Article

sempre

ociety for Education, Music and Psychology Research

# Detection of pitch errors in well-known songs

Psychology of Music 1 - 16© The Author(s) 2022  $\odot$   $\odot$ Article reuse guidelines:

sagepub.com/journals-permissions DOI: 10.1177/03057356221087447 journals.sagepub.com/home/pom

# Michael W Weiss<sup>1</sup> and Sandra E Trehub<sup>2</sup>

#### Abstract

We examined pitch-error detection in well-known songs sung with or without meaningful lyrics. In Experiment 1, adults heard the initial phrase of familiar songs sung with lyrics or repeating syllables (*la*) and judged whether they heard an out-of-tune note. Half of the renditions had a single pitch error (50 or 100 cents); half were in tune. Listeners were poorer at pitch-error detection in songs with lyrics. In Experiment 2, within-note pitch fluctuations in the same performances were eliminated by auto-tuning. Again, pitch-error detection was worse for renditions with lyrics (50 cents), suggesting adverse effects of semantic processing. In Experiment 3, songs were sung with repeating syllables or *scat* syllables to ascertain the role of phonetic variability. Performance was poorer for scat than for repeating syllables, indicating adverse effects of phonetic variability, but overall performance exceeded Experiment 1. In Experiment 4, listeners evaluated songs in all styles (repeating syllables, scat, lyrics) within the same session. Performance was best with repeating syllables (50 cents) and did not differ between scat or lyric versions. In short, tracking the pitches of highly familiar songs was impaired by the presence of words, an impairment stemming primarily from phonetic variability rather than interference from semantic processing.

#### **Keywords**

pitch, singing, lyrics, semantic processing, music cognition

What factors influence our detection of pitch deviations or out-of-tune pitches in songs? Often, Western melodies are aligned with the framework of a scale, such as the major or minor scales in Western tonal music, and have a realized or implied tonal center or tonic. Pitch deviations could take the form of a note that is mistuned in the context of the prevailing scale or a note that is outside of the scale. Enculturation to tonal expectations occurs early in life through passive exposure to music (Trainor & Hannon, 2013). Moreover, when a melody is

**Corresponding author:** 

<sup>&</sup>lt;sup>1</sup>International Laboratory for Brain, Music, and Sound Research, University of Montreal, Montreal, QC, Canada <sup>2</sup>University of Toronto Mississauga, Mississauga, ON, Canada

Michael W Weiss, International Laboratory for Brain, Music, and Sound Research, University of Montreal, Pavillon Marie-Victorin 90 avenue Vincent d'Indy, Montreal, QC H3C 3J7, Canada. Email: michael.william.weiss@gmail.com

highly familiar, listeners have strong expectations of the successive intervals. Perception of a pitch deviation could be influenced by the magnitude of the change or other features of the note such as its duration (Micheyl et al., 2012), timbre (Allen & Oxenham, 2014), or intra-note pitch variability (van Besouw et al., 2008). Melodic and tonal context can also influence the perception of pitch (Marmel et al., 2008). In short, detection of a pitch deviation rests on the strength of expectations and on the salience of the event.

The current study focuses on vocal melodies. The voice is probably the oldest instrument, and its presence is ubiquitous in music across cultures. It is more dynamic and expressive than other sound sources (Schubert & Wolfe, 2016; Sundberg, 1994) primarily because of the flexibility of vocal-motor physiology (Wolfe et al., 2020). Like many other instruments, pitched vocal tones are generated by a vibrating sound source, forcing air past the vocal folds in the larynx. Unlike other instruments, the shape of the instrument (vocal tract), and hence its resonant properties, can be changed rapidly (Sundberg, 1977). Early research on singing revealed vocal tones to be more variable in acoustic features than instrumental tones (Simon, 1926). Deviations of intonation from equal tempered tuning are typical within notes, even for professional singers (Sundberg, 2013). For listeners, the ability to detect pitch deviations in sung tones (e.g., /a/) is reduced relative to instrumental tones, resulting in what is known as "vocal generosity" (Hutchins et al., 2012). Expressiveness in pitch and timbre contributes to the unique status of the voice, even among instruments considered "voice-like" (Schubert, 2019).

The voice differs from other instruments in its primary evolutionary function as a biological, conspecific signal, and in its communication of identity and emotional state. More importantly, the voice is the only instrument with the capacity to transmit the full gamut of verbal linguistic information. The connection between music and language processing is evident in the speech-to-song illusion, whereby successive repetitions of a spoken phrase shift perception of the phrase from speech to singing (Deutsch et al., 2011). Impairments in the intelligibility of lyrics in sung versus spoken stimuli or in different genres of music highlight the interference of musical information on linguistic processing (Condit-Schultz & Huron, 2015; Johnson et al., 2013).

Here, we ask whether the lyrical content of songs influences the salience of their pitches. The question is of considerable importance because vocal music with and without lyrics seems to be universal, so understanding the mechanisms of song perception has implications for a major class of music. It is also of theoretical interest to ascertain whether the interference between music and linguistic information is bidirectional. Most well-known songs have words, but the songs are easily recognized and produced without the words. On one hand, the words of songs increase listeners' and singers' processing demands because of variable linguistic (phonetic, semantic) and acoustic (fundamental frequency, amplitude, timbre) features. Young school-age children generally sing songs with words, but they achieve greater pitch accuracy when singing on a neutral syllable (Pereira & Rodrigues, 2019), highlighting the potential consequences of linguistic processing on melody production. On the other hand, listeners, especially adults, may engage in automatic semantic processing of highly overlearned songs, with no impact on pitch processing. Nevertheless, changes in higher-level linguistic features (semantic) coincide with changes in lower-level features (phonetic), and there are demonstrable consequences of vowel changes on the perception of interval size (Russo et al., 2019). In the case of infants, 11-month-olds more readily detect pitch changes in sung sequences when the component syllables are uniform rather than variegated (Lebedeva & Kuhl, 2010). To our knowledge, however, there have been no investigations of the influence of lyrics on adults' detection of pitch errors.

In the present study, we examined the influence of singing style on listeners' perception of pitch errors in familiar songs. In each of four experiments, participants heard sung excerpts

from well-known melodies that were perfectly in tune or had pitch deviations on one note (Figure 1). Experiment 1 asked whether the detection of pitch errors was affected by the presence of meaningful lyrics. We hypothesized that meaningful lyrics relative to repetitive syllables (*la la*) would impair performance, presumably due to the burden of additional processing of dynamic linguistic information (phonetic, semantic). Experiment 2 asked whether impairment related to lyrics could be attributed to pitch dynamics alone. We hypothesized that a true effect of lyrics would remain evident even when pitch dynamics were removed. Experiment 3 asked whether the effect of lyrics could be attributed to phonetic dynamics. We hypothesized that the detection of pitch errors would be affected by meaningless but variable syllabic content (i.e., scat singing), but it was unclear whether singing with lyrics impairs performance relative to singing with meaningless syllables, whether uniform or variable (i.e., *la la*, scat, lyrics). We hypothesized that pitch-error detection would differ across levels, in keeping with the additional processing of dynamics for both phonetic and semantic information.

#### General method

#### Participants

Participants were recruited from an introductory psychology course or by flyers posted on campus. Demographic information is provided in the context of each experiment. All participants were undergraduate students at an English-language university that reports a majority of Canadian citizens in enrollment statistics (76%), and we assumed that our samples had a similar distribution. Listening habits were not surveyed, but prior samples from this population have reported listening overwhelmingly to genres that follow the conventions of Western tonal music (e.g., pop, rock), and all participants confirmed their familiarity with stimulus melodies. Participants received partial course credit or modest remuneration. Participants who performed below chance, indicating inattention or pitch-perception difficulties, were excluded from the sample. Written informed consent was obtained prior to testing. The study was approved by the Office of Research Ethics, University of Toronto (protocol #30622).

#### Stimuli

An amateur female vocalist recorded 10 excerpts from songs with well-known lyrics and melodies (Figure 1). Melodies were chosen based on their cultural familiarity and included different genres and rhythms. Each melody was performed in three ways: with original lyrics, the same syllable (*la*) on every note, and a scat version with alternating syllables (e.g., *doo bah dee bah*). The melodies were recorded with Logic (Apple), with the singer listening to subsequently discarded MIDI backing and click tracks to promote accurate tuning and timing (120 beats per minute). Recordings were imported into Melodyne (Celemony, Inc.), which allowed for pitch correction by simultaneously (1) centering each note to the correct frequency and (2) correcting the note's pitch drift (i.e., if the note drifted out of tune over time). This method of pitch correction results in natural-sounding, high-quality recordings, with the singer's timbre and vibrato unaffected. Sample stimuli are provided as Supplementary Materials online.

To create the mistuned stimuli, the same original recordings were imported into Melodyne and corrected, as described, but one note was subsequently mistuned. Two notes in each melody were selected for mistuning, one in the first half of the melody and one in the second half, with restrictions of (1) no mistuning of the first note in the melody and (2) no mistuning that



**Figure 1.** Stimuli. Note: Excerpts of 10 familiar melodies sung in different styles across experiments, including singing with lyrics, singing with *la* on every note, and scat singing with alternating nonsense syllables. All singing was pitch-corrected in Melodyne. On half of the trials, pitch errors were applied to a single note either earlier or later in the excerpt (red notes). Singing style, magnitude of the pitch error (50, 100 cents), direction of pitch error (up, down), and placement of pitch error (earlier, later) was balanced for each individual, with melody assigned at random.

resulted in repeated notes. As a result, the selected notes varied in duration and placement. Eight varieties of mistuning were created for each melody: early versus late, upward (sharp) versus downward (flat), and quarter tone (50 cents) versus semitone (100 cents). Melodyne produces pitch shifts of this magnitude or greater without noticeable artifacts.

To prime listeners' expectations about the melody to follow, each song name was announced 2 s before song onset. Such priming was necessary because mistuned notes could occur as early as the second note in the melody. In principle, differences in singing style could generate notes that differed systematically in duration. The durations of the 20 target notes in each singing style were measured with PRAAT (Boersma & Weenink, 2020). A repeated-measures analysis of variance (ANOVA) revealed no difference in duration across singing styles, F < 1. Average and standard deviations of duration (ms) were similar for repeated syllable (M = 556, SD = 324), lyric (M = 533, SD = 301), and scat versions (M = 532, SD = 266), for an average value of 540 ms across all notes. A follow-up analysis compared the difference between the repeated syllable version of each target note and the versions with lyrics or scat syllables. Confirming the results of the ANOVA, the mean duration difference (ms) across styles was negligible (lyrics: M = 23, SD = 65; scat: M = 23, SD = 113) relative to the overall distribution of note durations (M = 540, SD = 293).

A similar analysis was carried out on within-note pitch fluctuations. For each note, continuous pitch, as measured in PRAAT (fundamental frequency every 0.01 s), was converted to semitones (MIDI format) and the standard deviation was calculated for each target note. A repeated-measures ANOVA found no difference in within-note pitch fluctuations across singing styles, F < 1. Average and standard deviations of pitch fluctuations (cents) were similar for singing with a repeating syllable (M = 51, SD = 20), lyrics (M = 57, SD = 36), and scat syllables (M = 50, SD = 19), for an average value of 52 cents (0.52 semitones) across all notes. In short, despite opting for natural, ecologically valid materials, there was no evidence of unintended acoustic consequences of the manipulations.

#### Procedure

Participants were tested individually in a double-walled sound-attenuating booth (Industrial Acoustics). Stimuli were presented over high-quality headphones (Sony) at a comfortable supra-threshold volume using the program PsyScript. After obtaining written consent, participants were instructed to (1) listen to short melodies that may or may not contain a single incorrect note, which could occur at any point in the melody, and (2) indicate, after the melody concluded, whether or not they heard an incorrect note (i.e., binary response). Participants were not required to indicate the position of incorrect notes. Responses via button press were recorded automatically. Participants were told that the melodies would be familiar and that song titles would be announced before each trial. The experimenter remained in the booth while participants completed two practice trials—one with a pitch error, one without—that required correct responses to proceed. The practice melody ("O Canada") was not used in the rest of the experiment. If the participant had no further questions after the practice trials, the experimenter left the booth and allowed the participant to begin.

In Experiments 1 to 3, trials were structured in four continuously presented blocks of 32 melodies (n=128 trials total), with blocks comprising the eight error types in each of two singing styles plus a matching number of in-tune melodies. A given trial could be perfectly in tune (n=64 trials) or contain a single note that was mistuned by 50 or 100 cents, in a sharp or flat direction, and heard early or late in the melody (n=8 trials per cell). The 10 melodies were assigned randomly to error types, and the set was presented in a semi-random fashion, with the constraint that the same melody could not occur on successive trials (regardless of the presence

of error or type of error). Experiment 4 differed in the number of singing styles (3 rather than 2), with an accompanying increase in the overall number of trials (n = 192 rather than 128).

After completing the task, participants provided information about their demographic background, musical training, and music listening habits, and were fully debriefed.

# Data analysis

Difference scores (hits minus false alarms) were the measure of analysis, such that a score of 1 indicated perfect (error-free) performance and a score of 0 indicated chance performance. Hits (i.e., correctly indicating that an error was present) were derived by taking the proportion of correctly identified errors of one type (e.g., trials with repeated syllables and errors of 50 cents). False alarms (i.e., reporting an error when no error occurred) represented the proportion of erroneously identified errors in that same singing style (e.g., trials with repeated syllables and no pitch errors). In other words, both levels of errors (50, 100 cents) used a common baseline within each singing style. A series of repeated-measures ANOVAs assessed whether performance differed by singing style and (1) magnitude of pitch error (50, 100 cents), (2) location of pitch error (early, late), and (3) direction of pitch error (higher, lower). A final test (4) compared singing style against error types for notes of different duration, regardless of location or magnitude. First, the 20 notes of each singing type were split at the median value and assigned the status of "shorter" or "longer." Next, difference scores were calculated as above, with the exception that the number of trials used to calculate percentages was determined dynamically because melodies were assigned randomly to error types. This corrected for the fact that, because of random assignment of melody to pitch-error type, some participants heard more "longer" notes, while others heard more "shorter" notes.

# Experiment I

Listeners heard short melodies sung with or without meaningful lyrics to examine whether obligatory processing of semantic information, by virtue of competition for shared resources, impairs the ability to identify mistuned notes in a melody.

# Method

*Participants.* There were 24 adult participants (13 female; M = 20.0, SD = 1.8 years). More than half of the sample received at least 1 year of musical training (n = 13), and the average was positively skewed (M = 2.0, SD = 2.6, range = 0-8 years). Musical training information was missing for one participant. One additional participant was excluded for overall performance below chance.

*Procedure*. As noted above, half of the melodies were sung with lyrics and half sung with the syllable *la* on each note.

## Results and discussion

Figure 2(a) displays the primary results. A 2  $\times$  2 repeated-measures ANOVA (Table 1) comparing difference scores (hits minus false alarms) for different pitch-error magnitudes (50 cents, 100 cents) sung in the two styles (*la la*, lyrics) revealed a main effect for error magnitude, with better performance for larger errors (M=.477, SD=.217) than smaller errors (M=.188,



**Figure 2.** Pitch-Error Detection for Familiar Melodies Across All Experiments. Note: Plots display correct identification of errors (hits) minus false alarms (chance=0), separately for error magnitude (50 or 100 cents) and singing style. Both error magnitudes used the same trials as reference (i.e., shared false alarms). Within each experiment, all factors were within-participant. Across all experiments, 100-cent errors were more readily detected. Panel A shows better performance on trials with *la la* than trials with lyrics, with no interaction of error magnitude. Panel B shows better performance on autotuned *la* trials than autotuned trials with lyrics for 50-cent errors but not 100-cent errors. Higher performance relative to Experiment I is attributable to reduced pitch dynamics from autotuning. Panel C shows better performance on *la la* trials than scat trials, and no interaction with error magnitude. Results in Experiment 3 were similar to Experiment 1, but performance was higher overall. Panel D replicates the effects of Experiments I and 3 in a single experiment, but only for 50-cent errors. There was no difference between singing with lyrics or scat, implying no effect of semantics. Boxplots visualize median (line), 25–75 percentiles (hinges), and 1.5 \* interquartile range beyond hinge (whiskers). Individual data are visualized as semi-translucent points jittered horizontally.

SD = .103). Listeners more readily detected errors in melodies sung with la (M = .392, SD = .149) than with lyrics (M = .273, SD = .167). There was no interaction between singing style and pitch-error magnitude.

The results revealed that singing with lyrics impairs the ability to detect pitch errors in familiar melodies relative to singing with repeated syllables, with an overall effect size of 0.74

Predictor	df	F	р	$\eta_p^2$
Singing style	1,23	17.25	<.001***	.43
Magnitude	1,23	57.01	<.001***	.71
Singing style $ imes$ Magnitude	1,23	0.06	.812	<.01

Table 1. ANOVA Results for 2 (Singing Style: *la la*/Lyrics)  $\times$  2 (Magnitude: 50/100 Cents).

Note: \*\*\*p < .001, \*\*p < .01, \*p < .05.

(Cohen's *d*). This result did not interact with error magnitude, note placement, direction of pitch error, or duration (see Supplementary Materials online), even though several of those features showed a clear influence on pitch-error detection (magnitude, placement, duration). However, vocal music is inherently dynamic in pitch, and singing with lyrics could interact with pitch dynamics in unforeseen ways, for example, by the nature of pitch scoops that begin or end notes (Larrouy-Maestri & Pfordresher, 2018) or by pitch vibrato (Sundberg, 2013), which could be heightened in singing with lyrics. The second experiment examined this possibility by equalizing pitch variability across singing styles.

# **Experiment 2**

Here, we asked whether the impairment of pitch-error detection in songs with lyrics would persist if within-note pitch fluctuations were removed. As in Experiment 1, pitch-error detection in singing with lyrics was compared to pitch-error detection in singing on a repeated syllable (*la*), but with all within-note pitch fluctuations removed. The resulting notes sounded unnaturally level in pitch, as with autotuned melodies, but the performances were recognizably vocal and the lyrics clearly audible. We expected singing with lyrics to impair pitch-error detection because of interference from semantic processing and spectral variability.

## Method

*Participants.* There were 29 adult participants (22 female; M = 20.0, SD = 2.6 years), most of whom had at least 1 year of musical training (n = 22). Overall, years of musical training were similar to participants in Experiment 1 (M = 2.0, SD = 2.4, range = 0-11 years).

*Procedure.* We used the same melodies and singing styles as in Experiment 1 (*la la* vs. lyrics), except that all notes were "autotuned" in Melodyne using the pitch-modulation tool (i.e., pitch modulation reduced to 0%), which removes natural fluctuations of pitch within each note. All other aspects of the procedure were identical.

## Results and discussion

Figure 2(b) displays the primary results. Difference scores for detection of pitch errors (hits minus false alarms) were analyzed in a 2 × 2 repeated-measures ANOVA (Table 2) according to error magnitude (50, 100 cents) and singing style (autotuned *la la*, autotuned lyrics). As expected, there was a main effect for error magnitude, with better performance for 100-cent errors (M=.716, SD=.175) than 50-cent errors (M=.590, SD=.214). In addition, there was

Predictor	df	F	р	$\eta_p^2$
Singing style	1,28	6.36	.018*	.19
Magnitude	1,28	32.64	<.001***	.54
Singing style $\times$ Magnitude	1,28	5.80	.023*	.17

**Table 2.** ANOVA Results for 2 (Singing Style: Autotuned *la la*/Autotuned Lyrics)  $\times$  2 (Magnitude: 50/100 Cents).

Note: \*\*\*p < .001, \*\*p < .01, \*p < .05.

a main effect of singing style, with better performance for autotuned *la la* melodies (M=.678, SD=.195) than for autotuned melodies with lyrics (M=.628, SD=.192). Unexpectedly, there was an interaction between pitch-error magnitude and singing style. Separate paired-samples *t*-tests for each magnitude (Bonferroni–Holm) showed that singing type had a significant effect on 50-cent errors, t(28)=2.99, p=.011, d=.37, but not on 100-cent errors, p>.4.

Singing with lyrics impaired the ability to detect pitch errors in familiar melodies even when pitch dynamics were reduced, with an overall effect size of d=.26. As in Experiment 1, this result did not interact with target-note location or pitch-error direction, but it interacted with note duration (see Supplementary Materials online). Moreover, the deficit was limited to the smaller pitch deviations (50 cents). These results occurred in the context of considerably better overall performance (M=.657, SD=.185) relative to Experiment 1 (M=.333, SD=.142), perhaps because performance approached ceiling on trials with larger pitch deviations, reducing the main effect size of singing style. In any case, singing with lyrics impaired performance even in the absence of differences in within-note pitch dynamics.

In principle, poorer performance on songs sung with lyrics in Experiments 1 and 2 could stem from obligatory processing of semantic information, with adverse consequences for pitch perception. Alternatively, poorer performance could result from the spectral changes that accompany phonetic changes. The subsequent experiment explored this possibility by means of meaningless stimuli with variable phonetic content.

## **Experiment 3**

In the present experiment, listeners heard the same short melodies sung without meaningful lyrics. Half of the melodies, as in Experiment 1, were sung with repeating *la*, while the other half were sung in a somewhat more variable *scat* style with alternating syllables (see Figure 1). Both singing styles lacked semantic information, enabling us to ask whether the results of Experiments 1 and 2 were attributable to the acoustic dynamics of singing with lyrics.

#### Method

*Participants.* There were 25 adult participants (16 female; M = 20.8, SD = 2.4 years), most of whom had at least 1 year of musical training (n = 24), positively skewed (M = 4.8, SD = 4.0, range = 0.5-16 years).

*Procedure.* We compared melodies sung with repeating *la* to those sung with alternating *scat* or nonsense syllables (*doo bah dee bah*). All other aspects of the procedure were identical to Experiments 1 and 2.

				2
Predictor	df	F	р	$\eta_p^2$
Singing style	1,24	15.53	<.001***	.39
Magnitude	1,24	128.67	<.001***	.84
Singing style $ imes$ Magnitude	1,24	0.64	.639	.03

Table 3. ANOVA Results for 2 (Singing Style: *la la*/Scat)  $\times$  2 (Magnitude: 50/100 Cents).

Note: \*\*\*p < .001, \*\*p < .01, \*p < .05.

#### Results and discussion

Figure 2(c) displays the primary results. As in Experiments 1 and 2, difference scores (hits minus false alarms) for detection of pitch errors were first analyzed in a 2 × 2 repeated-measures ANOVA (Table 3) according to their magnitude (50, 100 cents) and singing style (*la la*, scat). There was a main effect of error magnitude, with better performance for 100-cent errors (M=.608, SD=.188) than 50-cent errors (M=.291, SD=.221). In addition, there was a main effect of singing style, with better performance for melodies sung with the identical, repeating syllable *la* (M=.509, SD=.166) than with alternating *scat* syllables (M=.390, SD=.240). There was no interaction between pitch-error magnitude and singing style.

The results of Experiment 3 imply that variable phonetic information, even in the absence of semantic information, impedes the ability to detect pitch errors in familiar melodies. Again, this result did not interact with pitch-error magnitude, target-note location, direction of pitch deviation, or duration (see Supplementary Materials online), even as most of these features influenced pitch perception (magnitude, location, duration). The present experiments, taken together, offer no support for the notion that semantic processing of familiar song lyrics reduces pitcherror detection beyond the reduction attributable to changing phonetic information. Performance was poorer for singing with lyrics in Experiment 1 than for scat singing in Experiment 3 when directly comparing those trials in an independent-samples t-test, t(47) = 1.96, p = .056, d = .56. However, overall performance was lower in Experiment 1 than in Experiment 3, t(47) = 2.41, p = .020, d = .69, and there were roughly comparable effect sizes for singing with lyrics (d = .74) and scat singing (d=.52) relative to singing with repeated syllables within each experiment. A parsimonious interpretation of these results is that group differences across studies generated differences in overall performance, with singing with lyrics and scat singing having comparable effects. Indeed, years of musical training were somewhat higher in Experiment 3 than in Experiment 1, and years of training correlated with overall performance across both experiments, r(46) = .49, p < .001. Another possibility is that semantic information had task-wide effects, reducing performance on adjacent trials with repeated syllables in Experiment 1. A direct comparison of all three singing styles could resolve these issues.

# **Experiment 4**

In the present experiment, participants evaluated the three styles of singing (lyrics, scat, repeating *la*) to permit direct comparisons. Two between-participant conditions were also included to assess the influence of adjacent trials. In the mixed condition, singing style was assigned randomly from trial to trial, as in the previous experiments. In the blocked condition, stimuli were grouped according to style.

Predictor	df	F	р	$\eta_p^2$
Group	1,34	1.22	.278	.03
Singing style	2,68	3.26	.045*	.09
Magnitude	1,34	222.52	<.001***	.87
Group $ imes$ Singing style	2,68	1.82	.170	.05
Group  imes Magnitude	1,34	0.22	.639	.01
Singing style $\times$ Magnitude	2,68	5.57	.006**	.14
Three-way interaction	2,68	0.52	.599	.02

**Table 4.** ANOVA Results for 2 (Group: Mixed, Blocked Trials)  $\times$  3 (Singing Style: *la la*/Scat/Lyrics)  $\times$  2 (Magnitude: 50/100 Cents).

Note: \*\*\*p < .001, \*\*p < .01, \*p < .05.

## Method

*Participants.* There were 36 participants in the final sample (n=18 per condition). Inadvertently, demographic information gathered from participants was inaccessible, but all participants were recruited from the same population and in the same manner as the three other experiments. Two additional participants were excluded for performance below chance.

*Procedure.* We used the same melodies as in Experiments 1 and 3, but with three singing styles (lyrics, scat, *la la*) instead of two, resulting in 3 blocks of 64 trials. Two between-participants conditions compared the effect of singing style adjacency. In the mixed condition, singing style was randomized from trial to trial, as in the previous experiments. In other words, each block of trials contained all three styles of singing. In the blocked condition, stimuli were grouped according to style and each block contained one singing style. The order of blocks was counterbalanced across participants. All other aspects of the task were identical to the general procedure.

## Results and discussion

Figure 2(d) displays the primary results. Difference scores (hits minus false alarms) for detection of pitch errors were first analyzed in a  $2 \times 2 \times 3$  mixed-model ANOVA (Table 4) with a between-participants factor of blocking condition (mixed, blocked) and within-participants factors of error magnitude (50 cents, 100 cents) and singing style (*la la*, scat, lyrics). There was no main effect and there were no interactions related to the between-participants manipulation of blocking, indicating that the adjacency of trials with lyrics to trials in other singing styles did not influence performance. As in Experiments 1 and 3, there was a main effect of error magnitude, with better performance for 100-cent errors (M=.622, SD=.182) than 50-cent errors (M=.273, SD=.147). In addition, there was a main effect of singing style, with performance highest for *la la* singing (M=.491, SD=.201), intermediate for scat singing (M=.431, SD=.186), and poorest for singing with lyrics (M=.418, SD=.163). Unlike Experiments 1 and 3, however, these main effects were qualified by a significant interaction between singing style and error magnitude. In separate 2 × 3 repeated-measures ANOVAs for each error magnitude, there was no main effect of singing style on trials with 100-cent errors, F(2, 70)=6.36, p=.003,

 $\eta_p^2 = .15$ . For 50-cent errors, there was a significant difference between *la la* singing and scat singing, t(35) = 3.61, p = .003, d = 0.56, and between *la la* singing and singing with lyrics, t(35) = 2.67, p = .023, d = 0.58 (Bonferroni–Holm). Importantly, there was no difference what-soever between singing with lyrics (M = .234, SD = .179) and scat syllables (M = .234, SD = .193) for 50-cent errors. In short, singing with lyrics and scat syllables had comparable effect sizes relative to singing on a uniform syllable (*la*).

Experiment 4 replicated the effect of dynamic singing (lyrics, scat) relative to singing with identical repeating syllables observed in Experiments 1 and 3, but as in Experiment 2, only for smaller pitch deviations. Singing style did not interact with direction of the pitch error, but it did interact with note duration as a matter of degree, and once again, there were clear effects of error magnitude, placement in the melody, and duration (see Supplementary Materials online).

It is unclear why there was no effect of singing style on 100-cent deviations, but the overall main effect of singing style (i.e., collapsed across magnitudes) was significant and followed the pattern observed in the other experiments ( $la \ la > lyrics$ ,  $la \ la > scat$ ). More importantly, there was no difference between singing with lyrics and scat syllables, which suggests that semantic information impairs pitch-error detection in singing largely because of acoustic dynamics.

The previous experiments found that lyrics (Experiments 1 and 2) and scat singing (Experiment 3) impaired the ability to detect pitch errors relative to uniform syllables. Thus, the results of Experiment 4 provide a replication of that primary finding. Moreover, the findings clarify that differences in overall performance between those experiments (e.g., better performance in Experiment 1 [lyrics] than 3 [scat]) could be attributed to individual differences in samples rather than differences in processing meaningful versus meaningless syllable changes. In other words, there was no support for the hypothesis that semantic processing entails costs on pitch perception.

#### **General discussion**

In the present study, we investigated the impact of lyrics on the detection of mistuned notes in highly familiar songs. Across all experiments, performance was affected by pitch-error magnitude (100 > 50 cents), placement of the error (later > earlier), and duration of the target note (longer > shorter), but none of these features interacted consistently with singing style. Singing with lyrics impaired the detection of pitch errors in well-known songs relative to singing on a single, repeated syllable, even when naturally occurring, within-note pitch fluctuations were removed. Comparable reductions in pitch-error detection were evident when the songs were sung with alternating nonsense syllables (scat singing). Taken together, these findings indicate that our ability to detect errors in highly familiar tunes is compromised by the presence of lyrics but not because of interference from semantic processing. Our findings suggest, instead, that the source of difficulty arises from rapidly changing phonetic information and associated spectral dynamics.

Questions about integrated or independent perception of linguistic and musical information in song are related to larger issues involving shared processing resources for music and language (Patel, 2010; Peretz & Coltheart, 2003). Here, we found no evidence that semantic information interfered with pitch-error detection for highly familiar songs. The result is noteworthy for two reasons. First, the songs in the present study are highly overlearned for North American listeners, and they are typically sung with lyrics. Participants in the present study may have heard or even sung these songs hundreds or thousands of times during their lifetime. After such extensive exposure, semantic processing of the lyrics may become automatic or essentially irrelevant to pitch tracking, at least with regard to additional effects beyond phonetic processing. Second, these highly familiar melodies would evoke strong musical expectations with regard to the unfolding of pitches, including the sequence of intervals and their relation to the tonal center. Presumably listeners' knowledge of the pitch relations of these familiar songs is strengthened because of exposure to everyday renditions at various pitch levels. In addition, the melodies in the present set of songs are fairly simple (i.e., limited number of notes, familiar, monophonic). Accordingly, our findings do not rule out the possibility that semantic processing could interfere with pitch processing when melodies and lyrics are less familiar or more complex.

In principle, well-known lyrics reduce the linguistic processing demands on listeners relative to less familiar or semantically incongruous lyrics, and well-known melodies reduce the music processing demands relative to less familiar melodies. In other words, it takes extra effort to process unfamiliar materials. Other studies of songs with behavioral and neural (electroencephalography [EEG]) measures have introduced levels of complexity or incongruities in linguistic (phonology, semantics, syntax) and musical (melody, harmony) information (Besson et al., 1998; Bigand et al., 2001; Bonnel et al., 2001; Fedorenko et al., 2009; Koelsch et al., 2005; Slevc et al., 2009). Further demands of parsing unfamiliar, complex, or incongruous lyrics could differentiate singing with lyrics from scat singing, but those manipulations would reduce the ecological validity of the materials. Alternatively, one could manipulate the influence of familiarity of the lyrics on pitch perception by using novel, repeated materials, along the lines of the "speech-to-song" illusion (Deutsch et al., 2011).

The voice is unique among instruments in the degree to which its resonance characteristics—and hence its timbre—change rapidly during performance. Although timbre perception is secondary to pitch perception in most melodic contexts, listeners cannot readily ignore timbral information. Our finding that pitch perception is less accurate in the context of alternating syllables than repeating syllables is consistent with the demonstrated influence of vowel brightness or spectral centroid on perceived interval size (Russo et al., 2019). Systematic manipulations of pitch (FO) and brightness variance in synthesized tones result in symmetrical interference from one dimension to the other, at least when the variations are perceptually equivalent across dimensions (Allen & Oxenham, 2014). In the present study, spectral content was not manipulated systematically or controlled acoustically. Nevertheless, all materials were produced by the same amateur singer in a relatively neutral (i.e., non-expressive) manner, without embellishment or excessive vibrato, and there were no differences in within-note pitch variance or note duration across singing styles. In short, the current study indicates that timbre variation arising from natural singing that incorporates syllable variability reduces the accuracy of pitch perception.

Finally, songs provide an excellent opportunity for probing the overlap and independence of music and language systems (Patel, 2010; Peretz & Coltheart, 2003; Schön et al., 2005). Indeed, the neural processing of songs seems to be distinct from the neural processing of instrumental music or speech (Norman-Haignere et al., 2020). Moreover, the ubiquity of songs with lyrics within and across cultures underlines their importance as a topic worthy of continuing study, one that can yield important insights about music cognition.

#### Limitations

The current research was limited by a lack of information about listeners' listening habits (across studies), details on specific musical training, and demographics (Experiment 4), all of which may influence exposure to Western music and, in turn, sensitivity to pitch errors

in that system. Future studies could examine pitch deviations in an unfamiliar system (e.g., microtonal) or replicate the study in other cultures with locally familiar songs. The number of levels of pitch error (50 or 100 cents) was limited by the study design and manual creation of stimuli. These magnitudes were chosen for their relevance to equal temperament tuning (quartertone, semitone), and may not always coincide with listeners' categorization of "out-of-tune." Future research could use programmatically generated stimuli with a more continuous distribution of pitch errors to determine thresholds on an individual basis.

# Conclusion

In sung renditions of familiar melodies, a pitch error is more easily detected when the syllables are unchanging (*la la*) than when they change from note to note. This effect is observed to a similar degree in changing syllables that are semantically meaningful (lyrics) and in changing syllables that lack meaning (scat). We conclude that, for well-known songs at least, phonetic variability interferes with precise pitch tracking.

# Implications

The singing voice, presumably the original instrument, is prevalent with and without lyrics across cultures. The effect of singing style (changing or unchanging syllables) on pitch tracking has implications for pitch and melody perception in music in everyday life that incorporates vocal elements. The current findings are equally relevant for music listeners and music makers. Vocalists may assume that listeners are generous when evaluating the intonation of scat or lyrical renditions. Lyricists and songwriters may consider the implications of different singing styles in passages that contain expressive pitch or difficult-to-perform intervals. For students, pitch-based ear training may be more effortful for materials sung with dynamic rather than repeating syllables.

## Acknowledgements

We thank Susan Morris for musical performances, Bradley Marks for assistance in data collection, and Anne-Marie Bissonnette and Margot Charignon for assistance in figure preparation.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Natural Sciences and Engineering Research Council (SET) and Fonds de Recherche du Quebec: Nature et Technologies (MWW).

## ORCID iD

Michael W Weiss (D) https://orcid.org/0000-0001-5208-9957

## Supplemental material

Supplemental material for this article is available online.

## References

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *Journal of the Acoustical Society of America*, 135(3), 1371–1379. https://doi.org/10.1121/1.4863269

- Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, *9*(6), 494–498.
- Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, 81(1), B11–B20. https://doi.org/10.1016/ S0010-0277(01)00117-2
- Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer. http://www.praat.org/
- Bonnel, A.-M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, 63(7), 1201–1213. https://doi.org/10.3758/BF03194534
- Condit-Schultz, N., & Huron, D. (2015). Catching the lyrics. *Music Perception*, 32(5), 470–483. https:// doi.org/10.1525/mp.2015.32.5.470
- Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America*, 129(4), 2245–2252.
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, 37(1), 1–9. https://doi. org/10.3758/MC.37.1.1
- Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception*, 30(2), 147–159. https://doi.org/10.1525/mp.2012.30.2.147
- Johnson, R. B., Huron, D., & Collister, L. (2013). Music and lyrics interactions and their influence on recognition of sung words: An investigation of word frequency, rhyme, metric stress, vocal timbre, melisma, and repetition priming. *Empirical Musicology Review*, 9(1), 2–20. https://doi.org/10.18061/ emr.v9i1.3729
- 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
- Larrouy-Maestri, P., & Pfordresher, P. Q. (2018). Pitch perception in music: Do scoops matter? *Journal of Experimental Psychology: Human Perception and Performance*, 44(10), 1523–1541. https://doi.org/10.1037/xhp0000550
- Lebedeva, G. C., & Kuhl, P. K. (2010). Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behavior and Development*, 33(4), 419–430. https://doi.org/10.1016/j.infbeh.2010.04.006
- Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics*, 70(5), 841–852. https://doi.org/10.3758/PP.70.5.841
- Micheyl, C., Xiao, L., & Oxenham, A. J. (2012). Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level. *Hearing Research*, 292(1), 1–13. https://doi. org/10.1016/j.heares.2012.07.004
- Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., Schalk, G., & Kanwisher, N. (2020). Intracranial recordings from human auditory cortex reveal a neural population selective for song. *bioRxiv*. https://doi.org/10.1101/696161
- Patel, A. D. (2010). Music, language, and the brain. Oxford University Press.
- Pereira, A. I., & Rodrigues, H. (2019). The relationship between Portuguese children's use of singing voice and singing accuracy when singing with text and a neutral syllable. *Music Perception*, 36(5), 468–479. https://doi.org/10.1525/mp.2019.36.5.468
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. Nature Neuroscience, 6(7), 688–691. https://doi.org/10.1038/nn1083
- Russo, F. A., Vuvan, D. T., & Thompson, W. F. (2019). Vowel content influences relative pitch perception in vocal melodies. *Music Perception*, 37(1), 57–65. https://doi.org/10.1525/ mp.2019.37.1.57
- Schön, D., Gordon, R. L., & Besson, M. (2005). Musical and linguistic processing in song perception. Annals of the New York Academy of Sciences, 1060, 71–81. https://doi.org/10.1196/annals.1360.006
- Schubert, E. (2019). Which nonvocal musical instrument sounds like the human voice? An empirical investigation. *Empirical Studies of the Arts*, 37(1), 92–103. https://doi.org/10.1177/0276237418763657

- Schubert, E., & Wolfe, J. (2016). Voicelikeness of musical instruments: A literature review of acoustical, psychological and expressiveness perspectives. *Musicae Scientiae*, 20(2), 248–262. https://doi. org/10.1177/1029864916631393
- Simon, C. (1926). The variability of consecutive wavelengths in vocal and instrumental sounds. *Psychological Monographs*, *36*(1), 41–83. https://doi.org/10.1037/h0093223
- Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2), 374–381. https://doi.org/10.3758/16.2.374
- Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*, 236(3), 82–91. https://doi. org/10.1038/scientificamerican0377-82
- Sundberg, J. (1994). Perceptual aspects of singing. Journal of Voice, 8(2), 106–122. https://doi. org/10.1016/S0892-1997(05)80303-0
- Sundberg, J. (2013). Perception of singing. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 69–105). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00003-1
- Trainor, L. J., & Hannon, E. E. (2013). Musical development. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 423–497). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00011-0
- van Besouw, R. M., Brereton, J. S., & Howard, D. M. (2008). Range of tuning for tones with and without vibrato. *Music Perception*, *26*(2), 145–155. https://doi.org/10.1525/mp.2008.26.2.145
- Wolfe, J., Garnier, M., Henrich Bernardoni, N., & Smith, J. (2020). The mechanics and acoustics of the singing voice. In F. A. Russo, B. Ilari & A. J. Cohen (Eds.), *The Routledge companion to interdisciplinary studies in singing* (1st ed., pp. 64–78). Routledge. https://doi.org/10.4324/9781315163734-5