




# Roles of bilingualism and musicianship in resisting semantic or prosodic interference while recognizing emotion in sentences

Cassandra Neumann<sup>1,2</sup> , Anastasia Sares<sup>1,2</sup>, Erica Chelini<sup>1,2</sup> and Mickael Deroche<sup>1,2</sup>

## Research Article

This article has earned badge for transparent research practices: Open Data. For details see the Data Availability Statement.

**Cite this article:** Neumann, C., Sares, A., Chelini, E., & Deroche, M. (2023). Roles of bilingualism and musicianship in resisting semantic or prosodic interference while recognizing emotion in sentences. *Bilingualism: Language and Cognition*, 1–15. <https://doi.org/10.1017/S1366728923000573>

Received: 13 June 2022  
Revised: 16 June 2023  
Accepted: 27 June 2023

**Keywords:**  
bilingualism; musicianship; prosody;  
semantics; vocal emotion recognition

**Corresponding author:**  
Cassandra Neumann,  
Email: [cassandra.neumann@concordia.ca](mailto:cassandra.neumann@concordia.ca)

<sup>1</sup>Laboratory for Hearing and Cognition, Psychology Department, Concordia University, 7141 Sherbrooke St. West, Montreal, QC H4B 1R6, Canada and <sup>2</sup>Centre for Research on Brain, Language & Music, Montreal, Quebec, Canada

### Abstract

Listeners can use the way people speak (prosody) or what people say (semantics) to infer vocal emotions. It can be speculated that bilinguals and musicians can better use the former rather than the latter compared to monolinguals and non-musicians. However, the literature to date has offered mixed evidence for this prosodic bias. Bilinguals and musicians are also arguably known for their ability to ignore distractors and can outperform monolinguals and non-musicians when prosodic and semantic cues conflict. In two online experiments, 1041 young adults listened to sentences with either matching or mismatching semantic and prosodic cues to emotions. 526 participants were asked to identify the emotion using the prosody and 515 using the semantics. In both experiments, performance suffered when cues conflicted, and in such conflicts, musicians outperformed non-musicians among bilinguals, but not among monolinguals. This finding supports an increased ability of bilingual musicians to inhibit irrelevant information in speech.

## 1. Introduction


Prosody can communicate a speaker’s intent, attitude, or emotion with the use of acoustic variables such as pitch, intensity, and duration of speech segments (Botinis et al., 2001; Cutler et al., 1997; Lehiste, 1970). For example, anger is typically characterized by high pitch (often with a descending contour), high intensity levels, and a rapid and variable speech rate (Preti et al., 2016). Children begin to recognize prosody early in infancy (Friend, 2001; Mastropieri & Turkewitz, 1999), which has been linked to better social development, communication skills, and empathy (Baron-Cohen et al., 1985). However, it is not until age 5 that children begin to consistently label a speaker’s emotional state using their tone of voice (Aguert et al., 2010, 2013; Sauter et al., 2013).

In daily conversation, the emotional prosody of speech can sometimes conflict with the semantic context (or the choice of words used to portray an emotion). In such cases, the understanding of emotional prosody is vital for understanding the true message of speech. For example, the utterance “What a great day” has positive or happy semantic content. However, if said in a sarcastic tone of voice, it would indicate the speaker’s discontent. Thus far, the literature shows that when presented with incongruent semantic and prosodic cues to emotions in spoken sentences and specifically asked to use prosody, 4-year-old children will make judgements about a speaker’s emotions based on semantic cues (Friend, 2000; Friend & Bryant, 2017; Morton & Trehub, 2001). However, at 10 years of age, children begin to grant more weight to prosodic (over semantic) cues in such situations (Morton & Trehub, 2001). Thus, at 10 years old there is a shift in the salience of semantic and paralinguistic cues to vocal emotions, moving toward a more adult-like ability. This is surprising given that prosody can be recognized very early in life, before children learn to speak and understand the semantic context of speech (Friend, 2001). One possible interpretation is that young children have a rudimentary understanding of the communicative role of vocal emotions and therefore grant more weight to semantic cues which may be utilized more easily for communication (Morton & Trehub, 2001). This raises the question of how the progressive mastery in a language along with general maturation effects eventually offset the balance between the use of semantic and prosodic cues to emotions.

While children are developing the ability to recognize speech prosody, many are also being exposed to a second language (Grosjean, 2010). Being bilingual or multilingual comes with

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



 This article has earned badge for transparent research practices: Open Data. For details see the Data Availability Statement.

several advantages. As bilinguals must learn when to use each language depending on the context, they are constantly making linguistic decisions, and may be better at handling conflicting demands. In addition to being able to communicate with more people, research has shown that being bilingual may contribute to better metalinguistic awareness, as well as advantages in several executive functions including inhibition, monitoring, and working memory (Adesope et al., 2010; Bialystok, 2015; Bialystok & Craik, 2010; Christoffels et al., 2013; Kroll & Bialystok, 2013; Yow & Markman, 2015, 2016). Executive functions are defined as high level cognitive control abilities that are involved in mental activity (Lehtonen et al., 2018) that arguably involve the pre-frontal and parietal regions of the brain (Chung et al., 2014; Kang et al., 2022). Inhibitory control, or the ability to selectively attend to relevant information while ignoring irrelevant information, is one executive function of particular interest in the current study, as participants are required to attend to one cue while inhibiting another (Bialystok et al., 2005). However, the literature on bilingualism has not always shown an advantage in executive functions, such as working memory, conflict monitoring, and inhibitory control, among bilinguals (Lehtonen et al., 2018; Paap, 2019; Paap & Greenberg, 2013). There are also deficits seen in bilinguals' linguistic abilities across the lifespan (Bailey et al., 2020; Bialystok & Craik, 2010; see meta-analysis by Donnelly et al., 2019). Furthermore, previous research has failed to find a bilingual advantage or found a bilingual disadvantage in some cognitive tasks including metacognitive processing, attentional control, inhibitory control, etc. (e.g., see Folke et al., 2016; Paap & Greenberg, 2013; Paap et al., 2015, 2017, 2018). Thus, this body of knowledge does not point to a clear cognitive or executive functioning advantage for bilinguals, but perhaps specific advantages in prosody perception or reliance on prosody.

It is well established that both languages known by a bilingual are activated in all contexts, requiring selection mechanisms to attend to the appropriate language in a given listening situation (Bialystok, 2017; Lehtonen et al., 2018). Thus, bilinguals are constantly inhibiting one of their competing languages, while actively using the other. To reduce mental load, bilingual children may progressively learn to use a cue that is more consistent across individuals and languages: prosody. In a study where 4 year old children were presented with sentences with conflicting prosodic and semantic cues, bilinguals showed an earlier ability to use prosodic cues than monolinguals (Yow & Markman, 2011). These results may be interpreted as a bilingual advantage in executive functioning (Bialystok, 1999; Costa et al., 2008; Kovács & Mehler, 2009) but could also be interpreted as a prosodic bias. For example, in Champoux-Larsson and Dylman's (2019) study, when asked to identify the emotion in the content (i.e., semantics) of words while ignoring prosody, bilingual children made more mistakes than monolingual children. When asked to identify the emotion in the prosody while ignoring the content, bilingual children made fewer mistakes than monolingual children and this difference increased both with age and with increased bilingual experience. Thus, bilingual 6–9-year-olds demonstrated a prosodic bias whereby they used prosodic cues to detect vocal emotions, even when prosody was the distractor. To delve further into the role of language exposure and the nature of the exposure received, some researchers have used questionnaires, such as the LEAP-Q (Marian et al., 2007), but in Champoux-Larsson and Dylman's (2019) study they used the Language and Social Background Questionnaire (LSQ; Luk & Bialystok, 2013) to rate participants (from 1: exposed to and used only one language to 4: equal

exposure to and use of two languages) and classify them as bilingual or monolingual. Thus, these differing methods of measuring bilingualism in the field may explain the lack of consistency in finding a bilingual advantage. While these studies focus on child populations, more recently, Champoux-Larsson and Dylman (2021) found that this prosodic bias may continue into adulthood, but only under some experimental conditions. However, one factor not considered in these studies was the effect of musical training.

Musical training has been shown to have many cognitive benefits. For example, musicians have better working memory (George & Coch, 2011) and executive functions (D'Souza et al., 2018; Moradzadeh et al., 2015) than non-musicians. Musical training might also help the encoding of speech and more globally for processing language (Coffey et al., 2017; Patel, 2011; Shook et al., 2013; Tierney & Kraus, 2013; Tierney et al., 2013). This is not surprising given that music and language share many common features (Besson et al., 2011; Hausen et al., 2013; Peretz et al., 2015), including the communication of emotions (Paquette et al., 2018). Emotions can be recognized in music as their acoustic properties are similar to emotions depicted in speech (Juslin & Laukka, 2003). Musicians are also better than non-musicians at detecting pitch fluctuations in both music and language (Sares et al., 2018; Schön et al., 2004). The fact that these findings hold for both speech and music is promising as it suggests that this musician advantage (in perceiving the "musicality" of speech) may be robust to linguistic (and therefore semantic) influences in the speech materials. In real life, people convey emotions partly through prosody and partly through their choice of words, the latter being arguably more straightforward (Pell & Skorup, 2008; Shakuf et al., 2022). There is also some mixed evidence in the literature that suggests that musicians may be better at recognizing the emotional prosody in speech (Lima & Castro, 2011; Trimmer & Cuddy, 2008), depending on whether emotional intelligence is accounted for or not. To date, it is unknown whether adult musicians and non-musicians differ in their use of prosody versus semantics for emotion processing, but there could well be a musician advantage in parsing these cues when they conflict. Finally, it is important to note that in this field (just like in the bilingualism field), the definition of musicianship is one of the possible reasons why mixed results have been found. For example, George and Coch (2011) defined a musician as someone who has studied music for 9 or more years, began playing prior to age 10, continuously studies the same instrument, and actively studies music. D'Souza et al. (2018) defined a musician as someone who has at least 8 years of experience playing and performing music, began training around 7 years old, and practices regularly. Moradzadeh et al.'s (2015) musicians had an average of 12 years of formal musical training, 90% had music theory training, 83% had ear training, and on average rated themselves 3.25 or having "good" sight-reading ability on a 5-point scale (1 = "beginner" and 5 = "expert"). The advantages that musicians exhibit in different tasks could rely more on some of the aforementioned variables than others, which could explain the differing results.

To the best of our knowledge, very few research groups have looked at the individual and combined effects of bilingualism and musicianship on vocal emotion recognition abilities in a single study and in an adult sample. Bialystok and DePape (2009) used an auditory Stroop task, where listeners were instructed to attend to prosody or to semantics of single words (and not sentences) with an emotional meaning. They found that adult

musicians (monolingual) responded more quickly than bilinguals and monolinguals (both non-musicians) in the prosody task, but there was no group difference in the semantics task. Similarly, Graham and Lakshmanan (2018) largely replicated the same design but only included a prosody task. They found that adult musicians (monolingual) had reduced reaction times on incongruent trials and smaller cognitive costs compared to the bilinguals (non-musicians and non-tone second language) but did not differ from monolinguals (non-musicians) or tone language bilinguals. However, neither study looked at the combined effects of bilingualism and musicianship. This is a striking gap given that both factors could facilitate the recognition of emotional prosody: being both a bilingual and musician may have additive effects.

The current study addresses this gap using an orthogonal 2×2 design to examine the contribution of each factor (bilingualism and musicianship) and their possible interaction, in relying on prosody versus semantics (or vice-versa) when recognizing emotions in sentences. More specifically, in situations of conflict, we hypothesized that bilingual adults would either demonstrate a prosodic bias as seen in children (whereby they would outperform monolinguals when asked to use prosody but perform worse than monolinguals when asked to use semantics) or they would demonstrate an inhibitory control advantage making them more resistant than monolinguals to distractors in both tasks. For musicians, the idea of a prosodic bias received mixed evidence so we favored the inhibitory control idea: we hypothesized that musicians would outperform non-musicians both when asked to use prosodic cues and when asked to use semantic cues to emotions, as long as the cues conflicted with one another. To test this, we designed two separate studies to mirror each other, with participants either attending to prosody (Experiment 1) or semantics (Experiment 2) to report the emotion contained in sentences – a sort of emotional Stroop task. Note that these experiments were run for two independent sets of subjects to avoid: 1) the same participant switching between the two tasks and changing listening/communicative strategies; and 2) to not expose a participant to the same sentence twice (a within-subject design would have halved the number of trials per task with such a constraint).

## 2. Methods

### 2.1. Participants

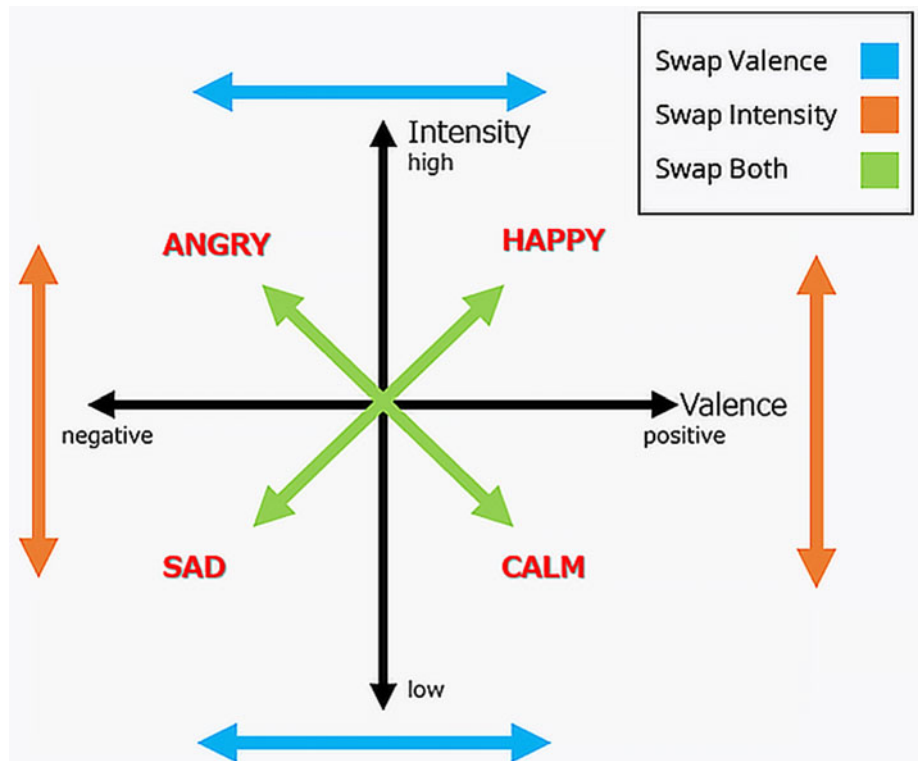
A total of 1086 participants across two experiments were recruited through Prolific (<https://prolific.co/>), an online recruitment platform. Recruitment was open only to specific English-dominant countries (Australia, Ireland, New Zealand, United Kingdom, and United States). Four separate batches were collected for each experiment (for a total of 8 batches): bilingual musicians, bilingual non-musicians, monolingual musicians, and monolingual non-musicians. The batches were based on filters for bilingualism and musicianship available in Prolific. For bilinguals, this meant answering “English” to the question “What is your first language?” (just like for monolinguals) and answering, “native +1 or native +2 other languages” to the question “Apart from your native language, do you speak any other languages fluently?”. For musicianship, this meant answering “Yes. For 5+ years.” to the question “Do you play a musical instrument, if so for how many years?” Forty-five participants either had technical difficulties (e.g., downloading the materials or browser issues) or did not complete their respective experiment and were thus excluded from the analyses. None of the participants had concerns about

their hearing, but two participants (0.19%) reported having mental health issues (still included). The final sample included 526 participants ( $N_{Females} = 271$ ,  $N_{Males} = 253$ , and  $N_{Prefer\ not\ to\ say} = 2$ ) in Experiment 1 and 515 participants ( $N_{Females} = 298$  and  $N_{Males} = 217$ ) in Experiment 2. All participants were between the ages of 18–41 years old (Experiment 1:  $M = 25.35$ ,  $SD = 5.94$ ; Experiment 2:  $M = 24.23$ ,  $SD = 4.95$ ).

Within our experimental interface, participants were asked about their language and musical background, and based on these answers (not their answers on the Prolific filters), they were divided into four groups: bilingual musicians ( $N_{Experiment\ 1} = 177$ ,  $N_{Experiment\ 2} = 171$ ) bilingual non-musicians ( $N_{Experiment\ 1} = 114$ ,  $N_{Experiment\ 2} = 101$ ), monolingual musicians ( $N_{Experiment\ 1} = 138$ ,  $N_{Experiment\ 2} = 144$ ), and monolingual non-musicians ( $N_{Experiment\ 1} = 97$ ,  $N_{Experiment\ 2} = 99$ ). Participants were asked “How many languages do you know in total?” and then required to give the name of each language. For each language entered, participants were then asked “At what age did you begin learning this language?”, “How proficient are you in this language?”, and “In the past year, how much have you used this language in daily life? 0 = Never, 10 = Exclusively.” The same questions were then asked, replacing the word *language* by *instrument*. The group classification was intentionally simple: monolinguals were participants who reported knowing only one language, English, while bilinguals reported knowing two or more languages (including English, their first language). Similarly, non-musicians were participants who did not play any musical instrument, while musicians reported playing one or more instruments. Note that this is not to deny the considerable variability within these groups. There is a notorious heterogeneity among bilinguals (e.g., de Bruin, 2019; Luk, 2015) and among musicians (Daly & Hall, 2018), so the information we recorded about their age of acquisition, proficiency, and use of each of their languages or musical instruments could allow us to probe further into the roles of bilingualism and musicianship. For example, it is known that early-trained musicians (before age 7) have behavioral benefits in auditory tasks (Bailey et al., 2020) and changes in cortical and sub-cortical networks compared to late-trained musicians (Penhune, 2019; Shenker et al., 2022; Vaquero et al., 2020). In our sample, roughly 30% reported learning their first instrument before age 7. On this basis, one might be tempted to narrow down our musician group definition by one demographic variable (and the same holds for bilingualism). However, given the dangers of dichotomizing such continuous variables (MacCallum et al., 2002) and the many possibilities for this dichotomization, we did not reclassify participants using arbitrary cut-offs (or manually dichotomize participant) from these metrics (i.e., age of acquisition, proficiency, or use). Instead, our variables were based on whether the participants reported a second language or not, or reported playing an instrument or not. In addition, we explored bilingualism and musicianship as continuous variables in regression approaches (see Figures S5 and S6 in Appendix S6).

### 2.2. Protocol

The participants recruited were bilinguals, monolinguals, musicians, and non-musicians interested in participating in an online study for compensation. Individuals interested in participating were redirected from Prolific to the experimental interface hosted on Pavlovia (an online platform for behavioural experiments), that was designed using the PsychoPy software (Peirce et al., 2019). All participants provided informed consent online in



**Figure 1.** Three different block types in the test phase. The blue arrows show a swap in valence, the orange arrows show a swap in intensity, and the green arrows show a swap in both intensity and valence.

accordance with the Institutional Review Board at Concordia University (ref: 30013650) and were compensated £3.90 for their participation.

Written instructions were given to explain the task. Participants were asked to adjust the volume of their device to a comfortable level before beginning a practice block. In each experiment, the practice block consisted of 16 trials of auditory stimuli, half of which were congruent (matching semantics and prosody), and the other half were incongruent (differing semantics and prosody). Participants were asked to attend to the prosody of each sentence for those recruited in Experiment 1, or the semantics for those recruited in Experiment 2, such that there was no confusion (or switch) in the goal of the task. After the presentation of each sentence, the participants were asked to click on the word of the emotion that was expressed out of four possible options displayed in four quadrants of the screen: angry (top-left), calm (bottom-right), happy (top-right), or sad (bottom-left). To pass the practice block, the participants had to obtain a minimum of 75% correct (12 out of 16 trials correct). If this was not obtained, participants continued repeating the practice block until 75% was attained. Feedback on performance was provided for practice trials but not for test trials. After completing the practice, participants moved on to the test phase.

In each experiment, the test phase consisted of 144 trials of auditory stimuli split across three blocks (48 trials per block). In each block, half of the trials (24) were congruent, and the other half were incongruent (24). Trials were equally divided into the four emotions: angry, calm, happy, or sad. Participants were presented with audio recordings of the sentences and asked to choose which emotion was expressed out of the four possible buttons (same quadrants as in the practice). Each of the three blocks differed in the way in which the semantic and prosodic cues to emotions were swapped in the incongruent trials (see Figure 1). In the *swap valence* block, the valence, or positive-negative dimension, of the emotions was

swapped (e.g., a semantically angry sentence enacted with a happy prosody). In the *swap intensity* block, the intensity, or high-low energy dimension, of the emotions was swapped (e.g., a semantically happy sentence enacted with a calm prosody). Finally, in the *swap both* block, both the intensity and valence of the emotions were swapped (e.g., a semantically angry sentence enacted with a calm prosody). The order in which these three blocks were presented was counterbalanced across participants.

The experiment took on average 25 minutes,  $SD = 11$ , to complete. The amount of time taken to complete the experiment did not differ by group,  $F(3,1033) = 0.86$ ,  $p = .462$ ,  $\eta^2 = 0.002$ , or by study,  $F(1,1033) = 0.49$ ,  $p = .483$ ,  $\eta^2 < .001$ , nor was there an interaction between the two,  $F(3,1033) = 1.29$ ,  $p = .277$ ,  $\eta^2 = 0.004$ .

### 2.3. Stimuli

All stimuli were created by the experimenters. They were produced and recorded by four speakers (2 males and 2 females) to generate variability and prevent listeners from learning speaker-specific manners of conveying emotions (either through their voice characteristics or speaking style). The list of 144 sentences can be found in Appendix S1 and contained 36 semantically angry sentences (e.g., “My sister gets on my nerves”), 36 semantically calm sentences (e.g., “Baths are relaxing”), 36 semantically happy sentences (e.g., “Let’s go to Disneyland”), and 36 semantically sad sentences (e.g., “His grandmother died”). These four emotions were selected as we wanted one emotion from each quadrant (see Figure 1) to have emotions of both positive and negative valence, and high and low intensity. This also allowed us to do the block type analyses (see Appendix S6). The speakers read each sentence with the prosody of all four emotions to create congruent and incongruent stimuli, resulting in 576 recordings from each speaker. Thus, the full set consisted of 2304 stimuli in total. Of these, 144 were randomly selected

for each participant, with no repetition of sentences. Each sentence was between 1.2 and 3.0 seconds long ( $M = 2.0$ ,  $SD = 0.3$ ).

We conducted an analysis on the semantics of each sentence using the *word2vec* algorithm to ensure that each sentence depicted its intended emotion (see Appendix S2 for more details on this analysis). This analysis confirmed that, overall, each set of sentences contained a semantic content that reflected the intended emotion. However, this was somewhat difficult to demonstrate and can perhaps be improved with more advanced packages (Raji & de Melo, 2020). Similarly, we conducted an analysis on the prosody of each sentence, demonstrating that emotions were enacted by the four speakers as expected: angry productions were particularly fast and dynamic in their intensity contours, while sad productions were slow and more stationary; happy productions were particularly high in pitch and well intonated, while sad and calm productions were low and more monotonous. In each metric, however, it is clear that speakers had their own style (see Appendix S3 for more detail) and were only partially consistent with one another in how they conveyed emotions.

#### 2.4. Equipment

Given that the present experiments took place online, we did not have rigorous control over the equipment and quality of sound. To address this limitation, we asked participants to indicate the audio device they were using (headphones, earbuds, external speakers, or default output from their PC/laptop), and asked them to rate the quality of their audio from 0-10 (where 0 is poor and 10 is excellent). There were no differences between groups in audio quality,  $F(3, 1037) = 0.22$ ,  $p = .881$ ,  $\eta^2 < .001$ ;  $M = 6.3$ ,  $SD = 0.86$ . There were also no differences between groups in the type of audio device used,  $\chi^2(9, N = 1041) = 16.52$ ,  $p = .057$ , with about 26% of participants listening through headphones, 19% through earbuds, 17% through speakers, and 38% through their default computer output.

#### 2.5. Analyses

##### Demographic analyses

Separate 2-by-2 ANOVAs (*musicianship* by *bilingualism*) were run on the combined data from both studies to analyze whether the groups differed in the language metrics collected (age of acquisition, proficiency, and use) for first language, second language (if applicable), and first instrument (if applicable). Additionally, Chi-Squared tests were used to compare other demographic variables, such as age, sex, employment status, and student status, between groups. Finally, all three metrics related to second language (L2; proficiency, use, and age of acquisition) were correlated with each other and all three metrics related to first instrument (I1; proficiency, use, and age of acquisition) were also correlated with each other.

##### Performance analyses

The measures of performance focused on sensitivity ( $d'$  values) and reaction times. Participants' responses (accuracy) were collapsed into confusion matrices, which were translated into hits and false alarm rates for each emotion. From these rates, we calculated  $d'$  ( $d'$ ) values for each participant, which were then used as the dependent variable in linear mixed effects models to examine the recognition of emotional prosody in Experiment 1 and the recognition of emotional semantics in Experiment

2. There were two between-subject fixed factors *musicianship* and *bilingualism*, where participants were either classified as a musician or as a non-musician and classified as a bilingual or as a monolingual. Finally, there was a within-subject fixed factor of *trial type* (incongruent or congruent condition). These models always contained random intercepts by subject, and random intercepts by emotion. Chi-square tests were conducted, after each fixed term was progressively added to the model to evaluate main effects and interactions. The analyses were run separately for Experiment 1 and Experiment 2. Scores were also analyzed on a trial-by-trial basis (using logistic regressions; see Appendix S4) and the findings were consistent with the main analysis.

In the aforementioned analysis, the type of incongruity was ignored (i.e., block type was not considered). However, we designed this experiment such that the emotions portrayed by the semantic and prosodic cues were swapped in a particular fashion in each block: valence-based, intensity-based, or both valence and intensity (see section 2.2 and Figure 1). To examine this factor,  $d'$  values by block type were also used as the dependent variable in linear mixed effects models to examine the differences in performance by block type and group allocation averaged across the four emotions. For simplicity (i.e., to avoid complex 4-way interactions), we used the interference effect in  $d'$  units (congruent-incongruent) as the dependent variable, with *musicianship*, *bilingualism*, and *block type* (swap valence, swap intensity, or swap both) as fixed factors. This model contained random intercepts by subject. See Appendix S6 for the results and discussion of the block type results.

Finally, the logarithm of the reaction time was used as the dependent variable in linear mixed effects models to examine how quickly participants responded as a function of *trial type*, *musicianship*, and *bilingualism* as fixed factors. This model again contained random intercepts by subject, and random intercepts by emotion. Each model was run using the *lme4* package in *r* (Bates et al., 2015) and was run separately for Experiment 1 and Experiment 2. The *emmeans* package in *r* (Lenth, 2023) was used for all post-hoc comparisons with Tukey's HSD adjustment to control for the inflation of Type I error in multiple comparisons.

### 3. Results

#### 3.1. Demographics

First, we present the demographic data for the total sample (combining Experiments 1 and 2). These are not the main results of the present study – however, given our large sample size, they are valuable in that they may be generalizable to bilinguals and musicians overall (or at least those that can be found online). The means and standard deviations for each language variable (proficiency, use, and age of acquisition of first and second languages) and instrument variables (proficiency, use, and age of acquisition of first instrument) by experiment are presented in Table 1.

##### Group differences in language and instrument variables

Although all participants had English as first language (L1), an interesting observation was a main effect of bilingualism on the age of acquisition of L1,  $F(1, 1037) = 16.65$ ,  $p < .001$ ,  $\eta^2 = 0.016$ , where bilinguals learned it 0.38 years later than monolinguals,  $SE = 0.092$ ,  $p < .001$  (see Table 1). This might seem surprising, but perhaps point to a different (more nuanced) understanding of what age of acquisition means for bilinguals than for monolinguals. On the other hand, there was no main effect of *musicianship*,

**Table 1.** The means and standard deviations for each language and instrument variable by experiment

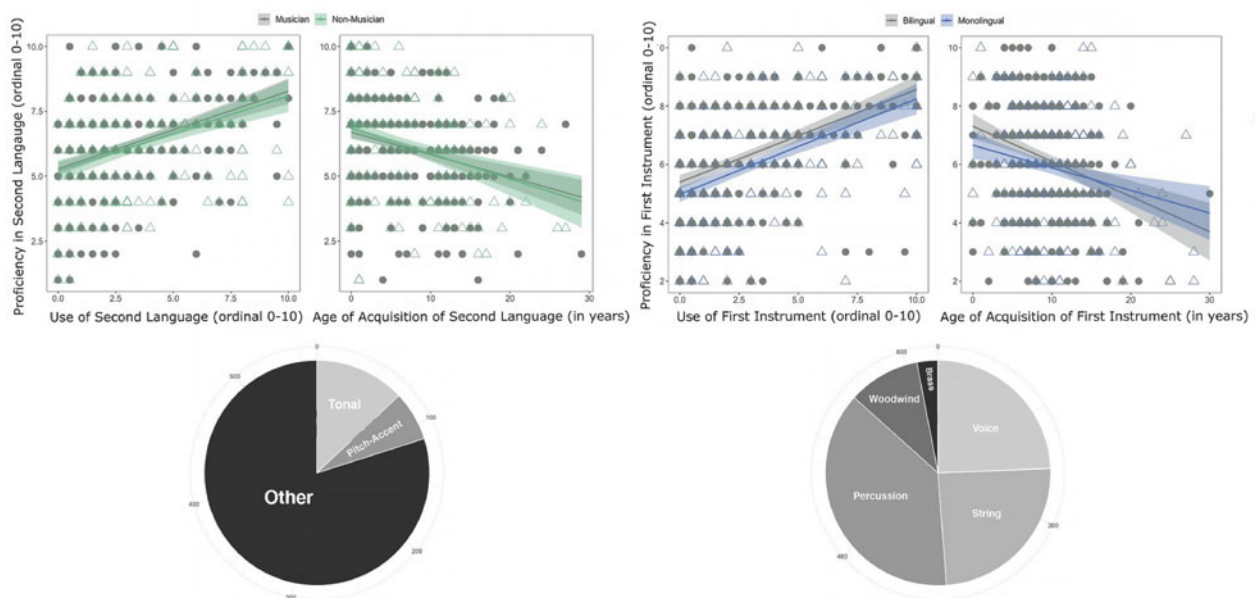
Groups	n	Mean Proficiency (SD)			Mean Use (SD)			Mean Age of Acquisition (SD)		
		L1	L2	I1	L1	L2	I1	L1	L2	I1
Experiment 1										
Monolingual Musicians	138	9.52 (0.76)	–	5.99 (1.79)	9.93 (0.25)	–	2.84 (2.75)	0.31 (0.81)	–	9.57 (4.64)
Monolingual Non-musicians	97	9.71 (0.59)	–	–	9.96 (0.18)	–	–	0.34 (0.73)	–	–
Bilingual Musicians	177	9.81 (0.60)	6.18 (1.82)	6.16 (1.90)	9.36 (0.92)	2.67 (2.59)	2.63 (2.79)	0.48 (1.31)	5.73 (5.87)	8.47 (4.56)
Bilingual Non-Musicians	114	9.56 (0.91)	6.40 (2.01)	–	9.24 (1.35)	3.79 (2.79)	–	1.06 (2.25)	6.13 (6.44)	–
Experiment 2										
Monolingual Musicians	144	9.58 (0.68)	–	5.84 (1.83)	9.95 (0.19)	–	2.88 (2.69)	0.30 (0.85)	–	9.65 (4.86)
Monolingual Non-musicians	99	9.56 (1.00)	–	–	9.92 (0.24)	–	–	0.33 (0.77)	–	–
Bilingual Musicians	171	9.70 (0.67)	6.12 (1.75)	6.40 (1.80)	9.42 (0.90)	3.11 (2.56)	2.96 (2.85)	0.64 (2.22)	6.99 (6.63)	8.67 (3.78)
Bilingual Non-Musicians	101	9.66 (1.10)	6.10 (2.06)	–	9.49 (0.92)	3.51 (2.83)	–	0.57 (1.20)	5.89 (6.13)	–

Note: L1 = first language; L2 = second language; I1 = first instrument

$F(1, 1037) = 2.72, p = .099, \eta^2 = 0.003$ , and no interaction,  $F(1, 1037) = 1.71, p = .192, \eta^2 = 0.002$ , on the age of acquisition of L1. For proficiency in L1, there was no main effect of *bilingualism*,  $F(1, 1037) = 3.35, p = .067, \eta^2 = 0.003$ , no main effect of *musicianship*,  $F(1, 1037) = 0.49, p = .483, \eta^2 < .001$ , but surprisingly there was an interaction,  $F(1, 1037) = 5.29, p = .022, \eta^2 = 0.005$ . Bilingual musicians rated themselves as more proficient in their L1 than monolingual musicians,  $M_{Difference} = 0.21, SE = 0.063, p = .006$ , while no other group comparison reached significance ( $p$  ranges from .124-.990). For use of L1, there was expectedly a main effect of *bilingualism*,  $F(1, 1037) = 139.72, p < .001, \eta^2 = 0.12$ , where monolinguals used their first language more than

bilinguals,  $M_{Difference} = 0.57, SE = 0.048, p < .001$ , but no main effect of *musicianship*,  $F(1, 1037) = 0.087, p = .77, \eta^2 = 0.000074$ , and no interaction,  $F(1, 1037) = 0.18, p = .676, \eta^2 < .001$ .

As demonstrated in **Figure 2** (top two panels) there is diversity in participants' L2s and musical instruments. For their L2, bilingual musicians and bilingual non-musicians did not differ in age of acquisition,  $F(1, 561) = 0.38, p = .54, \eta^2 = 0.00067$ , or proficiency,  $F(1, 561) = 0.46, p = .50, \eta^2 = 0.00082$ , but they did differ in their use,  $F(3, 561) = 11.13, p < .001, \eta^2 = 0.0019$ , as bilingual non-musicians used their L2 more often than bilingual musicians,  $M_{Difference} = 0.77, SE = 0.23, p < .001$ . Here again, this finding is not intuitive, and it is not clear whether it is a peculiarity of



**Figure 2.** Demographic Data. Top left: Correlations between proficiency and use, or proficiency and age of acquisition for their second language. Top right: Correlations between proficiency and use, or proficiency and age of acquisition of their first instrument. Bottom left: Pie chart of types of second languages. Bottom right: Pie chart of classes of first instruments.

our online samples or whether this could reflect a generalizable tendency. Previous research has shown that musical training positively impacts second language proficiency (see review by Zeromskaite, 2014; Slevc & Miyake, 2006), so it is rather puzzling as to why it would have an opposite effect for use (especially knowing how correlated proficiency and use are). Perhaps, this relates to the amount of free time: by engaging in extracurricular activities, musicians may have simply less time to practice their L2 compared to non-musicians.

Next, we analyzed the two musician groups. Monolingual musicians acquired their first instrument 1.04 years ( $SE = 0.36$ ) later than bilingual musicians,  $F(1, 628) = 8.44, p = .004, \eta^2 = 0.0143$ . Additionally, bilingual musicians were more proficient in their first instrument than monolingual musicians,  $M_{Difference} = 0.36, SE = 0.15; F(1, 628) = 6.13, p = .014, \eta^2 = 0.010$ . However, they did not differ in use,  $F(1, 628) = 0.097, p = .76, \eta^2 = 0.00016$ . Thus, bilingual musicians learned their instrument earlier and were more proficient in it than monolingual musicians. Once again, we could not identify any similar finding in the literature, so it is unclear whether they are a peculiarity of our samples or truly generalizable, but they further support the need to cross-investigate both factors in demographic analyses. We surmise that they might reflect environmental factors (home support, culture, and diligence regarding musicianship) which were not captured here by any other variable.

#### Group differences on other demographic variables

There were no differences in sex between monolinguals (combining musicians and non-musicians) and bilinguals (combining musicians and non-musicians),  $\chi^2(1, N = 1039) = 2.736, p = .098$ , nor between musicians (combining bilinguals and monolinguals) and non-musicians (combining bilinguals and monolinguals),  $\chi^2(1, N = 1039) = 0.070, p = .791$ . There was no difference in employment status between musicians and non-musicians,  $\chi^2(1, N = 1017) = 0.38, p = .54$ , but more monolinguals (65%) were employed than bilinguals (54%),  $\chi^2(1, N = 1017) = 12.92, p < .001$ . A related finding was no difference in student status between musicians and non-musicians,  $\chi^2(1, N = 1027) = 1.38, p = .24$ , but more bilinguals (57%) were students than monolinguals (38%),  $\chi^2(1, N = 1027) = 37.04, p < .001$ . Finally, the language groups differed in age,  $F(1, 1036) = 34.70, p < .001, \eta^2 = 0.032$ , such that monolinguals were slightly older than bilinguals,  $M_{Difference} = 2.02$  years,  $SE = 0.34, p < .001$ . The music groups did not differ in age,  $F(1, 1036) = 1.62, p = .203, \eta^2 = 0.002$ , but there was an interaction between *musicianship* and *bilingualism* because the age difference between monolinguals (older) and bilinguals (younger) was slightly larger in musicians.

#### Language and instrument variable correlations

All three metrics related to L2 (proficiency, use, and age of acquisition) were correlated with each other with an  $R^2$  above .092,  $p < .001$  (see Figure 2 top left). These relationships held within bilingual musicians and within bilingual non-musicians,  $R^2$  above .075,  $p < .001$ . In contrast, only some of the first instrument (I1) metrics were correlated with each other (I1; proficiency, use, and age of acquisition). Proficiency was correlated with use and age of acquisition,  $R^2$  above .061,  $p < .001$  (see Figure 2 top right). These relationships held within bilingual musicians and within monolingual musicians,  $R^2$  above .041,  $p < .001$ . However, use and age of acquisition of first instrument were not correlated,  $R^2 = .0044, p = .100$ , and even though this link existed within bilingual musicians, it was weak,  $R^2 = .022, p = .005$ .

### 3.2. Experiment 1 – Performance in emotional prosody

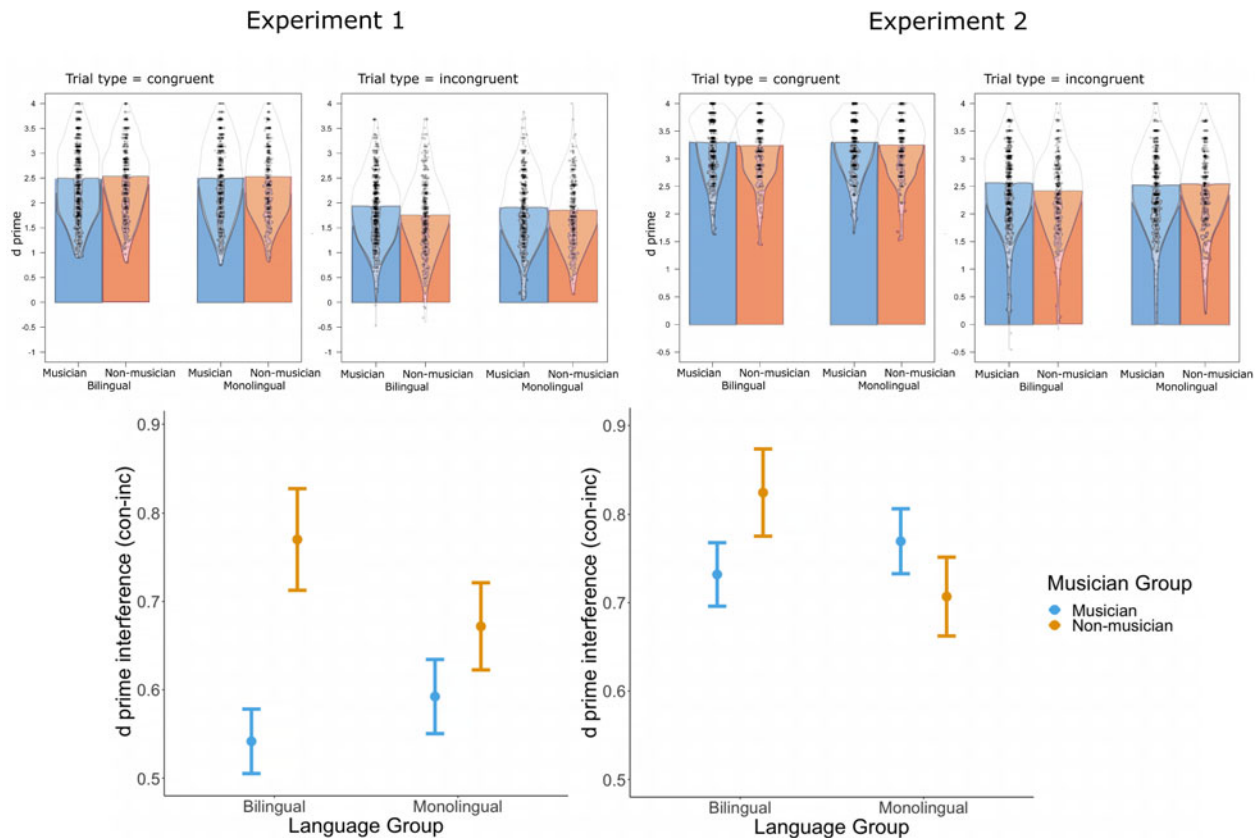
Figure 3 depicts the  $d'$  results of Experiments 1 and 2. As a reminder, for Experiment 1, participants were instructed to respond to the prosody (and ignore semantic cues). There was a main effect of *trial type*, confirming that  $d'$  decreased for incongruent stimuli compared to congruent stimuli, thus demonstrating that the task worked as participants found it challenging to completely ignore semantics (see Table 2 for all model results). There was no main effect of *bilingualism*, no main effect of *musicianship*, and no interaction between the two. There was a two-way interaction between *trial type* and *musicianship*, no interaction between *trial type* and *bilingualism*, and a three-way interaction.

Dissecting the three-way interaction between *trial type*, *bilingualism*, and *musicianship*, there were no differences in performance between any of the groups on the congruent trials,  $p$  always above .963; differences were only seen on the incongruent trials. This confirms the idea that the factors of interest (*bilingualism* and *musicianship*) acted upon the resistance to semantic interference (i.e., correctly attending to prosody), but not on basic emotion recognition. More precisely, there was a differential effect of *musicianship* among monolinguals compared to bilinguals in this resistance: bilingual musicians were better able to resist the semantic interference than bilingual non-musicians,  $p < .001$ , whereas *musicianship* had no effect among monolinguals,  $p = .948$ . On the other hand, there was no effect of *bilingualism* among non-musicians,  $p = .338$ , or among musicians,  $p = .993$ , suggesting that, controlling for musicianship, *bilingualism* had no role. To summarize, musicians were good at attending to prosody and could thus resist semantic interference compared to non-musicians, but this effect appeared to be driven by bilinguals. However, this interaction may be driven by the bilingual non-musicians exhibiting the poorest performance compared to all other groups.

### 3.3. Experiment 2 – Performance in emotional semantics

For Experiment 2, participants responded to semantics (and ignored prosodic cues). There was a main effect of *trial type*, confirming that  $d'$  decreased for incongruent stimuli compared to congruent stimuli, thus demonstrating that the task worked as participants found it challenging to completely ignore prosody (see Table 2 for all model results). There was no main effect of *bilingualism*, but the main effect of *musicianship* was statistically significant. Additionally, there was no interaction between *bilingualism* and *musicianship*. There was no interaction between *trial type* and *musicianship*, nor between *trial type* and *bilingualism*, but a significant 3-way interaction.

Mirroring Experiment 1, there were no differences between groups on the congruent trials,  $p$  is always above .915, but group differences on the incongruent trials, confirming the idea that the factors of interest (*bilingualism* and *musicianship*) acted upon the resistance to prosodic interference (i.e., correctly attending to semantics). More specifically, there was a differential effect of *musicianship* among bilinguals and not among monolinguals in this resistance. That is, bilingual musicians were better able to resist the prosodic interference than bilingual non-musicians,  $p = .0194$ , whereas *musicianship* had no role among monolinguals,  $p > 0.999$ . On the other hand, there was no effect of *bilingualism* among non-musicians,  $p = .179$ , or among musicians,  $p = .983$ . To summarize, musicians were also good at attending to semantics and could thus resist prosodic interference compared to non-musicians, but this effect appeared rather exclusive to



**Figure 3.**  $d'$  results.  $d'$  data by group and trial type for Experiment 1 (top left panel) and Experiment 2 (top right panel). Interaction between musicianship and bilingualism on the interference effect (congruent minus incongruent trials) expressed in  $d'$  units in Experiment 1 (bottom left panel) and Experiment 2 (bottom right panel), where lower  $d'$  units indicate better performance.

bilinguals. However, contrary to the first experiment, this interaction is less driven by the bilingual non-musician group.

### 3.4. Reaction Time

In both experiments, reaction time was delayed in incongruent compared to congruent trials (see Table A4.2 in Appendix S4.2). Specifically, it was about 3.00 versus 2.87 seconds in Experiment 1 and 3.24 versus 3.00 seconds in Experiment 2 (see Figure 4), but this 130-240ms difference was not sensitive to group allocation.

## 4. Discussion

The goal of the present study was to examine how bilinguals and musicians would recognize vocal emotions based on prosodic or semantic cues, compared to monolinguals and non-musicians. As intended, all groups showed a performance reduction accompanied by a delayed reaction time in incongruent compared to congruent trials. Consistent with the literature, we found a musician advantage in both experiments, whereby musicians were less prone to interference of the distracting cue (be it prosodic or semantic). However, this advantage was only found when also bilingual (i.e., in bilingual musicians). As for bilingualism on its own, we failed to observe a prosodic bias like the one seen in children (i.e., advantage in using prosodic cues and disadvantage in ignoring them) and we failed to see a bilingual advantage across the two tasks independent of musicianship. Furthermore, in

Experiment 1, the interaction seems to be driven by the poorest performance exhibited by bilingual non-musicians. Taken together, these results do not point to differences in cue weighting across these four groups rather differences in executive functioning among musicians and non-musicians, that somehow are exacerbated when also bilingual.

Regarding the protocol as an emotional Stroop task, it worked as expected and successfully created interference in processing in the incongruent trials. This was demonstrated by a reduction in accuracy of 10-20% and a delayed reaction time (by about 200 ms) in incongruent compared to congruent trials. Incongruent trials are of interest as they require listeners to pit two cues against each other, similar to situations of sarcastic speech encountered in everyday life. Previous studies on vocal emotion recognition have shown similar interference effects, where performance suffers and reaction times are delayed (Dupuis & Pichora-Fuller, 2010; Nygaard & Queen, 2008; Wurm et al., 2001). Our findings support the idea that experience with language and music can modulate the degree of confusion or challenges posed by this sort of ambiguous communicative modes.

### 4.1. Previous research in children

Based on findings in bilingual children, we had hypothesized that, even in adults, experience with multiple languages would influence the domain (prosody over semantics) primarily recruited in conflicting situations. Indeed, bilingual children begin using prosodic cues earlier than monolingual children (Yow &



**Table 2.** Model Results of the linear mixed effects models with  $d'$  as the dependent variable.

Fixed Effects	$\chi^2$	DF	$p$	95% CI	
				Lower	Upper
Experiment 1					
Intercept					
Trial Type	1387.5	1	<.001***	-0.59	-0.49
Bilingualism	0.49	1	.484	-0.057	0.096
Musicianship	3.14	1	.0766	-0.042	0.12
Trial Type x Bilingualism	0.039	1	.844	-0.13	0.027
Trial Type x Musicianship	26.7	1	<.001***	-0.31	-0.15
Bilingualism x Musicianship	1.40	1	.236	-0.13	0.11
Trial Type x Bilingualism x Musicianship	5.68	1	.0172*	0.026	0.27
Experiment 2					
Intercept					
Trial Type	2216.6	1	<.001***	-0.78	-0.69
Bilingualism	0.22	1	.639	-0.078	0.079
Musicianship	4.34	1	.0373*	-0.14	0.031
Trial Type x Bilingualism	0.67	1	.414	-0.11	0.031
Trial Type x Musicianship	0.42	1	.519	-0.17	-0.017
Bilingualism x Musicianship	2.28	1	.131	-0.12	0.14
Trial Type x Bilingualism x Musicianship	7.67	1	.00562**	0.045	0.27

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

Markman, 2011) and show a prosodic bias in situations where prosodic and semantic cues to emotions conflict (Champoux-Larsson & Dylman, 2019). The present findings did not replicate the same pattern, suggesting that in young adulthood, bilingualism alone does not lead to greater reliance on prosodic cues. We speculate that with greater cognitive maturation and language development, bilinguals can offset their early bias towards prosody and change their listening strategies to make an appropriate use of emotional cues in speech. However, the current results clearly highlight the importance of controlling for both language and musical experience in these types of designs.

#### 4.2. Previous research on the effects of bilingualism and musicianship individually

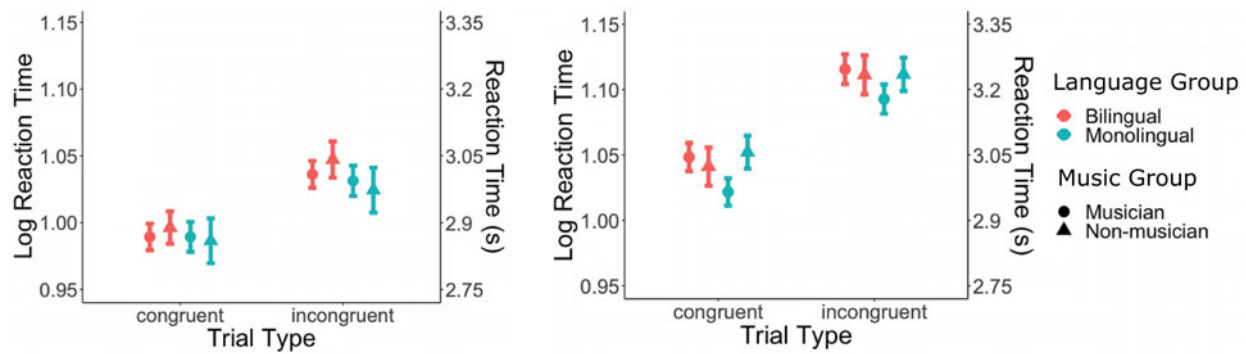
The present study accounts for both bilingualism and musicianship individually, as well as their combined effects. We added a group of bilingual musicians for a fully orthogonal sampling structure, which is rarely done in studies on vocal emotion recognition. This turned out to be critical as our findings generally support a musician advantage effect that is largely exaggerated among (if not exclusive to) bilinguals. Previous studies looking at each factor separately had revealed a musician advantage in a prosody task (Bialystok & DePape, 2009; Graham & Lakshmanan, 2018) but not in a semantics task (Bialystok & DePape, 2009). We think that this latter discrepancy may be due partly to the rudimentary nature of the semantic material used in these previous

studies (i.e., using the words “high” vs “low” and not emotionally loaded sentences). Thus, group differences could have been missed in the role of semantics for reasons related to task complexity. If this interpretation is correct, the musician advantage may be found in either domain (prosody or semantics) but would be easier to observe when placing participants in richer linguistic environments, which would surely have ecological relevance. Notably, however, the musician advantage effect that we observed among bilinguals was slightly smaller in the semantics task than in the prosody task. This difference is therefore going in a direction consistent with the contrast highlighted by Bialystok and DePape (2009). So, the nature of the task is important but perhaps more important is the need to control for language exposure among musicians and non-musicians.

One interesting avenue to try and make sense of the difference between bilingual musicians and non-musicians comes from insight gleaned from Schwartz and Kroll (2006). In cognitive tasks involving language, both languages are activated and influence performance even if the intention is to process information in one language only, a process referred to as non-selective lexical activation. Schwartz and Kroll (2006) found that non-selectivity is reduced when sentences provide richer semantic context, as if the brain was primed to navigate within a targeted language. In our study, Experiment 2 did guide participants towards semantics in general and could have limited this non-selectivity, but our first experiment would have done precisely the opposite. The fact that bilingual non-musicians (and not bilingual musicians) showed a particularly poor performance in Experiment 1 but less so in Experiment 2 suggest that they may be especially prone to non-selectivity. It follows that music training could mitigate the impact of non-selective lexical activation among bilinguals. Exactly how is unclear but perhaps by organizing parallel (rather than common) networks in each language separately. This further emphasizes the importance of controlling for both language exposure and musical experience in cognitive tasks involving language.

#### 4.3. Previous research on the combined effects of bilingualism and musicianship

In the few studies that did investigate bilingualism and musicianship simultaneously, findings are rather consistent with the present ones. Namely, it is musical training and not bilingualism that is more likely associated with benefits, specifically in task switching and dual-task performance tasks (Moradzadeh et al., 2015). Furthermore, Schroeder et al. (2016) disambiguated a “true” interference (by looking at a neutral condition minus incongruent trials) from a facilitation effect (congruent minus neutral trials), and Simon effects (congruent minus incongruent trials, as in the present study) but on a non-linguistic visual-spatial Simon Task in the same four groups. They found that bilingual musicians had a smaller Simon effect compared to all other groups, consistent with the present findings. However, bilingual musicians, bilingual non-musicians, and monolingual musicians had all smaller interference effects compared to monolingual non-musicians. There were no differences in facilitation effects once confounding variables such as IQ and age were accounted for. Their results suggest an enhanced ability to suppress interfering cues shared among bilinguals, musicians, and bilingual musicians, but they propose that the Simon effect (congruent minus incongruent) is a more convoluted metric encompassing both facilitation and interference effects making it



**Figure 4.** Reaction time results by trial type. Reaction time by trial type shown both with log reaction time and reaction time in seconds in Experiment 1 (top left) and Experiment 2 (top right) and by group in Experiment 1 (bottom left) and Experiment 2 (bottom right).

harder to interpret. In the present study, we did not include semantically neutral sentences or sentences spoken with a neutral prosody, so we are unable to disentangle these different effects. It would be interesting to see whether the unique advantage of the combined musician and bilingual profile taps more into the facilitation than the interference effect. It is important to note though that these studies did not focus on vocal emotion recognition, but rather executive functioning among these groups. However, based on the results of these studies, we could speculate that the present results may be due to better executive functioning among bilingual musicians.

#### 4.4. The role of executive functions

While we see differences in performance between groups, they do not reflect differences in cue weighting, but differences in executive functioning. A difference in cue weighting would have resulted in bilingual musicians outperforming the other groups on one task and performing worse on the other task. For example, if they weighted prosody more heavily, then bilingual musicians' performance would have been best when asked to use prosody to detect vocal emotions as they would easily ignore anything unrelated to prosody (i.e., the semantic meaning of the sentence). Additionally, their performance would have been worse when prosody served as a distractor because they would still rely on these salient prosodic cues that do not necessarily help in the task of deciphering the semantic content of the sentences. Since we never observed this sort of advantage/disadvantage reversal between the two tasks, we must interpret the results based on elements that were common across the two tasks, hence a general advantage in executive functioning when making judgements about vocal emotions. Bilingual musicians were able to use the correct cue regardless of the task and did not favour one listening strategy over another. This may reflect better response inhibition, cognitive control, or cognitive flexibility that have been previously shown to be advantages associated with being bilingual (Bialystok & Craik, 2010; Costa et al., 2008; Krizman et al., 2012; Wiseheart et al., 2016) or being a musician (Bialystok & DePape, 2009; Strong & Mast, 2019; Zuk et al., 2014). However, previous research has been somewhat inconclusive on whether bilingualism and musicianship have benefits that extend beyond the realm of language and music, respectively, into other executive functions. Neither bilingualism nor musical experience (D'Souza et al., 2018; Lehtonen et al., 2018) has been

unequivocally shown to facilitate executive functioning in adults. Based on the current results, we speculate that this might be partly because the other factor (bilingualism or musicianship) was not controlled for. Given their individual roles, it makes sense that the interaction between these two skills provides additional benefits in executive functioning in certain situations. In simpler terms, the effects may be additive. However, executive functioning was not specifically measured in the present study, so this idea is only one possible interpretation. An alternative interpretation is that the musician advantage in executive functioning transfers to the language domain more easily in bilinguals.

#### 4.5. Transfer effects

Overlap between music and language has been noted in acoustic properties (Besson et al., 2011; Hausen et al., 2013; Peretz et al., 2015) and the communication of emotions (Paquette et al., 2018). Also, there is substantial overlap in brain regions that process language and music (Fedorenko et al., 2009; Levitin, 2003; Maess et al., 2001; Patel & Iversen, 2007). So, one could speculate that the benefits of experience in one would transfer to the other. Cross-domain transfer effects have been reported from music to language (Besson et al., 2011; Bidelman et al., 2011; Moreno, 2009; Patel, 2011) and from language to music (Deroche et al., 2019a; Krishnan & Gandour, 2009), but the causality of music training – as opposed to inherent perceptual or cognitive aptitudes – is highly debated (Mankel & Bidelman, 2018; Penhune, 2019, and also McKay, 2021 for a review of this question within the hearing-impaired world). Patel (2011) argues that musical training leads to neuroplasticity in brain networks responsible for speech processing resulting in better encoding of several features of speech, but this would occur only under certain conditions. Specifically, music training must allow for precise processing and discrimination of auditory information in these networks, connect to emotional rewards, be associated with focused attention, and be repeated frequently, for such transfers to occur. These criteria are all common to language learning and good reasons why musical training may benefit the acquisition of a second language (Chobert & Besson, 2013). In sum, individuals who receive musical training and learn multiple languages might have a unique opportunity to develop neural networks that are critical to the encoding of certain aspects of speech (perhaps particularly affective cues) necessary to decode emotions in sentences. However, once again, this is only speculative and further

research needs to be done to better understand why such a transfer from music training to the language domain would not occur (or not as easily) in monolinguals.

#### 4.6. Emotional intelligence

There are other variables that may account for, or mediate, some of the current results: emotional intelligence. Not surprisingly, higher emotional intelligence has been linked to better recognition of emotions. Alqarni and Dewaele (2020) found that participants who have higher trait emotional intelligence (i.e., the construct that relies more on perception of one's own emotions) were better at perceiving and interpreting emotions from audio-visual recordings. Crucially, they found that bilinguals had higher trait emotional intelligence than monolinguals. However, the effect sizes for each of these results were small (Cohen's *d* of about 0.30). Furthermore, Trimmer and Cuddy (2008) found that emotional prosody discrimination was related to emotional intelligence scores but not musical training (contradicting other reports – see introduction). Also, musical training has not been linked to higher emotional intelligence (Glenn Schellenberg, 2011; Trimmer & Cuddy, 2008) and, to our knowledge, there are no studies on emotional intelligence in individuals who are both a musician and bilingual. Thus, if differences in emotional intelligence were a concern for this study, one might have expected it to enhance performance among bilinguals but not musicians, which was not what we observed. Also, we would expect this variable to affect performance on congruent trials as well, whereas group differences were exclusive to incongruent trials here. For these reasons, we suspect that it is unlikely to explain the current results.

#### 4.7. Socioeconomic status

We might equally wonder whether socioeconomic status (SES) could partially explain the results as SES is known to affect research on bilingualism particularly. Some studies have found SES to be a potential confound when assessing a bilingual advantage in the Simon Task (Morton & Harper, 2007), while others have controlled for SES and continued to find a bilingual advantage in inhibitory control (Emmorey et al., 2008; Filippi et al., 2022; Nair et al., 2017). So, this is an on-going debate. But perhaps most relevant here, Naeem et al. (2018) found that being a bilingual compared to a monolingual had no effect on performance (in the Simon task) among individuals with high SES, but bilinguals outperformed monolinguals (on both congruent and incongruent trials) among individuals with low SES. As musicians are likely to have higher SES than non-musicians (Swaminathan & Schellenberg, 2018), one could have expected (from a SES-based interpretation) bilingualism to have little role among musicians, but a beneficial role among non-musicians. Again, this is not what we observed, and as for emotional intelligence, such an interpretation would affect both trial types whereas our findings pointed specifically to the incongruent trials. Thus, the present findings do not align easily with an interpretation based on SES differences, though more research should be done to account for this variable.

#### 4.8. Limitations

Some limitations to the current study should be acknowledged. Given the nature of online studies: 1) there was generally a lack of control over the stimulus delivery as it was not administered

in a controlled environment, 2) the quality of bilinguals and musicians and the reliability of their self-reports could be questioned, and 3) the generalizability of online findings should be verified. Here we respond to each of these concerns.

In response to the first concern, we asked participants to rate the quality of their audio and did not find any group difference in this regard. Also, performance on congruent trials (including reaction times) was overall decent and similar to previous studies (e.g., see Bialystok & DePape, 2009; Champoux-Larsson & Dylman, 2019; Moradzadeh et al., 2015). Thus, poorer audio quality than in a lab or a general lack of interest and attention towards an online task would be unlikely to explain the group difference found in this study.

Second, we relied on participants' self-reports to allocate them as either a bilingual or monolingual, and as musician or non-musician. Tomoschuk et al. (2019) found that objective measures of language proficiency (e.g., picture naming or proficiency interviews) are better than self-rating of language proficiency, while other studies found self-report measures to be just as reliable as objective measures (Lim et al., 2008; Shameem, 1998). Thus, this concern is debatable – however, our analytical approach did not rely on precise estimates of age of acquisition, proficiency, and use, since we followed a categorical approach for the groups' definition. In other words, inaccuracies in self-reports would have had no consequence on our conclusion (but would have affected slightly the findings of Appendix S5 where continuous variables were used).

Third, the validity of online studies has been investigated in recent years. As outlined in the review by Chandler and Shapiro (2016), there are notable differences between the general population and online convenience samples. Several issues are relevant here, such as the realization that online samples tend to be younger than the general population and some samples may be either over- or under-represented (i.e., more participants tend to be Caucasian and Asian, and participants tend to be more educated). In this study specifically, we found that bilinguals were younger and more of them were students and unemployed compared to monolinguals. However, the extent to which these characteristics might be biased by being online samples is very difficult to assess. Of note, Eyal et al. (2021) found that the online platform Prolific (the one used here) provided higher quality data in terms of comprehension, attention, and dishonesty, than MTurk (the online platform used in Chandler & Shapiro, 2016). We also see certain advantages to conducting the present study online: having a very large sample size that could reflect the heterogeneity of musicians and bilinguals; and being able to easily recruit English-speaking monolinguals (a fairly difficult thing to do in-person in Québec). Thus, we believe that the benefits outweigh the disadvantages of using online platforms in some experimental designs, such as in the present study.

## 5. Conclusions and future directions

In conclusion, musical training appears to benefit the recognition of vocal emotions, either when semantic cues or when prosodic cues are providing conflicting information, but only among bilinguals. We did not see a difference in cue weighting (e.g. a prosodic bias) when identifying vocal emotions among the groups, as previously seen in bilingual and monolingual children. Instead, differences may be due to enhanced executive functioning in bilingual musicians that results in better performance in incongruent trials. We speculate that this is because the enhanced executive functions of musicians are somehow strengthened, or transfer more easily to

the language domain, in bilinguals than they do in monolinguals. This may be due to bilinguals being more flexible in their listening strategies or still figuring out the different ways to resolve conflictual situations of communicative intent.

This research has implications for educational and linguistic fields, but also for clinical areas such as in individuals growing up with degraded hearing. For example, school-aged children with cochlear implants or with hearing aids typically perform worse than their normal hearing counterparts on tasks of emotional prosody (Barrett et al., 2020; Chatterjee et al., 2015; Deroche et al., 2019b; Lin et al., 2022; Most & Peled, 2007). Deficits in these tasks are often linked to poor pitch perception, but it may well be that these children also develop alternative strategies to recognize emotions in sentences. Some of these strategies could involve a stronger reliance on semantics and weaker reliance on prosody, or a different weighting among prosodic cues (e.g., using temporal and intensity cues more than pitch cues). Thus, understanding the particular circumstances or participant profiles that result in enhanced vocal emotion recognition may be beneficial to understanding how to improve these abilities in hearing-impaired and cochlear implanted children and adults. Experiments are under way to run this exact paradigm in cochlear implant users.

**Acknowledgements.** We would like to thank all participants on Prolific who gave their time to complete this study. We also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada's (NSERC) Discovery Grant awarded to M.D. (ref: DGEGR-2020-00106), NSERC's Canada Graduate Scholarship awarded to C.N., les Fonds de recherche du Québec - Nature et technologies (FRQNT) Scholarship awarded to C.N. (#301964), and the Center for Research on Brain, Language, and Music (CRBLM) Scholarship awarded to C.N. The CRBLM is funded by the Government of Quebec via the Fonds de Recherche Nature et Technologies and Société et Culture.

**Competing interests.** The authors declare none.

**Supplementary material.** For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S1366728923000573>

**List of Supplementary material.** Appendix S1: Transcripts of all sentences  
List of sentence material used as the stimuli in the present study.  
Appendix S2: Confirming adequate semantics  
An analysis testing whether each sentence was semantically close to the intended emotion.  
Appendix S3: Confirming adequate prosody  
An analysis of the acoustic characteristics of each sentence to ensure that they contained the expected prosodic features of their intended emotion.  
Appendix S4: Trial by Trial analyses  
Logistic mixed effects models to analyse performance results and log reaction time by trial.  
Appendix S5: Bilingualism and Musicianship as continuous variables  
Linear mixed effects models to analyze performance results with bilingualism and musicianship as continuous variables (as opposed to categorical).  
Appendix S6: Block Type  
A more detailed description of the different block types used in the present study and an analysis of performance by block type in each experiment.

**Data availability.** The data that support the findings are openly available in OSF at [https://osf.io/nb2wv/?view\\_only=b101c8962aae41968a7465161b6b59ff](https://osf.io/nb2wv/?view_only=b101c8962aae41968a7465161b6b59ff).

## References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, 80(2), 207–245. <https://doi.org/10.3102/0034654310368803>

Aguert, M., Laval, V., Le Bigot, L., & Bernicot, J. (2010) Understanding Expressive Speech Acts: The Role of Prosody and Situational Context in French-Speaking 5-to 9-Year-Olds. *Journal of Speech Language and Hearing Research*, 53(6), 1629–1641. doi: 10.1044/1092-4388(2010/08-0078)

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Bigot, L. L. (2013). Inferring emotions from speech prosody: Not so easy at age five. *PLoS ONE*, 8(12). <https://doi.org/10.1371/journal.pone.0083657>

Alqarni, N., & Dewaele, J.-M. (2020). A bilingual emotional advantage? An investigation into the effects of psychological factors in emotion perception in Arabic and in English of Arabic-English bilinguals and Arabic/English monolinguals. *International Journal of Bilingualism*, 24(2), 141–158. <https://doi.org/10.1177/1367006918813597>

Bailey, C., Venta, A., & Langley, H. (2020). The bilingual [dis]advantage. *Language and Cognition*, 12(2), 225–281. <https://doi.org/10.1017/langcog.2019.43>

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37–46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)

Barrett, K. C., Chatterjee, M., Caldwell, M. T., Deroche, M. L. D., Jiradejvong, P., Kulkarni, A. M., & Limb, C. J. (2020). Perception of child-directed versus adult-directed emotional speech in pediatric cochlear implant users. *Ear and Hearing* 41(5), 1372–1382. <https://doi.org/10.1097/AUD.0000000000000862>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00094>

Bialystok, E. (1999). Cognitive Complexity and Attentional Control in the Bilingual Mind. *Child Development*, 70(3), 636–644. <https://doi.org/10.1111/1467-8624.00046>

Bialystok, E. (2015). Bilingualism and the Development of Executive Function: The Role of Attention. *Child Development Perspectives*, 9(2), 117–121. <https://doi.org/10.1111/cdep.12116>

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233–262. <https://doi.org/10.1037/bul0000099>

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, 19(1), 19–23. <https://doi.org/10.1177/0963721409358571>

Bialystok, E., & DePape, A. M. (2009). Musical Expertise, Bilingualism, and Executive Functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 565–574. <https://doi.org/10.1037/a0012735>

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, 9(1), 103–119. <https://doi.org/10.1177/13670069050090010701>

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, 77(1), 1–10. <https://doi.org/10.1016/j.bandc.2011.07.006>

Botinis, A., Granström, B., & Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Communication*, 33(4), 263–296. [https://doi.org/10.1016/S0167-6393\(00\)00060-1](https://doi.org/10.1016/S0167-6393(00)00060-1)

Champoux-Larsson, M. F., & Dylman, A. S. (2019). A prosodic bias, not an advantage, in bilinguals' interpretation of emotional prosody. *Bilingualism*, 22(2), 416–424. <https://doi.org/10.1017/S1366728918000640>

Champoux-Larsson, M.-F., & Dylman, A. S. (2021). Bilinguals' inference of emotions in ambiguous speech. *International Journal of Bilingualism*, 25(5), 1297–1310. <https://doi.org/10.1177/13670069211018847>

Chandler, J., & Shapiro, D. (2016). Conducting Clinical Research Using Crowdsourced Convenience Samples. *Annual Review of Clinical Psychology*, 12(1), 53–81. <https://doi.org/10.1146/annurev-clinpsy-021815-093623>

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research* 322, 151–162. <https://doi.org/10.1016/j.heares.2014.10.003>

- Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences*, 3(2), 923–940. <https://doi.org/10.3390/brainsci3020923>
- Christoffels, I. K., Kroll, J. F., & Bajo, M. T. (2013). Introduction to Bilingualism and Cognitive Control. *Frontiers in Psychology*, 4, 1664–1078. <https://doi.org/10.3389/fpsyg.2013.00199>
- Chung, H. J., Weyandt, L. L., & Swentosky, A. (2014). *The Physiology of Executive Functioning*. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (pp. 13–27). Springer New York. [https://doi.org/10.1007/978-1-4614-8106-5\\_2](https://doi.org/10.1007/978-1-4614-8106-5_2)
- Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>
- Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, 106(1), 59–86. <https://doi.org/10.1016/j.cognition.2006.12.013>
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, 40(2), 141–201. <https://doi.org/10.1177/002383099704000203>
- Daly, H. R., & Hall, M. D. (2018). Not all musicians are created equal: Statistical concerns regarding the categorization of participants. *Psychomusicology: Music, Mind, and Brain*, 28(2), 117–126. <https://doi.org/10.1037/pmu0000213>
- de Bruin, A. (2019). Not All Bilinguals Are the Same: A Call for More Detailed Assessments and Descriptions of Bilingual Experiences. *Behavioral Sciences*, 9(3), 33. <https://doi.org/10.3390/bs9030033>
- Deroche, M. L. D., Felezou, M., Paquette, S., Zeitouni, A., & Lehmann, A. (2019a). Neurophysiological Differences in Emotional Processing by Cochlear Implant Users, Extending Beyond the Realm of Speech: *Ear and Hearing*, 40(5), 1197–1209. <https://doi.org/10.1097/AUD.0000000000000701>
- Deroche, M. L. D., Lu, H.-P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S.-C., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2019b). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports*, 9(1), 109. <https://doi.org/10.1038/s41598-018-36393-1>
- Donnelly, S., Brooks, P. J., & Homer, B. D. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review*, 26(4), 1122–1147. <https://doi.org/10.3758/s13423-019-01567-z>
- D'Souza, A. A., Moradzadeh, L., & Wiseheart, M. (2018). Musical training, bilingualism, and executive function: Working memory and inhibitory control. *Cognitive Research: Principles and Implications*, 3(1). <https://doi.org/10.1186/s41235-018-0095-6>
- Dupuis, K., & Pichora-Fuller, M. K. (2010). Use of affective prosody by young and older adults. *Psychology and Aging*, 25(1), 16–29. <https://doi.org/10.1037/a0018777>
- Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The Source of Enhanced Cognitive Control in Bilinguals: Evidence From Bimodal Bilinguals. *Psychological Science*, 19(12), 1201–1206. <https://doi.org/10.1111/j.1467-9280.2008.02224.x>
- Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 24, 1643–1662. <https://doi.org/10.3758/s13428-021-01694-3>
- 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>
- Filippi, R., Ceccolini, A., Booth, E., Shen, C., Thomas, M. S. C., Toledano, M. B., & Dumontheil, I. (2022). Modulatory effects of SES and multilingual experience on cognitive development: A longitudinal data analysis of multilingual and monolingual adolescents from the SCAMP cohort. *International Journal of Bilingual Education and Bilingualism*, 25(9), 3489–3506. <https://doi.org/10.1080/13670050.2022.2064191>
- Folke, T., Ouzia, J., Bright, P., De Martino, B., & Filippi, R. (2016). A bilingual disadvantage in metacognitive processing. *Cognition*, 150, 119–132. <https://doi.org/10.1016/j.cognition.2016.02.008>
- Friend, M. (2000). Developmental changes in sensitivity to vocal paralinguistic. *Developmental Science*, 3(2), 148–162. <https://doi.org/10.1111/1467-7687.00108>
- Friend, M. (2001). The transition from affective to linguistic meaning. *First Language*, 21(63), 219–243. <https://doi.org/10.1177/014272370102106302>
- Friend, M., & Bryant, J. B. (2017). A Developmental Lexical Bias in the Interpretation of Discrepant Messages. *Merrill Palmer Q*, 42(2), 342–369.
- George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, 49(5), 1083–1094. <https://doi.org/10.1016/j.neuropsychologia.2011.02.001>
- Graham, R. E., & Lakshmanan, U. (2018). *Tunes and Tones: Music, Language, and Inhibitory Control*. 18, 104–123. <https://doi.org/10.1163/15685373-12340022>
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Harvard University Press. <https://doi.org/libezproxy.concordia.ca/10.4159/9780674056459>
- Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: A common rhythm. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00566>
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Kang, W., Hernández, S. P., Rahman, M. S., Voigt, K., & Malvaso, A. (2022). Inhibitory Control Development: A Network Neuroscience Perspective. *Frontiers in Psychology*, 13, 651547. <https://doi.org/10.3389/fpsyg.2022.651547>
- Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, 106(16), 6556–6560. <https://doi.org/10.1073/pnas.0811323106>
- Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, 110(3), 135–148. <https://doi.org/10.1016/j.bandl.2009.03.005>
- Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*, 109(20), 7877–7881. <https://doi.org/10.1073/pnas.1201575109>
- Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25(5), 497–514. <https://doi.org/10.1080/20445911.2013.799170>
- Lehiste, I. (1970). *Suprasegmentals*. Massachusetts Inst. of Technology P.
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, 144(4), 394–425. <https://doi.org/10.1037/bul0000142>
- Lenth, R. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.7.
- Levitin, D. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, 20(4), 2142–2152. <https://doi.org/10.1016/j.neuroimage.2003.08.016>
- Lim, V. P. C., Liow, S. J. R., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Determining language dominance in English–Mandarin bilinguals: Development of a self-report classification tool for clinical use. *Applied Psycholinguistics*, 29(3), 389–412. <https://doi.org/10.1017/S0142716408080181>
- Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, 11(5), 1021–1031. <https://doi.org/10.1037/a0024521>
- Lin, Y., Wu, C., Limb, C. J., Lu, H., Feng, I. J., Peng, S., Deroche, M. L. D., & Chatterjee, M. (2022). Voice emotion recognition by Mandarin-speaking pediatric cochlear implant users in Taiwan. *Laryngoscope Investigative Otolaryngology* 7(1), 250–258. <https://doi.org/10.1002/lio2.732>
- Luk, G. (2015). Who are the bilinguals (and monolinguals)? *Bilingualism: Language and Cognition*, 18(1), 35–36. <https://doi.org/10.1017/S1366728914000625>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605–621. <https://doi.org/10.1080/20445911.2013.795574>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. <https://doi.org/10.1037/1082-989X.7.1.19>
- Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, 4(5), 540–545. <https://doi.org/10.1038/87502>

- Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences*, 115(51), 13129–13134. <https://doi.org/10.1073/pnas.1811793115>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. [https://doi.org/10.1044/1092-4388\(2007\)067](https://doi.org/10.1044/1092-4388(2007)067)
- Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology* 35(3), 204–214. [https://doi.org/10.1002/\(SICI\)1098-2302\(199911\)35:33.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2302(199911)35:33.0.CO;2-V)
- Mckay, C. M. (2021). No evidence that music training benefits speech perception in hearing-impaired listeners: A systematic review. *Trends in Hearing* 25, 233121652098567. <https://doi.org/10.1177/2331216520985678>
- Moradzadeh, L., Blumenthal, G., & Wiseheart, M. (2015). Musical Training, Bilingualism, and Executive Function: A Closer Look at Task Switching and Dual-Task Performance. *Cognitive Science*, 39(5), 992–1020. <https://doi.org/10.1111/cogs.12183>
- Moreno, S. (2009). Can Music Influence Language and Cognition? *Contemporary Music Review*, 28(3), 329–345. <https://doi.org/10.1080/07494460903404410>
- Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, 10(6), 719–726. <https://doi.org/10.1111/j.1467-7687.2007.00623.x>
- Morton, J. B., & Trehub, S. E. (2001). Children's Understanding of Emotion in Speech. *Child Development* 72(3), 834–843. <https://doi.org/10.1111/1467-8624.00318>
- Most, T., & Peled, M. (2007). Perception of suprasegmental features of speech by children with cochlear implants and children with hearing aids. *Journal of Deaf Studies and Deaf Education*, 12, 350–361. <https://doi.org/10.1093/deafed/enm012>
- Naeem, K., Filippi, R., Periche-Tomas, E., Papageorgiou, A., & Bright, P. (2018). The Importance of Socioeconomic Status as a Modulator of the Bilingual Advantage in Cognitive Ability. *Frontiers in Psychology*, 9, 1818. <https://doi.org/10.3389/fpsyg.2018.01818>
- Nair, V. K., Biedermann, B., & Nickels, L. (2017). Effect of socio-economic status on cognitive control in non-literate bilingual speakers. *Bilingualism: Language and Cognition*, 20(5), 999–1009. <https://doi.org/10.1017/S1366728916000778>
- Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 1017–1030. <https://doi.org/10.1037/0096-1523.34.4.1017>
- Paap, K. (2019). The Bilingual Advantage Debate: Quantity and Quality of the Evidence. In J. W. Schwieter & M. Paradis (Eds.), *The Handbook of the Neuroscience of Multilingualism* (1st ed., pp. 701–735). Wiley. <https://doi.org/10.1002/9781119387725.ch34>
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258. <https://doi.org/10.1016/j.cogpsych.2012.12.002>
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265–278. <https://doi.org/10.1016/j.cortex.2015.04.014>
- Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, 29(2), 89–112. <https://doi.org/10.1080/20445911.2016.1248436>
- Paap, K. R., Anders-Jefferson, R., Mason, L., Alvarado, K., & Zimiga, B. (2018). Bilingual Advantages in Inhibition or Selective Attention: More Challenges. *Frontiers in Psychology*, 9, 1409. <https://doi.org/10.3389/fpsyg.2018.01409>
- Paquette, S., Takerkart, S., Saget, S., Peretz, I., & Belin, P. (2018). Cross-classification of musical and vocal emotions in the auditory cortex: Cross-classification of musical and vocal emotions. *Annals of the New York Academy of Sciences*, 1423(1), 329–337. <https://doi.org/10.1111/nyas.13666>
- Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology*, 2, 1–14. <https://doi.org/10.3389/fpsyg.2011.00142>
- Patel, A. D., & Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11(9), 369–372. <https://doi.org/10.1016/j.tics.2007.08.003>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, 50(6), 519–530. <https://doi.org/10.1016/j.specom.2008.03.006>
- Penhune, V. B. (2019). Musical Expertise and Brain Structure: The Causes and Consequences of Training. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford Handbook of Music and the Brain* (pp. 417–438). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198804123.013.17>
- Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1664), 20140090. <https://doi.org/10.1098/rstb.2014.0090>
- Preti, E., Suttora, C., & Richetin, J. (2016). Can you hear what I feel? A validated prosodic set of angry, happy, and neutral Italian pseudowords. *Behavior Research Methods*, 48(1), 259–271. <https://doi.org/10.3758/s13428-015-0570-7>
- Raji, S., & de Melo, G. (2020). What Sparks Joy: The AffectVec Emotion Database. *Proceedings of The Web Conference 2020*, 2991–2997. <https://doi.org/10.1145/3366423.3380068>
- Sares, A. G., Foster, N. E. V., Allen, K., & Hyde, K. L. (2018). Pitch and Time Processing in Speech and Tones: The Effects of Musical Training and Attention. *Journal of Speech, Language, and Hearing Research*, 61(3), 496–509. [https://doi.org/10.1044/2017\\_JSLHR-S-17-0207](https://doi.org/10.1044/2017_JSLHR-S-17-0207)
- Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113. <https://doi.org/10.1111/j.2044-835X.2012.02081.x>
- Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence: Music lessons and intelligence. *British Journal of Psychology*, 102(3), 283–302. <https://doi.org/10.1111/j.2044-8295.2010.02000.x>
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language: Music and prosody: An ERP study. *Psychophysiology*, 41(3), 341–349. <https://doi.org/10.1111/1469-8986.00172.x>
- Schroeder, S. R., Marian, V., Shook, A., & Bartolotti, J. (2016). Bilingualism and Musicianship Enhance Cognitive Control. *Neural Plasticity*, 2016, 1–11. <https://doi.org/10.1155/2016/4058620>
- Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, 55(2), 197–212. <https://doi.org/10.1016/j.jml.2006.03.004>
- Shakuf, V., Ben-David, B., Wegner, T. G. G., Wesseling, P. B. C., Mentzel, M., Drefen, S., Allen, S. E. M., & Lachmann, T. (2022). Processing emotional prosody in a foreign language: The case of German and Hebrew. *Journal of Cultural Cognitive Science*, 6(3), 251–268. <https://doi.org/10.1007/s41809-022-00107-x>
- Shameem, N. (1998). *Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians*. 15(1), 86–108. <https://doi.org/10.1177/026553229801500104>
- Shenker, J. J., Steele, C. J., Chakravarty, M. M., Zatorre, R. J., & Penhune, V. B. (2022). Early musical training shapes cortico-cerebellar structural covariation. *Brain Structure and Function*, 227(1), 407–419. <https://doi.org/10.1007/s00429-021-02409-2>
- Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R. (2013). Musical Experience Influences Statistical Learning of a Novel Language. *The American Journal of Psychology*, 126(1), 95–104. <https://doi.org/10.5406/amerjpsyc.126.1.0095>
- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science* 17(8), 675–681. <https://doi.org/10.1111/j.1467-9280.2006.01765.x>
- Strong, J. V., & Mast, B. T. (2019). The cognitive functioning of older adult instrumental musicians and non-musicians. *Aging*,

- Neuropsychology, and Cognition*, 26(3), 367–386. <https://doi.org/10.1080/13825585.2018.1448356>
- Swaminathan, S., & Schellenberg, E. G. (2018). Musical Competence is Predicted by Music Training, Cognitive Abilities, and Personality. *Scientific Reports*, 8(1), 9223. <https://doi.org/10.1038/s41598-018-27571-2>
- Tierney, A., & Kraus, N. (2013). Music Training for the Development of Reading Skills. In *Progress in Brain Research* (Vol. 207, pp. 209–241). Elsevier. <https://doi.org/10.1016/B978-0-444-63327-9.00008-4>
- Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00855>
- Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, 22(3), 516–536. <https://doi.org/10.1017/S1366728918000421>
- Trimmer, C. G., & Cuddy, L. L. (2008). Emotional Intelligence, Not Music Training, Predicts Recognition of Emotional Speech Prosody. *Emotion*, 8(6), 838–849. <https://doi.org/10.1037/a0014080>
- Vaquero, L., Rousseau, P.-N., Vozian, D., Klein, D., & Penhune, V. (2020). What you learn & when you learn it: Impact of early bilingual & music experience on the structural characteristics of auditory-motor pathways. *NeuroImage*, 213, 116689. <https://doi.org/10.1016/j.neuroimage.2020.116689>
- Wiseheart, M., Viswanathan, M., & Bialystok, E. (2016). Flexibility in task switching by monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 19(1), 141–146. <https://doi.org/10.1017/S1366728914000273>
- Wurm, L. H., Vakoch, D. A., Strasser, M. R., Calin-Jageman, R., & Ross, S. E. (2001). Speech perception and vocal expression of emotion. *Cognition and Emotion*, 15(6), 831–852. <https://doi.org/10.1080/02699930143000086>
- Yow, W. Q., & Markman, E. M. (2011). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition*, 14(4), 562–569. <https://doi.org/10.1017/s1366728910000404>
- Yow, W. Q., & Markman, E. M. (2015). A bilingual advantage in how children integrate multiple cues to understand a speaker's referential intent. *Bilingualism: Language and Cognition*, 18(3), 391–399. <https://doi.org/10.1017/S1366728914000133>
- Yow, W. Q., & Markman, E. M. (2016). Children Increase Their Sensitivity to a Speaker's Nonlinguistic Cues Following a Communicative Breakdown. *Child Development*, 87(2), 385–394. <https://doi.org/10.1111/cdev.12479>
- Zeromskaitė, I. (2014). The potential role of music in second language learning: A review article. *Journal of European Psychology Students* 5(3), 78–88. <https://doi.org/10.5334/jeps.ci>
- Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and Neural Correlates of Executive Functioning in Musicians and Non-Musicians. *PLoS ONE*, 9(6), e99868. <https://doi.org/10.1371/journal.pone.0099868>