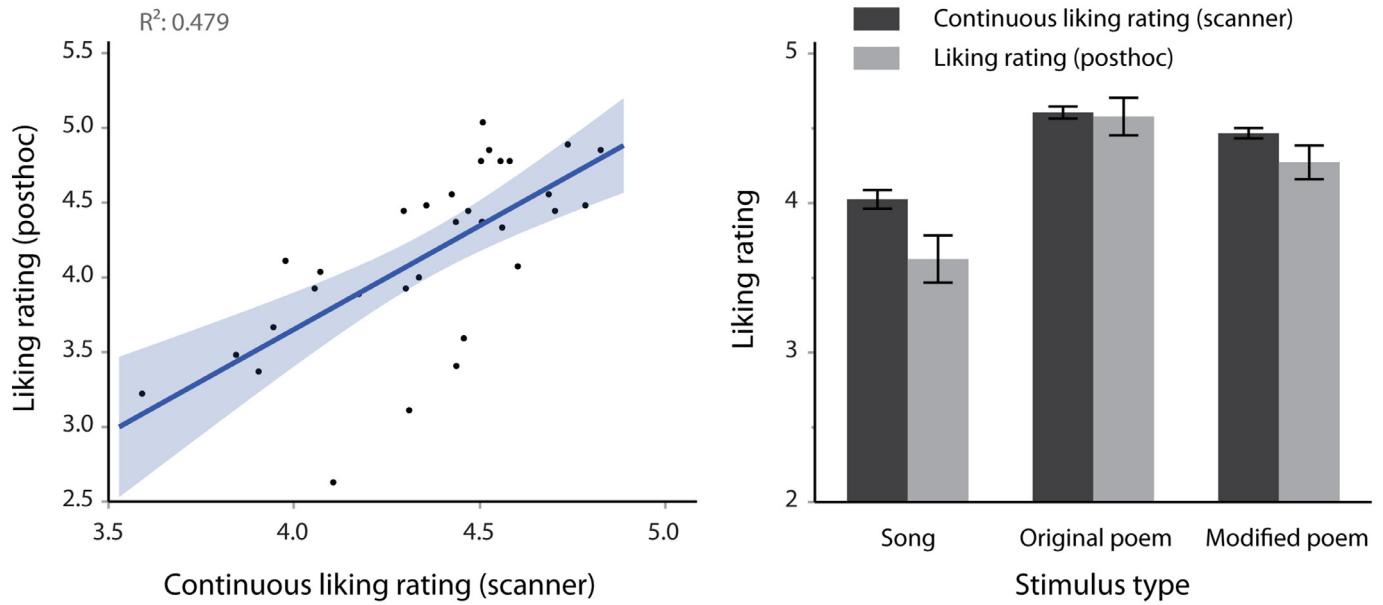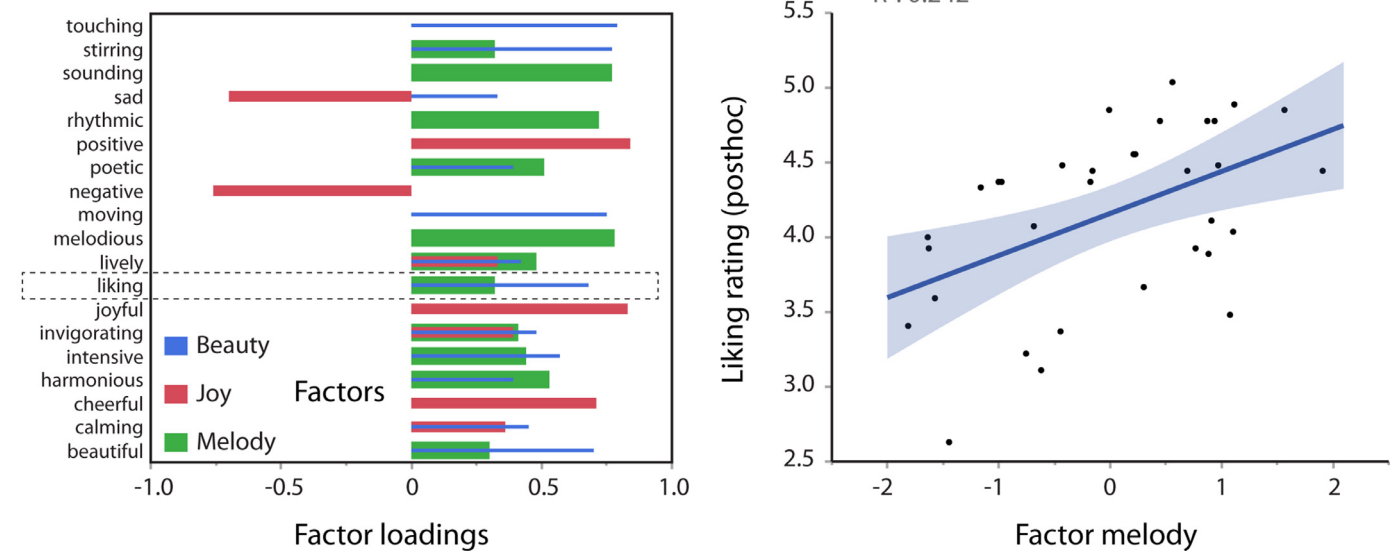
Liking ratings

Fig. 2. Comparison of data collected in the MRT-scanner (continuous liking ratings) with data collected post-hoc (liking ratings). Left top: correlations of post-hoc liking ratings (pooled across participants) and continuous liking ratings (pooled across time bins and all participants for which we had both fMRI and post-hoc data, N=27). Each dot therefore represents one stimulus; right top: Rating means as a function of stimulus type. Whiskers indicate the standard error of the mean. Right bottom: Overview of the factor loadings on the 19 scales of the post-hoc study. Left bottom: Correlation of mean liking ratings with the factor "melody".

tensive, sounding, lively, melodious, poetic, rhythmic and beautiful (see Fig. 2 and Supplementary Material Table S2). For all factors, we calculated scores by multiplying the z-transformed, summed scales with the respective positive or negative factor loadings (> 0.3). Importantly, the factor "melody" included the liking scale and was highly correlated with liking ratings ($r = 0.71$, $t = 29.65$, df=889, $p < 0.001$; aggregated data $r = 0.49$, $t = 3.14$, df=31, $p < 0.01$; see Fig. 2). With this correlation, we are able to relate some of the variance explained by the parametric modulator "liking" to melodiousness aspects of the stimuli. Effectively, this enabled us to consider "liking" as a proxy for melodiousness ratings that we could not collect in the scanner.

3.2. Pitch autocorrelations

Autocorrelations based on automatically extracted F0 (PSA) in time bins corresponding to EPI-epochs were significantly correlated to autocorrelations based on mean syllable pitch ($r = 0.34$, $t = 2.02$, $p < 0.05$, Fig. 3). This enables us to directly compare the F0-based values to previous work on autocorrelations. Further, F0-based autocorrelations differed between the three test conditions, as seen in the main effect *stimulus type* (F(2,30)=11.78, $p < 0.001$, Fig. 3). Planned comparisons showed higher PSA-values for songs than for original ($t = 3.43$, $p < 0.01$) and modified poems ($t = 4.69$, $p < 0.001$). Original poems had nominally
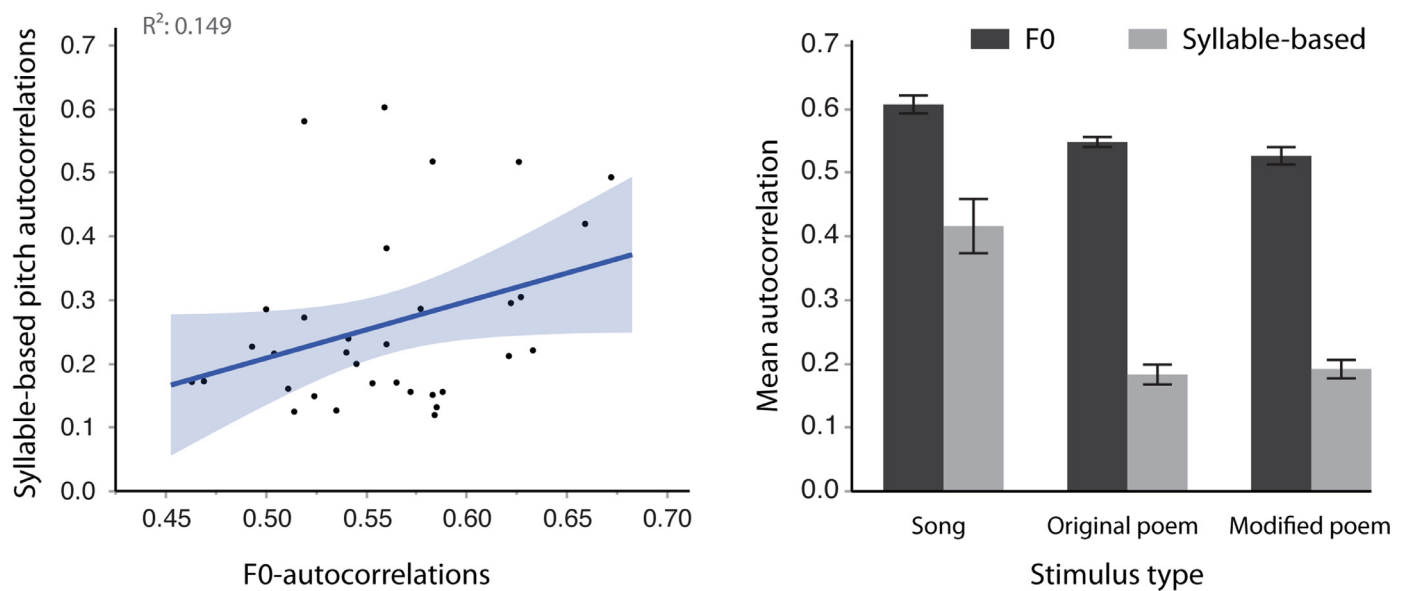
F0-Autocorrelations

Fig. 3. Comparison of autocorrelations (proportion of significant autocorrelation, PSA) between F0-based and syllable-based analyses. Left: Correlation between F0- and syllable-based autocorrelations. Right: Means of PSA-values (EPI- vs. syllable-based) as a function of stimulus type. Whiskers indicate standard errors of the mean.

higher PSA-values than modified poems; the difference is, however, not significant ($p > 0.1$).

3.3. Imaging data

3.3.1. Main contrasts

Compared to baseline, songs elicited activation in the bilateral superior temporal gyrus, extending from posterior parts through planum temporale and parts of Heschl's gyrus to anterior parts. Clusters in the left and right STG were almost symmetrical, with marginally larger clusters in the left. Original poems showed similar bilateral clusters along the superior temporal gyrus, including anterior and posterior parts and planum temporale. Importantly, the larger cluster in the left hemisphere comprised more area within Heschl's gyrus and additionally involved parts of superior temporal sulcus and middle temporal gyrus (MTG). Modified poems mainly elicited activation in a left temporal cluster, including the posterior STG and planum temporale. Only a small cluster in the right anterior STG additionally was recruited during listening to modified poems (cf. Table 2 and Fig. 4).

3.3.2. Comparisons

When comparing the test conditions directly, songs showed more activity in a right temporal cluster, involving Heschl's gyrus, planum temporale as well as posterior parts of STG. By contrast, original poems showed more activity in a left temporal cluster, including posterior STG and STS and parts of MTG (cf. Table 2 and Fig. 4). The comparison of original and modified poems led to a contrast image that did not survive our p-value and cluster-extent threshold. When selecting a more lenient threshold ($p < 0.01$, cluster-extent 17 voxels), original poems showed more activity than modified poems in a bilateral posterior STG, with the left cluster additionally including planum temporale and Heschl's Gyrus. The reverse contrast did not show any cluster activation.

3.3.3. Parametric modulations

F0 autocorrelations: F0-based PSA values of songs positively correlated with activity in several clusters. The two main clusters were found in the posterior STG and IFG (including pars opercularis) in the right

hemisphere. The temporal cluster also extended into the MTG and supramarginal gyrus, while the frontal cluster included the pars triangularis and the frontal operculum. Smaller cluster comprised the right precuneus and left precentral gyrus (cf. Table 3 and Fig. 5).

F0-based PSA values of original poems positively correlated with activity in the left posterior STG (including parts of planum temporale and Heschl's gyrus), right pSTG (also including parts of planum temporale and Heschl's gyrus), as well as the left lateral occipital cortex (including the occipital fusiform area, cf. Table 3 and Fig. 5).

Continuous liking ratings: Continuous liking ratings of the original poems covaried with activation in a larger network, some parts of which belong to the reward network. In particular, the parametric modulator positively correlated with activity in the bilateral posterior, mid- and anterior cingulate cortex, precuneus, cerebellum, Thalamus, mid-occipital cortex, supramarginal gyrus, angular gyrus as well as medial frontal and dorsal prefrontal cortex (cf. Table 3 and Fig. 5). Importantly, none of the clusters overlapped with any cluster from the main contrasts or autocorrelation activations. Furthermore, the continuous liking ratings of songs and modified poems did not show any significant correlations.

3.4. Statistics of ROI-based activation

From the main contrast analyses and the parametric modulator analysis involving F0-autocorrelations, we selected six regions of interest (ROIs) that showed the strongest activation in each analysis. This procedure aimed at exploring more details of activation patterns. The first two ROIs are centered around the left and right STG, including parts of the planum temporale, with the MNI coordinates [-/+60 -16 -1]; the third and fourth ROIs with the MNI coordinates [-/+49 -31 11] comprise posterior parts of left and right STG (also including parts of planum temporale and Heschl's gyrus). Finally, the last two ROIs with the MNI coordinates [-/+38 17 12] covered the left and right IFG, including the pars opercularis.

Percentage signal change (PSC) values (from the contrast analyses) and beta values (from the parametric modulation analyses) of all ROIs, conditions, runs and participants were analyzed in linear mixed effects models using the lmerTest package in R (Kuznetsova et al., 2014). *Par-*

Table 2

Peak coordinates, sizes and statistical values of clusters based on the whole-brain analysis (main contrasts). Anatomical descriptions are abbreviated as follows: STG – superior temporal gyrus; PT – planum temporale; HG – Heschl's gyrus; pSTG – posterior STG; aSTG – anterior STG; pSTS – posterior superior temporal sulcus.

Contrast	x	y	z	Cluster	Voxels	cm3	Peak t	Peak z	Description
Songs>Controls	-63	-16	4	1	119	357	8.45	6.36	STG, PT, HG
	60	-16	1	2	95	285	6.96	5.60	STG, PT, HG
	63	-4	-5	2			6.50	5.34	
Original poems>Controls	63	-4	-5	1	108	324	8.89	6.57	STG, PT, HG
	63	-16	1	1			7.48	5.88	
	48	-25	1	1			6.36	5.26	
	-63	-19	4	2	187	561	7.87	6.08	pSTG, pSTS, PT, HG
	-57	-34	1	2			7.85	6.07	
Modified poems>Control	-60	-13	1	1	80	240	6.95	5.60	STG, PT
	-57	-31	1	1			6.13	5.12	pSTG
	63	-1	-8	2	18	54	6.87	5.55	aSTG
Songs>Original poems	54	-7	4	1	131	393	5.06	4.42	STG, HG, PT
	66	-28	16	1	-	-	4.54	4.06	pSTG
					-	-			
Original poems>Songs	-54	-28	1	1	44	132	4.93	4.33	pSTS, STG
	-54	-40	4	1	-	-	4.67	4.15	PT

Table 3

Peak coordinates, sizes and statistical values of clusters based on the whole-brain analysis (parametric modulations). Anatomical descriptions are abbreviated as follows: pSTG – posterior STG, IFG – inferior frontal gyrus; PO – pars opercularis; OC – occipital cortex; PT – planum temporale; HG – Heschl's gyrus; precentral – precentral gyrus; mCingulate – middle cingulate cortex; pCingulate – posterior cingulate cortex; aCingulate – anterior cingulate cortex; MFG – medial frontal gyrus; PFC – prefrontal cortex; SMG – supramarginal gyrus; AG – angular gyrus.

Modulator	x	y	z	Cluster	Voxels	cm3	Peak t	Peak z	Description
F0 PSA Songs	57	-43	13	1	147	441	4.48	4.01	pSTG
	42	14	16	2	122	366	4.01	3.65	IFG, PO
	39	23	7	2			3.87	3.55	
	33	14	13	2			3.07	2.89	
	18	-46	40	3	30	90	3.33	3.11	precuneus
	21	-49	28	3			2.98	2.82	
	9	-55	40	3			2.81	2.67	
	-18	-28	40	4	49	147	3.3	3.09	precental
	-24	-43	46	4			3.11	2.93	
	-27	-28	49	4			2.93	2.77	
F0 PSA Original Poem	-24	-85	-2	1	30	90	4.14	3.75	OC
	-39	-31	13	2	30	90	3.8	3.49	pSTG, PT, HG
	51	-19	7	3	23	69	3.37	3.14	pSTG, PT, HG
Continous Liking Poems Original	12	-31	34	1	615	1845	4.13	3.75	mCingulate, precuneus
	21	-52	37	1			3.86	3.54	pCingulate
	-9	-37	46	1			3.53	3.27	
	6	-4	13	2	153	459	3.92	3.58	Thalamus
	-3	-10	34	2			3.5	3.25	dorsal aCingulate
	9	-10	22	2			2.98	2.81	aCingulate
	9	-46	-8	3	73	219	3.4	3.16	Cerebellum
	-6	-58	-11	3			3.21	3.01	
	39	17	43	4	18	54	3.31	3.09	MFG
	36	41	13	5	76	228	3.19	2.99	dorsal PFC
	39	29	10	5			3.11	2.93	
	39	44	25	5			3.04	2.87	
	-39	-55	28	6	33	99	3.18	2.98	AG
	30	-73	37	7	48	144	3.16	2.97	mid OC
	18	-76	40	7			2.95	2.79	
	-63	-31	25	8	24	72	3.14	2.95	SMG

ticipants were included in the model as the model as random factor, *condition* (original poem, modified poem, song) and *ROI* (3 left- and right fronto-temporal areas as defined above) as well as their interaction were included as fixed effects. *Run* (1–4) was added as covariate. Prior to these analyses, PSC and betas were tested against zero in one-sample *t*-tests. All betas significantly differed from zero (all $ts>3$, $p < 0.001$). However, PSC values in left and right IFG did not differ from zero (all $ts<1$, $p > 0.4$). This is not surprising given that the whole-brain contrast analyses did not reveal any significant clusters in bilateral IGF.

The PSC model revealed a main effect of ROI ($F(5,2893) = 68.71$, $p < 0.001$), with left STG showing the strongest activation. *ROI* interacted with *condition* ($F(10,2893) = 2.05$, $p < 0.05$), suggesting differences between conditions depended on brain regions. A decomposition of the interaction revealed a condition effect in left STG ($F(2,448) = 3.56$, $p < 0.05$), with original poems showing significantly
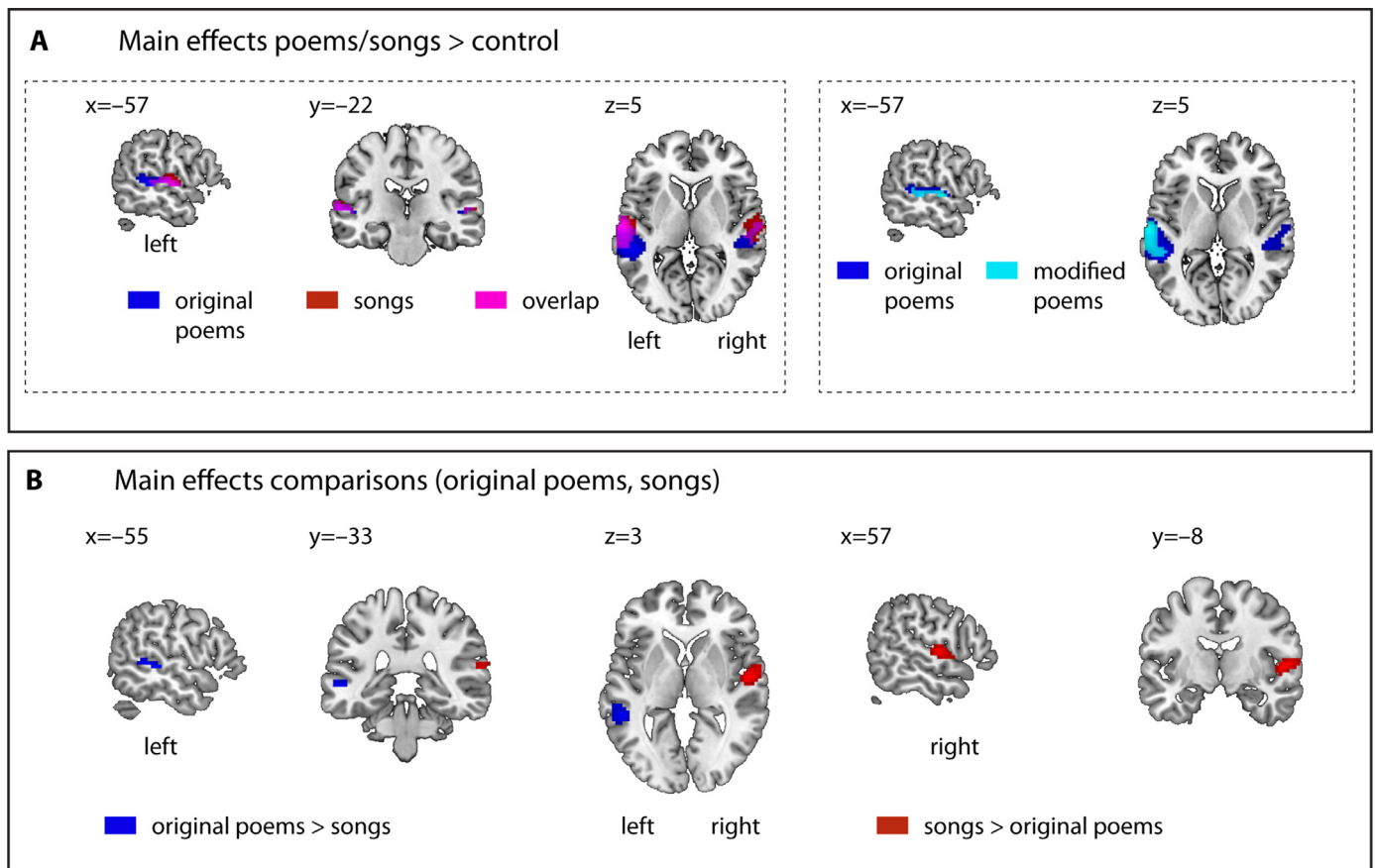
Fig. 4. Visualization of clusters emerging from the contrast between the test- and control conditions. (A) Main effects for original poems, modified poems and songs. For illustration purposes, overlapping areas are visually coded. Note that the left temporal cluster for modified poems lies entirely within the left temporal cluster for original poems. (B) Comparisons between activations of original poems and songs. All visualizations are based on thresholded contrast images (A: FWE corrected *p*-values at $\alpha < 0.05$; B: uncorrected at $\alpha < 0.001$) overlaid on a normalized canonical structural image in MRICRON.

higher PSC values than songs ($t = 2.59$, $p < 0.05$). In all other ROIs, there was no significant condition effect.

The beta model showed a main effect of *condition* (F(2,2893) = 16.33, $p < 0.001$), reflecting that modified poems elicited smaller betas than original poems ($t = 5.58$, $p < 0.001$) and songs ($t = 3.86$, $p < 0.001$), whereas songs and original poems did not differ ($t = 1.72$, $p > 0.08$). There was also an interaction of *condition* and *ROI* (F(10,2893) = 2.91, $p < 0.01$). Its decomposition showed main effects of condition in left STG (F(2,488) = 4.69, $p < 0.01$), posterior left STG (F(2,488) = 4.99, $p < 0.01$) and right STG (F(2,488) = 10.10, $p < 0.001$). In both left areas, higher betas were found for original poems than for modified poems ($t > 3.0$, $p < 0.01$). At the same time, modified poems showed lower beta values in the right STG than did songs ($t = 2.42$, $p < 0.05$). ROI-based activations are illustrated in Fig. 6.

3.5. Laterality of effects

On the basis of the above-mentioned ROIs, we computed laterality indices following the suggestions by Matsuo et al. (2012). The indices are bound between +1 (strongly left-lateralized) and -1 (strongly right-lateralized). Table 4 illustrates the average laterality indices for the whole-brain contrasts and modulations in the three selected ROIs.

4. Discussion

Spoken poems and songs based on the same words not only share structural, melodic properties, but their processing is also supported by partially overlapping neural processing areas. The most important find-

Table 4

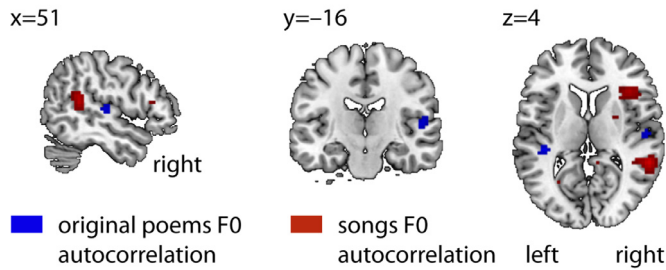Average laterality indices for contrasts/modulations in the three selected ROIs.

Contrast/Modulation	ROI	Average Laterality Index
Original poems>Controls	Heschl's Gyrus [±60 -16 -1]	0.38
Songs>Controls	Heschl's Gyrus [±60 -16 -1]	0.08
Original poems>Songs	STG [±49 -31 11]	0.91
Songs>Original poems	STG [±49 -31 11]	-0.93
F0 PSA Original Poems	STG [±49 -31 11]	0.57
F0 PSA Songs	STG [±49 -31 11]	-1
F0 PSA Original Poems	IFG [±38 17 12]	n/a
F0 PSA Songs	IFG [±38 17 12]	-0.73

ings of this study are the elicitation of a bilateral fronto-temporal network with hemispheric asymmetries, in line with previous research on music and language, and the covariation of activity within this fronto-temporal network with the pitch autocorrelations of the respective stimuli. We elucidate these points in reverse order in the following discussion. Importantly, all findings are based on ecologically valid stimuli, i.e., natural versions of spoken poems and their respective songs in their entire duration.

4.1. Neural sensitivity to pitch autocorrelations

The present study was conducted under the assumption that pitch autocorrelations provide a quantitative measure of melodic properties (Menninghaus et al., 2018; Scharinger et al., 2022). Previous studies that also capitalized on autocorrelations either focused on meter only (Brown, 1993; Eck, 2006; Toiviainen and Eerola, 2006; Vos et al., 1994),

A Parametric modulator: F0 autocorrelation

x=51   y=−16   z=4

right

■ original poems F0 autocorrelation   ■ songs F0 autocorrelation

left   right

B Parametric modulator: Continuous liking

x=10   y=−30   z=39

right

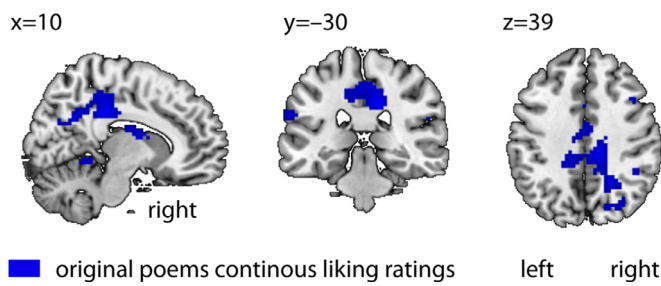■ original poems continous liking ratings

left   right

Fig. 5. Visualization of clusters emerging from the parametric modulators. (A) Correlations of F0 autocorrelation with BOLD. (B) Correlations of continuous liking ratings with BOLD. All visualizations are based on thresholded (uncorrected at *α*<0.005 with a cluster extent of 17 voxels) contrast images overlaid on a normalized canonical structural image in MRICRON.

their respective songs) to the hemodynamic response in specific brain regions as proxy of neural activity and processing.

Our endeavor was subject to substantial limitations. To start with, because the estimation of hemodynamic activity by means of the BOLD signal has a rather poor temporal resolution, we decided to analyze the pitch autocorrelations in epochs corresponding to the temporal extension of functional scans, i.e., in epochs of 2500 ms. This implies that we could not specifically focus on the autocorrelation structures dependent on higher-order building blocks of poems, i.e., individual verses and stanzas. Rather, we relied on the Proportions of Significant overall Autocorrelations (PSA) within these epochs. This implies that the selection of the 2500 ms epochs was not guided by the higher-order compositional structure of the songs.

Importantly, the positive and significant correlation of the PSA values obtained in this way are very similar to those obtained for the very same poems by the syllable-based analysis in Menninghaus et al. (2018), suggesting the validity of our approach. The higher PSA values for the songs compared to both versions of poems (original and modified; Fig. 3) furthermore support the feasibility of this approach. After all, the singing voice yields more time points with measurable f0 due to the ongoing vibration of the vocal folds. The absence of a difference between original and modified poems (Fig. 3), however, clearly shows the limitations of our approach that, by its very definition, cannot capture the relevant time-window of the entire stanzas. In the stanza-based autocorrelation analyses reported for the same original and modified poems in Menninghaus et al. (2018), the two conditions did differ.

Despite all these limitations, the PSA values of both original poems and songs significantly co-varied with the BOLD response in brain areas that support the processing of acoustic information. (Note that, for the modified poem versions, we obtained no such covariation with the BOLD response). Specifically, PSA values of original poems correlated with activity in bilateral posterior STG including parts of planum temporale and Heschl's gyrus. These areas define the region of the primary auditory cortex and surrounding structures, and are crucially involved in basic linguistic and musical pitch perception (Bianchi et al., 2017; De Angelis et al., 2018; Di Liberto et al., 2020; Hall and Plack, 2009; Patterson et al., 2016; Patterson et al., 2002; Wong et al., 2008). They support the basic stages of melody perception, if melody is parsimoniously defined as a structured sequence of pitches.

The brain areas identified by our parametric modulation analyses crucially comprise regions surrounding the primary auditory cortex, in-

or, if they also considered pitch (Liu et al., 2013), did not speculate on or directly test the neural bases of perceiving the autocorrelation structure. To our knowledge, our attempt is the first to directly relate autocorrelations of pitch in speech and music (i.e., spoken poems and

A Main effects (test > control)
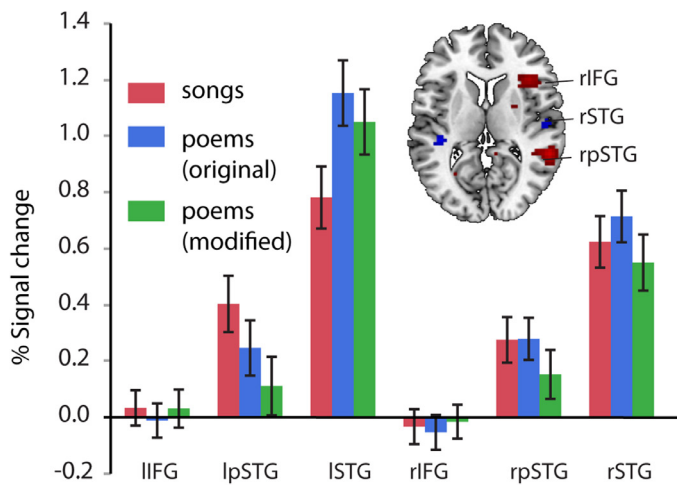
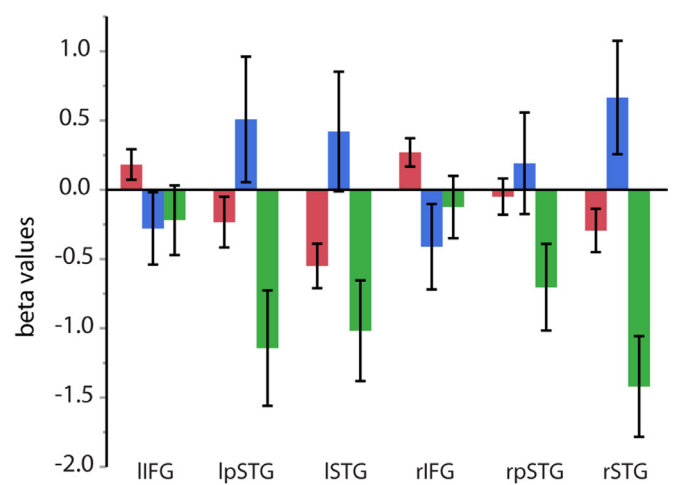## B Parametric modulator (F0 autocorrelation)

Fig. 6. Visualization of ROI-based activations. (A) Percentage signal change in 6 ROIs for the three test conditions. (B) Mean beta values from the parametric modulation analyses in 6 ROIs for the three test conditions. Whiskers represent the standard error of the mean.

cluding those regions that support the processing of more abstract acoustic properties such as pitch relations. Increasing distance from Heschl's gyrus is often equated with more abstract and complex auditory processing (Griffiths and Warren, 2002; Humphries et al., 2014), involving the support of auditory working memory (Jerde et al., 2011; Zatorre et al., 1994) and of the planum temporal as "computational hub" for auditory perception (Griffiths and Warren, 2002; Hickok and Saberi, 2012). We interpret the finding that PSA pitch values of poems activate this auditory network as evidence that melodic properties captured by pitch autocorrelations can be interpreted in a neurobiologically meaningful way.

The rather unexpected further cluster in the left occipital cortex, comprising parts of the fusiform face area (McCarthy et al., 1997; Rossion et al., 2003), might be interpreted by reference to visual imagery during listening to the recited poems in the scanner. Poetic language has long been credited with a unique power of evoking "vivid" mental imagery (for a collection of essays on this topic, see Avanessian and Völker 2009). Visual imagery is supported by the left fusiform area (Spagna et al., 2021), and it is plausible to assume that participants have either imagined the speaker of the poems, or the musical notation, or have otherwise had synaesthetic experiences (Satoh et al., 2015), perhaps also encouraged by the (optional) instruction to close their eyes. We deem it possible to speculate that the correlation with the pitch autocorrelation values might indicate the specific imagination of repetitive lip movements correlated with the autocorrelation structure of speech.

PSA values of songs, on the other hand, elicited a right-lateralized fronto-temporal network (see Table 3), with peaks in posterior STG and pars opercularis (and triangularis) in IFG. Again, the right-temporal area corresponds to auditory processing regions, with a posterior location in vicinity to the supramarginal gyrus. The supramarginal gyrus is considered to support pitch memory (Schaal et al., 2017; Schaal et al., 2015), thus playing an integral role in perceiving familiar melodies. The right IFG, on the other hand, is part of the prosodic network (Sammler et al., 2015), yet also plays a crucial role for "musical syntax", i.e., the processing of non-local, structural and hierarchical dependencies between tones of a melody (Bianco et al., 2016; Cheung et al., 2018; Koelsch, 2006, 2011; Kunert et al., 2015; Maess et al., 2001; Patel, 2005). Again, we interpret this finding as evidence that melodic properties of songs can well be captured by pitch autocorrelations in a neurobiologically plausible way.

The PSA values of songs correlated with the BOLD signal in a fronto-temporal network that overlaps with the melody network established by neuroimaging research on music perception, and partially also with a right-temporal area that we found to be implicated in the correlation of poem-based PSA values and the BOLD signal. Our results suggest that, due to the additional recruitment of the IFG (and, for that matter, left precentral cortex and precuneus), the higher-order processing of melodies is more pronounced in music than in speech. Perhaps, the musical settings carried more information for the participants than the spoken poems did. Altogether, our data suggest that pitch autocorrelations are a very promising quantitative measure to account for the neural processing of structured pitches in music and speech. Future research is necessary to further analyse to what degree the PSA values also encompass predictive information.

4.2. Hemispheric asymmetries in processing song and poems

The main effects we observed for the processing of songs and poems replicate a multitude of previous studies that found functional specializations in the left and right hemisphere depending on either linguistic or musical processing (e.g., Bianco et al. 2016; Brown et al. 2006; Koelsch et al. 2002; Levitin and Menon 2003; Zatorre and Baum 2012). Importantly, poems recruited a bilateral temporal network, with parts of planum temporale and Heschl's gyrus, but also comprising areas in middle temporal gyrus and superior temporal sulcus (STS) in the left hemisphere. The left STS has been established a "phonological" pro-

cessing hub and seems to support phonemic speech representations (Husain et al., 2006; Liebenthal et al., 2010; Scharinger et al., 2016). Its implication in poem processing is plausible, because phonemic decisions are required e.g., for evaluating rhyming words.

Songs likewise recruited a bilateral temporal network, in part overlapping with the poem network. Interestingly, the greater cluster (based on voxels) in the poem network was observed in the left, whereas the greater cluster in the song network was observed in the right hemisphere. This pattern was verified by the laterality indices, yielding a more left-biased index for poems than for songs. The asymmetry was further highlighted in the direct comparisons of poems and songs (and vice versa). Here, poems (compared to songs) showed more activation in the left posterior STG, STS and planum temporale, whereas songs (compared to poems) showed more activation in the right posterior STG, Heschl's gyrus and planum temporale. Again, this was corroborated by the laterality indices (Table 4). We interpret this asymmetry to reflect a greater need for phonemic processing (supported by the left STS) for poems, and a greater need for refined pitch processing (supported by the right Heschl's gyrus and right IFG discussed above) for songs.

The shared areas in STG are plausible, because in both cases, words — sung or spoken — were presented (see Merrill et al. 2012 for similar comparisons). Notably, modified poems recruited a smaller area in the left STG compared to the original poems, largely within the cluster of the original poems. It thus seems that the modified poems (due to the lack of meter and rhyme) elicit less intense processing than the original poems.

Given our high-level control condition (pitch-track derived from natural speech), the main contrasts reported above may be biased to some extent in that the pitch track from speech is closer to the poem than to the song condition. However, in both cases (poems and songs), the difference between the test and the control condition is to a large extent based on the contrast between speech and non-speech (i.e., intelligibility). The direct comparisons between original poems and songs are unaffected by such a potential confound, because the control condition is not considered in this contrast. We are therefore convinced that the possible confound introduced by the high-level baseline does not compromise our main finding.

4.3. Aesthetic perception in the scanner?

As with the PSA values which reflect objective phonetic properties of poem recitations, a study of subjective aesthetic experiences in an fMRI scanner requires some compromises. Because (melodic) autocorrelations of the sound properties of recited poems and songs can only be perceived over time, we decided to also capture the subjective responses with a time-sensitive measure, and hence with a continuous rating. And because we could not reasonably collect time-sensitive ratings for a variety of specific aesthetic virtues, we settled on collecting ratings only for overall liking to which the more specific dimensions of aesthetic evaluation—such as perceived beauty and melodiousness, being moved, etc.—make relevant contributions (for path analyses showing how such ratings differentially inform overall Liking ratings, see Menninghaus et al. 2020; Wassiliwizky et al. 2015). Moreover, liking ratings have previously been used in fMRI studies (e.g. Brattico et al. 2016; Zhao et al. 2020), and a review of uses of continuous aesthetically evaluative ratings during exposure to a variety of stimuli strongly suggests that such ratings, even though cognitively demanding, do not significantly alter aesthetic appreciation (Wagner et al., 2020). We therefore considered continuous liking ratings a good/viable option for measuring the aesthetically evaluative perception. Finally, the online-collected continuous liking ratings correlated reasonably with post-hoc liking ratings for a substantial subset of our participants who accepted our invitation to give another round of ratings six months after the fMRI study. This not only validates the continuous liking ratings as such but also the composite nature of their collection by means of the two pressure-sensitive buttons.

The continuous liking ratings correlated with brain activity in the reward network, in line with earlier empirical findings and theoretical suggestions (Bohrn et al., 2013; Sachs et al., 2016; Salimpoor et al., 2015; Wald, 2015; Wassiliwizky et al., 2015). Specifically, we found activation clusters in the prefrontal cortex, the anterior cingulate cortex and the thalamus, all key areas of the reward network. However, a far better match of the clusters obtained from the parametric modulator *liking ratings* was found with the default mode network (DMN). The default mode network, or "idling network" has recently received increased attention as a key player in aesthetic appreciation (Belfi et al., 2019; Vessel et al., 2019; 2013). Almost all clusters identified in our study – i.e., posterior and middle cingulate cortex, precuneus, medial frontal gyrus, dorsal prefrontal cortex and angular gyrus – are considered part of the DMN.

To be sure, previous evidence for the DMN's involvement in aesthetic appreciation is limited to studies that presented visual artworks. However, for the time being there is no reason to assume that the DMN should not play a similar role in responses to auditorily perceived and hence temporally unfolding artworks, such as poems and songs. Our findings can therefore be considered to extend earlier research into the role of the DMN in aesthetic appreciation of static images to two domains of the temporal arts, i.e., verbal art and music. By implication, this finding supports our goal to use ecologically valid stimuli in an fMRI environment and to still capture genuine aesthetic appreciation of the stimuli presented above EPI-noise.

An alternative interpretation would be that the correlations reflected the engagement or non-engagement in listening vs. focusing on the rating taks. It might be the case that increased engagement in listening was accompanied by decreased engagement in the rating task. Based on the present study, we cannot disentangle these potentially interacting factors. Future studies are therefore well-advised to bear this potential confound in mind.

The additional factor analysis of the post-hoc rating study provides further evidence that most participants were engaged in the aesthetic appreciation of the presented auditory stimuli. The set of scales used in the post-hoc study corresponded to previously used scales (Scharinger et al., 2022). The results of the factor analysis reflected a loading pattern that is quite similar to the one reported in this previous study. The correlation of the factor Melody with the continuous liking rating therefore validates our assumption that participants' liking responses in the scanner qualify as aesthetic responses that encompass several aspects of aesthetic appreciation with melodiousness playing a particularly prominent role. The other sets of groupings under the factors Beauty and Joy are also compatible with previous research on aesthetically evaluative ratings, specifically with regard to the fact that the beauty ratings correlate significantly with ratings for "being moved" (Menninghaus et al., 2015).

5. Conclusions

Research into commonalities between music and language has a long tradition. In this tradition, we investigated the role of melodic properties in poems and songs and operationalized the melodic properties in both domains by quantifying the degrees to which pitch sequences recur across recitations of the respective poems and songs. To reduce variance beyond the two conditions poem and song, we exclusively relied on songs that are based on poems and hence are identical in text and content. Importantly, the measure of pitch autocorrelations explained brain activity during our brain imaging experiment in bilateral temporal auditory areas, with additional activations in right-frontal areas for songs. We have thus provided evidence for the neurobiological plausibility of the autocorrelation measure for capturing melodic properties of music and language alike and simultaneously obtained patterns of activation that support the specificity of musical melodies at higher-order processing stages. Moreover, we extended earlier findings regarding the role of the DMN in the processing of images to the processing of temporally unfolding artworks in the domains of both language and music.

Supplementary materials

Supplementary materials associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119310.

References

Alkon, P.P., 1959. Joshua steele and the melody of speech. Lang. Speech 2, 154–174.

Angulo-Perkins, A., Concha, L., 2019. Discerning the functional networks behind processing of music and speech through human vocalizations. PLoS One 14, e0222796.

Ashburner, J., Good, C.D., Tofts, P., 2003. Spatial registration of images. In: Qualitative MRI of the Brain: Measuring Changes Caused by Disease. John Wiley & Sons, pp. 503–531.

Avanessian, A., Völker, J., 2009. Vita Aesthetica. Szenarien Ästhetischer Lebendigkeit. Diaphanes, Zürich.

Belfi, A.M., Vessel, E.A., Brielmann, A., Isik, A.I., Chatterjee, A., Leder, H., Pelli, D.G., Starr, G.G., 2019. Dynamics of aesthetic experience are reflected in the default-mode network. Neuroimage 188, 584–597.

Besson, M., Schön, D., 2001. Comparison between language and music. Ann. N.Y. Acad. Sci. 930, 232–258.

Bever, T.G., Chiarello, R.J., 1974. Cerebral dominance in musicians and nonmusicians. Science 185, 537–539.

Bianchi, F., Hjortkjær, J., Santurette, S., Zatorre, R.J., Siebner, H.R., Dau, T., 2017. Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians. Neuroimage 163, 398–412.

Bianco, R., Novembre, G., Keller, P.E., Kim, S.G., Scharf, F., Friederici, A.D., Villringer, A., Sammler, D., 2016. Neural networks for harmonic structure in music perception and action. Neuroimage 142, 454–464.

Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat. Neurosci. 8, 389–395.

Boersma, P., Weenink, D., 2019. PRAAT: Doing Phonetics by Computer (ver. 6.0.50). Institut for Phonetic Sciences, Amsterdam.

Bohrn, I.C., Altmann, U., Lubrich, O., Menninghaus, W., Jacobs, A.M., 2013. When we like what we know – a parametric fMRI analysis of beauty and familiarity. Brain Lang. 124, 1–8.

Brattico, E., Bogert, B., Alluri, V., Tervaniemi, M., Eerola, T., Jacobsen, T., 2016. It's sad but i like It: the neural dissociation between musical emotions and liking in experts and laypersons. Front. Hum. Neurosci. 9.

Brattico, E., Pearce, M., 2013. The neuroaesthetics of music. Psychol. Aesthet. Creat. Arts 7, 48–61.

Bretz, F., Hothorn, T., Westfall, P., 2011. Multiple Comparisons Using R. CRC Press.

Britton, B., Blumstein, S.E., Myers, E.B., Grindrod, C., 2009. The role of spectral and durational properties on hemispheric asymmetries in vowel perception. Neuropsychologia 47, 1096–1106.

Brown, J.C., 1993. Determination of the meter of musical scores by autocorrelation. J. Acoust. Soc. Am. 94, 1953–1957.

Brown, S., Martinez, M.J., Parsons, L.M., 2006. Music and language side by side in the brain: a PET study of the generation of melodies and sentences. Eur. J. Neurosci. 23, 2791–2803.

Burns, E.M., Deutsch, D., 1999. Intervals, scales, and tuning. In: The Psychology of Music. Academic Press, pp. 215–264.

Calma-Roddin, N., Drury, J.E., 2020. Music, language, and the N400: ERP interference patterns across cognitive domains. Sci. Rep. 10, 11222.

Cheung, V.K.M., Meyer, L., Friederici, A.D., Koelsch, S., 2018. The right inferior frontal gyrus processes nested non-local dependencies in music. Sci. Rep. 8, 1–12.

Chow, I., Brown, S., 2018. A musical approach to speech melody. Front. Psychol. 9, 247.

Christiansen, P., 2004. The meaning of speech melody for Leoš Janáček. J. Musicol. Res. 23, 241–263.

Clunies-Ross, K.L., Brydges, C.R., Nguyen, A.T., Fox, A.M., 2015. Hemispheric asymmetries in auditory temporal integration: a study of event-related potentials. Neuropsychologia 68, 201–208.

De Angelis, V., De Martino, F., Moerel, M., Santoro, R., Hausfeld, L., Formisano, E., 2018. Cortical processing of pitch: Model-based encoding and decoding of auditory fMRI responses to real-life sounds. Neuroimage 180, 291–300.

Di Liberto, G.M., Pelofi, C., Bianco, R., Patel, P., Mehta, A.D., Herrero, J.L., de Cheveigne, A., Shamma, S., Mesgarani, N., 2020. Cortical encoding of melodic expectations in human temporal cortex. eLife 9, e51784.

Eck, D., 2006. Identifying metrical and temporal structure with an autocorrelation phase matrix. Music Percept. 24, 167–176.

Farbood, M.M., Heeger, D.J., Marcus, G., Hasson, U., Lerner, Y., 2015. The neural processing of hierarchical structure in music and speech at different timescales. Front. Neurosci. 9, 157.

Friston, K.J., Frackowiak, R.S., Friston, K.J., Frith, C.D., Dolan, R.J., Price, C., Zeki, S., 2004. Experimental design and statistical parametric mapping. In: Human Brain Function. Academic Press, Amsterdam, pp. 599–632.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. Neuroimage 23, 344–357.

Garcea, F.E., Chernoff, B.L., Diamond, B., Lewis, W., Sims, M.H., Tomlinson, S.B., Teghipco, A., Belkhir, R., Gannon, S.B., Erickson, S., Smith, S.O., Stone, J., Liu, L., Tollefson, T., Langfitt, J., Marvin, E., Pilcher, W.H., Mahon, B.Z., 2017. Direct electrical stimulation in the human brain disrupts melody processing. Curr. Biol. 27, 2684–2691 e2687.

Gordon, R.L., Schön, D., Magne, C., Astésano, C., Besson, M., 2010. Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. PLoS One 5.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348–353.

Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A., Summerfield, A.Q., 2002. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 12, 140–149.

Hall, D.A., Plack, C.J., 2009. Pitch processing sites in the human auditory brain. Cereb. Cortex 19, 576–585.

Hickok, G., Saberi, K., Poeppel, D., Overath, T., Popper, A.N., Fay, R.R., 2012. Redefining the functional organization of the planum temporale region: space, objects and sensory–motor integration. In: The Human Auditory Cortex. Springer, New York, pp. 333–350.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185.

Humphries, C., Sabri, M., Lewis, K., Liebenthal, E., 2014. Hierarchical organization of speech perception in human auditory cortex. Front. Neurosci. 8, 1–12.

Huron, D.B., 2006. Sweet Anticipation: Music and the Psychology of Expectation. MIT Press.

Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., Horwitz, B., 2006. Neural bases of categorization of simple speech and nonspeech sounds. Hum. Brain Mapp. 27, 636–651.

Jakobson, R., Sebeok, T., 1960. Linguistics and poetics. In: Style in Language. MIT Press, Cambridge, pp. 350–377.

Janata, P., 2015. Neural basis of music perception. Handb. Clin. Neurol. 129, 187–205.

Jay, W., 1990. Expectancy and attention in melody perception. Psychomusicol. J. Res. Music Cognit. 9, 148–160.

Jenkinson, M., Beckman, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790.

Jerde, T.A., Childs, S.K., Handy, S.T., Nagode, J.C., Pardo, J.V., 2011. Dissociable systems of working memory for rhythm and melody. Neuroimage 57, 1572–1579.

Kawahara, H., Morise, M., 2011. Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. Sadhana 36, 713–727.

Koelsch, S., 2006. Significance of Broca's area and ventral premotor cortex for music-syntactic processing. Cortex 42, 518–520.

Koelsch, S., 2009. Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. Psychophysiology 46, 179–190.

Koelsch, S., 2011. Toward a neural basis of music perception – a review and updated model. Front. Psychol. 2, 110.

Koelsch, S., Gunter, T.C., v Cramon, D.Y., Zysset, S., Lohmann, G., Friederici, A.D., 2002. Bach speaks: a cortical "language-network" serves the processing of music. Neuroimage 17, 956–966.

Koelsch, S., Jentschke, S., Sammler, D., Mietchen, D., 2007. Untangling syntactic and sensory processing: an ERP study of music perception. Psychophysiology 44, 476–490.

Koelsch, S., Siebel, W.A., 2005. Towards a neural basis of music perception. Trends Cogn. Sci. 9, 578–584.

Koelsch, S., Vuust, P., Friston, K., 2019. Predictive processes and the peculiar case of music. Trends Cogn. Sci. 23 (1), 63–77.

Kreitewolf, J., Friederici, A.D., von Kriegstein, K., 2014. Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. Neuroimage 102, 332–344.

Kunert, R., Willems, R.M., Casasanto, D., Patel, A.D., Hagoort, P., 2015. Music and language syntax interact in Broca's area: an fMRI study. PLoS One 10, e0141069.

Kuznetsova, A., Bruun Brockhoff, P., Bojesen Christensen, R., 2014. lmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0–6. http://CRAN.R-project.org/package=lmerTest.

Lappe, C., Steinsträter, O., Pantev, C., 2013. Rhythmic and melodic deviations in musical sequences recruit different cortical areas for mismatch detection. Front. Hum. Neurosci. 7, 260.

Leipold, S., Brauchli, C., Greber, M., Jäncke, L., 2019. Absolute and relative pitch processing in the human brain: neural and behavioral evidence. Brain Struct. Funct. 224, 1723–1738.

Levitin, D.J., Menon, V., 2003. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann area 47 in temporal coherence. Neuroimage 20, 2142–2152.

Liebenthal, E., Desai, R., Ellingson, M.M., Ramachandran, B., Desai, A., Binder, J.R., 2010. Specialization along the left superior temporal sulcus for auditory categorization. Cereb. Cortex 20, 2958–2970.

Liu, L., Wei, J., Zhang, H., Xin, J., Huang, J., 2013. A statistical physics view of pitch fluctuations in the classical music from bach to chopin: evidence for scaling. PLoS One 8, e58710.

Luo, H., Poeppel, D., 2012. Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. Front. Psychol. 3, 170.

Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4, 540–545.

Matsuo, K., Chen, S.H.A., Tseng, W.Y.I., 2012. AveLI: a robust lateralization index in functional magnetic resonance imaging using unbiased threshold-free computation. J. Neurosci. Methods 205, 119–129.

McCarthy, G., Puce, A., Gore, J.C., Allison, T., 1997. Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9, 605–610.

Mehr, S.A., Singh, M., Knox, D., Ketter, D.M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A.A., Hopkins, E.J., Howard, R.M., Hartshorne, J.K., Jennings, M.V., Simson, J., Bainbridge, C.M., Pinker, S., O'Donnell, T.J., Krasnow, M.M., Glowacki, L., 2019. Universality and diversity in human song. Science 366, 1–17.

Menninghaus, W., Schindler, I., Wagner, V., Wassiliwizky, E., Hanich, J., Jacobsen, T., Koelsch, S., 2020. Aesthetic emotions are a key factor in aesthetic evaluation: reply to Skov and Nadal (2020). Psychol. Rev. 127, 650–654.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., Jacobsen, T., 2015. Towards a psychological construct of being moved. PLoS One 10, e0128451.

Menninghaus, W., Wagner, V., Knoop, C.A., Scharinger, M., 2018. Poetic speech melody: a crucial link between music and language. PLoS One 13, e0205980.

Menninghaus, W., Wallot, S., 2021. What the eyes reveal about (poetry) poetry. Poetics 85. doi:10.1016/j.poetic.2020.101526.

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., Friederici, A.D., 2012. Perception of words and pitch patterns in song and speech. Front. Psychol. 3, 76.

Nan, Y., Friederici, A.D., 2013. Differential roles of right temporal cortex and Broca's area in pitch processing: evidence from music and Mandarin. Hum. Brain Mapp. 34 (9), 2045–2054.

Obleser, J., Eisner, F., Kotz, S.A., 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J. Neurosci. 28, 8116–8123.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9, 97–113.

Omigie, D., Pearce, M., Lehongre, K., Hasboun, D., Navarro, V., Adam, C., Samson, S., 2019. Intracranial recordings and computational modeling of music reveal the time course of prediction error signaling in frontal and temporal cortices. J. Cogn. Neurosci. 31, 855–873.

Patel, A.D., 2005. The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. Ann. N.Y. Acad. Sci. 1060, 59–70.

Patel, A.D., 2010. Music, Language, and the Brain. Oxford University Press, New York.

Patterson, R.D., Andermann, M., Uppenkamp, S., Rupp, A., 2016. Locating melody processing activity in auditory cortex with magnetoencephalography. Adv. Exp. Med. Biol. 894, 363–369.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776.

Peretz, I., Vuvan, D., Lagrois, M.E., Armony, J.L., 2015. Neural overlap in processing music and speech. Philos. Trans. R. Soc. B Biol. Sci, 370, 68–75.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun. 41, 245–255.

Reber, R., Schwarz, N., Winkielman, P., 2004. Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? Pers. Soc. Psychol. Rev. 8, 364–382.

Revelle, W., 2020. Psych: procedures for personality and psychological research. R package version 2.1.6.

Rohrmeier, M.A., Koelsch, S., 2012. Predictive information processing in music cognition. A critical review. Int. J. Psychophysiol. 83, 164–175.

Rossion, B., Caldara, R., Seghier, M., Schuller, A.M., Lazeyras, F., Mayer, E., 2003. A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain 126, 2381–2395.

Sachs, M.E., Ellis, R.J., Schlaug, G., Loui, P., 2016. Brain connectivity reflects human aesthetic responses to music. Soc. Cogn. Affect. Neurosci. 11 (6), 884–891.

Salimpoor, V.N., Zald, D.H., Zatorre, R.J., Dagher, A., McIntosh, A.R., 2015. Predictions and the brain: how musical sounds become rewarding. Trends Cogn. Sci. 19, 86–91.

Sammler, D., 2020. Splitting speech and music. Science 367, 974–976.

Sammler, D., Elmer, S., 2020. Advances in the neurocognition of music and language. Brain Sci. 10 (8), 509.

Sammler, D., Grosbras, M.H., Anwander, A., Bestelmeyer, P.E.G., Belin, P., 2015. Dorsal and ventral pathways for prosody. Curr. Biol. 25, 3079–3085.

Satoh, M., Nagata, K., Tomimoto, H., 2015. Sound richness of music might be mediated by color perception: a PET study. Behav. Neurol. 2015, 241804.

Schaal, N.K., Pollok, B., Banissy, M.J., 2017. Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. Sci. Rep. 7, 424–456.

Schaal, N.K., Williamson, V.J., Kelly, M., Muggleton, N.G., Pollok, B., Krause, V., Banissy, M.J., 2015. A causal involvement of the left supramarginal gyrus during the retention of musical pitches. Cortex 64, 310–317.

Scharinger, M., Domahs, U., Klein, E., Domahs, F., 2016. Mental representations of vowel features asymmetrically modulate activity in superior temporal sulcus. Brain Lang. 163, 42–49.

Scharinger, M., Wagner, V., Knoop, C.A., Menninghaus, W., 2022. Melody in songs and poems: fundamental statistical properties predict aesthetic evaluation. Psychol. Aesthet. Creat. Arts doi:10.1037/aca0000465.

Schmuckler, M.A., 1989. Expectation in music: Investigation of melodic and harmonic processes. Music Percept. 7, 109–149.

Schön, D., Gordon, R.L., Besson, M., 2005. Musical and linguistic processing in song perception. Ann. N.Y. Acad. Sci. 1060, 71–81.

Schönwiesner, M., Rübsamen, R., von Cramon, D.Y., 2005. Spectral and temporal processing in the human auditory cortex–revisited. Ann. N.Y. Acad. Sci. 1060, 89–92.

Sievers, E., 1912. Rhythmisch-Melodische Studien. Rhythmisch-Melodische Studien. C. Winter, Heidelberg.

Slotnick, S.D., Moo, L.R., Segal, J.B., Hart, J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Brain Res. Cognit. Brain Res. 17, 75–82.

Spagna, A., Hajhajate, D., Liu, J., Bartolomeo, P., 2021. Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: a meta-analysis of neuroimaging evidence. Neurosci. Biobehav. Rev. 122, 201–217.

Steele, J., 1775. An Essay Towards Establishing the Melody and Measure of Speech to Be Expressed and Perpetuated By Peculiar Symbols. Bowyer and Nichols, London.

Steinbeis, N., Koelsch, S., 2008. Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. PLoS One 3 (5). doi:10.1371/journal.pone.0002226.

Toiviainen, P., Eerola, T., 2006. Autocorrelation in meter induction: the role of accent structure. J. Acoust. Soc. Am. 119, 1164–1170.

Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2005. Neural circuitry underlying sentence-level linguistic prosody. Neuroimage 28, 417–428.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.

Vessel, E.A., Isik, A.I., Belfi, A.M., Stahl, J.L., Starr, G.G., 2019. The default-mode network represents aesthetic appeal that generalizes across visual domains. Proc. Natl. Acad. Sci. USA 116, 19155–19164.

Vessel, E.A., Starr, G.G., Rubin, N., 2013. Art reaches within: Aesthetic experience, the self and the default mode network. Front. Neurosci. 7. doi:10.3389/fnins.2013.00258.

Vos, P.G., van Dijk, A., Schomaker, L., 1994. Melodic cues for metre. Perception 23, 965–976.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., Tervaniemi, M., 2012. The sound of music: differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. Neuropsychologia 50, 1432–1443.

Wagner, V., Scharinger, M., Knoop, C.A., Menninghaus, W., 2020. Effects of continuous self-reporting on aesthetic evaluation and emotional responses. Poetics 85, 101497.

Wald, C., 2015. Neuroscience: the aesthetic brain. Nature 526, S2–S3.

Wassiliwizky, E., Wagner, V., Jacobsen, T., Menninghaus, W., 2015. Art-elicited chills indicate states of being moved. Psychol. Aesthet. Creat. Arts 9, 405–416.

Westbury, C.F., Zatorre, R.J., Evans, A.C., 1999. Quantifying variability in the planum temporale: a probability map. Cereb. Cortex 9, 392–405.

Wingfield, P., 1992. Janacek's speech-melody theory in concept and practice. Camb. Opera J. 4, 281–301.

Winn, J.A., 1984. Unsuspected Eloquence: History of the Relations Between Poetry and Music. Yale University Press, New Haven, CT.

Wong, P.C.M., Warrier, C.M., Penhune, V.B., Roy, A.K., Sadehh, A., Parrish, T.B., Zatorre, R.J., 2008. Volume of left Heschl's gyrus and linguistic pitch learning. Cereb. Cortex 18, 828–836.

Yu, M., Xu, M., Li, X., Chen, Z., Song, Y., Liu, J., 2017. The shared neural basis of music and language. Neuroscience 357, 208–219.

Zajonc, R.B., 1968. Attitudinal effects of mere exposure. J. Pers. Soc. Psychol. 9, 1–27.

Zatorre, R.J., 2013. Predispositions and Plasticity in Music and Speech Learning: neural correlates and Implications. Science 342, 585–589.

Zatorre, R.J., Baum, S.R., 2012. Musical melody and speech intonation: singing a different tune. PLoS Biol. 10, e1001372.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11, 946–953.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14, 1908–1919.

Zhao, X., Wang, J., Li, J., Luo, G., Li, T., Chatterjee, A., Zhang, W., He, X., 2020. The neural mechanism of aesthetic judgments of dynamic landscapes: an fMRI study. Sci. Rep. 10, 20774.