Benefits of musical experience on whistled consonant categorization: analyzing the cognitive transfer processes

Anaïs Tran Ngoc [a,b,]*, Julien Meyer [b,c,a], Fanny Meunier [a]

[a] *Université Côte d'Azur, CNRS, BCL, France*
[b] *Université Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France*
[c] *Aula de Silbo, Universidad de Las Palmas de Gran Canaria, Spain*

ARTICLE INFO

ABSTRACT

In this study, we investigated the transfer of musical skills to speech perception by analyzing the perception and categorization of consonants produced in whistled speech, a naturally modified speech form. The study had two main objectives: (i) to explore the effects of different levels of musical skill on speech perception, and (ii) to better understand the type of skills transferred by focusing on a group of high-level musicians, playing various instruments. Within this high-level group, we aimed to disentangle general cognitive transfers from sound-specific transfers by considering instrument specialization, contrasting general musical knowledge (shared by all instruments) with instrument-specific ones. We focused on four instruments: voice, violin, piano and flute. Our results confirm a general musical advantage and suggest that only a small amount of musical experience is sufficient for musical skills to benefit whistled speech perception. However, higher-level musicians reached better performances, with differences for specific consonants. Moreover, musical expertise appears to enhance rapid adaptation to the whistled signal throughout the experiment and our results highlight the specificity of instrument expertise. Consistent with previous research showing the impact of the instrument played, the differences observed in whistled speech processing among high-level musicians seem to be primarily due to instrument-specific expertise.

1. Introduction

Music and speech share many similarities. From an acoustic perspective, both signals are complex, containing elements such as melody and rhythm, as well as following rules of syntax and building upon smaller units. From a cognitive perspective, music and speech rely on similar cognitive functions, including attention, memory or executive functions. Such processes are engaged when storing sounds and structures and are activated during production and perception. These similarities have led to studies exploring transfers of musical skills to language processing, demonstrating musically related advantages on various levels of speech (see the review by Besson et al., 2011). Such transfers include improvements in phonological awareness (Bhide et al., 2013), learning new words (Barbaroux et al., 2020) and speech perception in noise (Bidelman and Yoo, 2020; Strait and Kraus, 2011; Varnet et al., 2015). Musical training has also been associated with changes in the brain, including modifications in the development of the temporal and frontal areas (Gaser and Schlaug, 2003), as well as

increases in gray matter and cortical thickness (Bermudez et al., 2009). These changes can be linked to both acoustic and cognitive capacities: for example, changes in the temporal lobe - which includes the auditory cortex - are linked to improved sound perception, whereas increased gray matter in the frontal areas is associated with stronger executive functions. Ultimately, all of these adjustments help to create a "fine-tuned" auditory and cognitive processing system (Strait and Kraus, 2011; Smit et al., 2023), resulting in advantages in speech-related tasks for musicians.

When considering the auditory system, studies exploring psychoacoustic levels show several differences between non-musicians and musicians, including significant differences in auditory thresholds (Tervaniemi et al., 2005; Sharp et al., 2019) and auditory stream segregation thresholds (Johnson et al., 2021). This has led to a broad hypothesis regarding musical transfers to speech: as musical training enhances the processing of basic sound elements such as pitch, rhythm and timbre, these skills may transfer to speech processing through more effective bottom-up mechanisms, since speech also relies on these basic

---

acoustic elements (Barbaroux et al., 2020; Kraus and Chandrasekaran, 2010). Furthermore, it has been suggested that the enhancements in auditory processing depend on the style and instrument specialty of the musicians tested. Kyrtsoudi et al. (2023) demonstrated that classical musicians outperform both Byzantine musicians and non-musicians in a word in noise recognition task. However, Byzantine singers exhibited lower frequency discrimination thresholds than both classical musicians generally and percussionists. These findings suggest that auditory processing—and consequently, advantages in speech perception—may be shaped by the specific demands of musical practice, varying not only with the level of expertise but also with the type of musical experience. Yet, in parallel, musical training has also been shown to improve cognitive functions, such as working memory (Tierney et al., 2008 and review by Yurgil et al., 2020) and executive functions (Frischen et al., 2021), as well as to influence performance on intelligence tests (Degé et al., 2011; Criscuolo et al., 2019). Thus, the impact of musical training on higher cognitive capacities highlights the possibility that transfers towards speech come from general cognitive functions (referred to here as top-down transfer, see Dittinger et al., 2017). Theoretical propositions including both bottom-up and top-down types of transfers, have been proposed within a multimodal framework (see for example the OPERA Hypothesis; Patel, 2012).

In this study, we consider these issues by focusing on the phoneme level, consonants more specifically, and addressing the following questions: 1) which speech cues are impacted by musical experience? 2) If musical advantages are observed, are they due to top-down or bottom-up transfers?

Addressing these questions requires acknowledging the complexity of musical experience, where differences according to instrument and style have been shown to affect musical transfers. As noted by Smit et al. (2023), one challenge in studying musicians lies in the multidimensionality of their expertise, leading to inconsistencies in how it is measured. Musical experience encompasses both type (instrument, style, school of learning) and skill level, which can be operationalized in various ways, complicating the study of music-to-speech transfer. Most research relies on a binary distinction between musicians and non-musicians, despite its limitations. This approach oversimplifies the spectrum of musical training and overlooks individual differences in cognitive and auditory adaptations. Moreover, studies use inconsistent thresholds to define musician expertise, with criteria ranging from years of training (e.g., Zhang et al.'s, 2020, with the *six-year rule*) to musical diplomas, test scores, or starting age (Sala and Gobet, 2020; Cooper, 2020). Such variability underscores the need for a more refined approach. Rather than rigid categories, Smit et al. (2023) advocate for a continuous scale that accounts for the progressive nature of musical expertise, better capturing its influence on cognitive and perceptual abilities.

Considerations concerning the type of musical training received and one's expertise in certain skills due to this training have also been subject to similar unclear and non-unified approaches. For example, the importance of the specific format and organization of musical training in optimizing adaptability is often neglected in descriptions of the musical population, even though this adaptability can be illustrated by opposing different learning styles, such as self-taught musicians with conservatory-trained musicians (Patel, 2012 and Degé, 2021). Moreover, most "musicians" are essentially instrumentalists whose knowledge of their instrument quickly surpasses that of general musical training (ear-training, rhythm, note-reading etc.). Differences according to instrument specialization have even been observed on a neuro-functional level, where musicians showed specific cerebral activation (of motor control and auditory processing regions) and when listening to their own instrument (Pantev and Herholz, 2011; Pantev et al., 2001; Margulis et al., 2009; Choi et al., 2015). However, considering instrument or musical style differences often adds further complexity to studies. For example, Kyrtsoudi et al. (2023) compare Western classical musicians and Byzantine singers, but differentiate

participants in the Western classical musician group only by contrasting percussionists with other instrumentalists.

In our study, we aim to provide a nuanced perspective of musical experience and its impact on speech perception. This is why we choose to recognize both a tiered musical level system, which considers the type of training received, and a panel of instrument specializations among classical musicians (representing instrument families) to explore the multidimensional aspects of musical knowledge. In doing so, we tackle the hypotheses of musical transfer from a different angle, namely, by revisiting Barbaroux (2019)'s distinction between top-down transfers (from general cognitive functions) and bottom-up transfers (from acoustic analyses). We suggest that any differences found between high-level musicians specializing in different instruments would support the bottom-up approach, as we assume that domain general cognitive functions (i.e. memory, attention and executive functions), improved through musical practice (and potentially responsible for top-down transfers), would not vary according to instrument specialization.

To analyze the effect of musical experience on speech perception, and understand which specific cues are affected, we turned towards a form of naturally modified speech that simplifies the spoken signal into a whistled melodic line which remains intelligible to trained listeners: whistled speech (Moles, 1970; Busnel, 1970; Meyer et al., 2024). It is used by speakers of more than 80 languages worldwide for long-distance communication, primarily in mountainous regions and dense forests (Meyer, 2021; Diaz, 2017). This special speech type reduces the phonetic complexity of the spoken signal into whistles modulated in frequency and amplitude and situated within a narrow band of frequencies, between 1 and 4 kHz. Such signals propagate well in rural landscapes and are optimal for human audibility and sound discrimination (Meyer, 2021). While whistled speech allows for the production of the same vocabulary as the spoken mode, to the untrained ear, this speech form is characterized by a "melody" of whistled pitches which may sound like modulated musical notes that are difficult to decode. Yet, it has been shown that naïve listeners engaging in a categorization task based on whistled realizations of phonemes still perform a linguistic task, as their performances vary according to differences in the phonemic systems of their native language (Meyer et al., 2017). Therefore, decoding the speech content of the whistled production relies on the use of speech-specific cues preserved in the spoken-to-whistle transposition. Indeed, whistled speech adapts to the phonological characteristics present in each language by maintaining some salient articulatory and auditory features found in the consonants, vowels, and prosody. For non-tonal languages like French and Spanish, whistled speech primarily transposes the timbre of vowels –or vowel quality- into different pitches, and renders some of the acoustic cues available in the spectral formants of spoken modal speech. Typically, /i/ is whistled with the highest mean pitch values, /o/ the lowest, and /e/ and /a/ lie between the two. Like in spoken (modal) speech, consonants modify the vowels' rather stable frequencies through articulatory movements (see. Fig. 1), modulating the whistled vowel pitch (resembling formant transitions, for example for /s/ or /t/) and/or interrupting the whistled signal (for example for /k/ and /t/), (see Meyer, 2015 for a more detailed description). This selection of elements from the spectral and amplitude makeup of spoken vowels and consonants requires naïve listeners to rely on their knowledge of these cues in speech to identify the whistled productions.

Whistled speech is therefore a useful tool for exploring some aspects of music-to-speech transfers in relation to speech cues: by preserving key acoustic features in a simplified yet natural form, it serves as a natural laboratory for examining speech processing. This has been demonstrated in several previous studies, showing that naïve participants can categorize whistled speech significantly above chance, with specific response patterns depending, for instance, on the type of phoneme (consonant / vowel; Tran Ngoc et al. 2020a, 2020b). As argued in those papers, such results are comparable to insights obtained through studies using speech in noise or artificially reduced/modified speech signals (such as Sine-Wave Speech or Vocoded speech for example). This is also
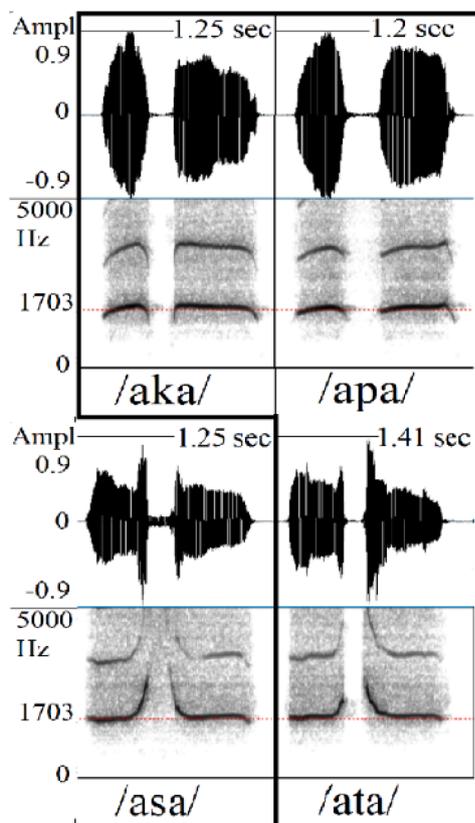
Fig. 1. Spectrogram and signal form for VCV target consonants. The three occlusive (which show an interruption) are outlined in black, highlighting the opposition with /asa/.

in line with studies showing that for ambiguous stimuli that resemble whistled speech—such as sine-wave speech— most naïve listeners initially perceive these sounds as non-speech; However, the same stimuli are processed and perceived as speech (even if not understandable) when the listener's attention is directed toward their phonetic properties—for instance, when the experimenter indicates that the stimuli are speech or asks the participant to perform a speech-related task such as phoneme categorization (see for example Bailey et al., 1977; Best et al., 1989, 1981; Serniclaes et al., 2001). In the present study, we continue to use whistled speech to investigate how such acoustic cues are relevant for phonological categorization, and more generally in the speech perception process, by testing listeners with no knowledge about whistled forms of languages (naïve) and with varying musical backgrounds. In doing so, we advance research in the interest of listener-specific experience by seeking to assess how training levels and instrumental skills modulate one's perception of specific cues present in both modal and whistled speech.

More specifically, the present experimental setting builds on two previous studies examining whistled consonant categorization among French speakers without considering their musical experience (Tran Ngoc et al., 2020a and Tran Ngoc et al., 2022). These studies, in line with what was reported for vowel categorization (Tran Ngoc et al., 2020b), demonstrate that naïve participants with no prior exposure to whistled speech can categorize whistled phonemes significantly above chance by using processes generally applied to modal speech. Important perceptual oppositions have been highlighted between vowels in terms of opening and placement (front/back), as well as between consonants, with a hierarchy between the 4 consonants tested (/k,p,s,t/ which were chosen to focus on different types of consonant transitions). These previous results showed that, among the consonants tested, /s/ and /t/ are better categorized than /k/ and /p/ (i.e., /s = t/ > /k = p/), thus favoring whistled

speech cues characterized by pitch changes mimicking formant transitions of modal spoken speech. A classification of speech cues and a comparison of their different effects has been also derived from perceptual results. The confusions observed could be explained both by common aspects in the frequency shapes of the consonant transitions but also in articulatory dynamics reflected in the amplitude rising times (see Tran Ngoc et al., 2022 for the detailed results).

Here, we include subgroups of semi-professional or professional classical musicians among the 66 total participants, who play one of the 4 target instruments: violin, piano, flute or voice. By focusing on classically trained French musicians —French speakers who studied at a conservatory— we ensured a degree of homogeneity in their musical experience, both culturally and through standardized institutional examinations. At the same time, this selection allowed for a clear distinction in musical skills based on instrumental specialization. We are therefore able to oppose this expert group with participants who have a lower skill level and did not follow or complete such forms of musical training. The 4 target instruments included in this study incorporate each of the different instrument families (string, percussion, wind and voice) which entail different musical skills for their instrument (both in terms of perception and production, notably due to their contrasting timbres). These instrument-based differences have previously been shown to produce different advantages in music-related tasks: singers, for example, have been shown to differ from other instrumentalists when tested for binaural processes (Nisha et al., 2023). We are therefore interested not only in differences in speech perception according to musical experience, but also in instrument specialization. If the effects of musical experience are essentially due to improvements in general cognitive functions, i.e. top-down transfers, all instrumentalists should exhibit similar performances; However, if the performances of the instrumentalists vary according to the musical instrument played, the advantages can be attributed to more specific instrument-based modifications affecting the perception of the signal, i.e. bottom-up transfers.

2. Experiment

2.1. Method

2.1.1. Stimuli

In this experiment, we studied four consonants ([p], [t], [s], [k]). In doing so, we included three occlusive consonants ([p]-bilabial, [t]-dental/alveolar, [k]-velar) and a fricative consonant ([s]-alveolar). These consonants were presented to French listeners in their whistled form and are noted here using the phonemic transcription, typically appearing between slashes "/" (this choice enables us to use the same symbol to represent both the spoken reference in the mind of the listeners and the altered transformations into whistles with which they are confronted). The second author recorded the stimuli (H4N Zoom recorder, using the built-in stereo microphone) produced in a soundproof room by an expert whistler in Silbo (whistled Spanish used in the Canary Islands). Despite the difference in language, the whistled Spanish consonants used here were chosen because of their similarity with French, allowing for a good categorization rate (well above chance, as demonstrated by Tran Ngoc et al., 2022). These consonants were produced and presented in the VCV form /aCa/ (using the vowel /a/ to reduce the variations stemming from co-articulation between different vowels and consonants). By placing the consonant in an intervocalic position, we also involve the highest number of consonant specific cues. These stimuli include oppositions that allow for characteristic distinctions between whistled forms of consonants: 'acute/grave' and 'interrupted/continuous'. 'Acute' refers to the whistled frequency rise from the vowel towards a higher frequency locus during the articulation of the consonant (similar to frequencies of the second formant of spoken modal speech in coronals, when surrounded by /a/). This contrasts with 'grave' where the whistled frequency does not rise from that of the vowel. As whistled frequencies are produced and articulated in the

upper resonant cavities of the mouth, they frequently reflect frequency shapes of the second or third formants of modal speech (F2 most frequently, but also F3 for high front vowels such as /i/) (Shadle, 1983; Meyer, 2015). In the case of coarticulation between back vowels and velar/uvular consonants, whistled pitch may also often resemble the F1 of modal speech (see, for example, /k/ in Fig. 1, and see Meyer et al., 2019). Among the selected consonants, we find /s/ to be acute and semi-continuous, where the important amplitude decay in the middle of the consonant resembles more of a "dip" rather than a cut, emulating the fricative aspect of the spoken [s] when produced by expert whistlers. Contrary to /s/, we find /t/ to be acute yet interrupted (lacking the semi-continuous aspect created by the fricative), and /k/ to be grave and interrupted, similar to /p/ which has a more gradual rise in amplitude (see Fig. 1).

The 16 recordings used as stimuli consist of 4 versions of the 4 sequences, /aka/, /apa/, /asa/ and /ata/. These recordings maintain a very consistent duration, with an average total length of 996 ms ($SD = 84.37$). Although all the VCVs have very similar lengths, the variability in duration is slightly more important for /asa/ (where $SD = 100$ ms). Using the program Praat, we calculated the average frequency of each of the vowels, before and after the consonant modulations. We find that, overall, the average frequency of the vowel /a/ is 1722.78 Hz, with little variation for the different consonants and productions ($SD = 69.37$). The frequency of the vowels preceding and following each consonant are also very consistent, deviating with an average of 48.9 Hz ($SD = 28.36$), with the final vowel being slightly higher than the first vowel (81 % of the time). The consistency in stimuli duration and vowel frequency allows us to consider that any categorization differences between consonants should be attributed to specific consonant cues. In addition to the main distinguishing cues, we notice small differences between consonants, notably between the maximum frequencies of the /s/ and /t/ productions (measured using Praat), where /asa/ reaches systematically higher frequencies than /ata/ (see also Leroy, 1970, Rialland, 2005).

2.1.2. Design

The design used here is identical to Tran Ngoc et al.'s (2022). The experiment consisted of 3 parts. In the first part of the experiment, we evaluated how naive participants (without any previous experience with whistled speech) performed on whistled consonant categorization. In this part, we randomly presented 40 stimuli to the participants, corresponding to a single production of each consonant played 10 times (for 4 consonants). Following this first part, a short section with feedback took place, comprised of 16 consonants, using the same 4 recordings as the first part, this time played 4 times each. In part 3, we tested participants' capacity for consonant categorization again, however, we included more variability in the stimuli. In this last part, the stimuli heard corresponded to 4 productions of each of the syllables produced. Thus, in part 3, participants heard 12 different recordings (3 additional versions of each consonant were added to those heard in part 1), and each recording was played 3 times, giving a total of 48 recordings presented. The experiment, programmed with PCIbex Farm, was proposed online. Therefore, participation took place at home using headphones, earbuds or speakers.

2.1.3. Procedure

Before starting this experiment, we asked participants to fill out a questionnaire indicating their musical experience, including the instrument played, the level achieved and their background in that instrument (this included the number of years of experience or the context in which they took lessons - music conservatory, music school -). We asked participants to choose between 1– beginner (*Débutant*), 2– amateur, 3– confirmed (*Confirmé*), 4– DEM Musical Diploma (*DEM – Diplôme d'Etudes Musicale*), 5– Superior University Diploma (*Diplôme Supérieur*) and 6– professional musician (*Professionnel*). These levels were chosen specifically to target French classical musicians and instrumentalists. We relied on the rigor and organization of French music conservatories to distinguish classical musicians with instrument-

specific diplomas (such as the DEM – received after completing French conservatories, but not equivalent to a university degree, the *Diplôme Supérieur*), from participants who did not have a strong instrument specialization, nor the common instruction required to obtain the DEM or Superior University diplomas.

Then, participants heard one recording of each of the four whistled consonants without any indication of which consonant was played, allowing time to familiarize themselves with the sound quality of whistled speech (and thus its similarities or differences with other instrument timbres). In addition, prior to the start of the experiment, participants were familiarized with the keyboard-based response layout, in which each consonant was assigned to a specific arrow key. The experiment itself then began with part 1, in which participants listened to 40 recordings and were asked to categorize the whistled consonants without any prior training. Then, in part 2, participants completed a short session where feedback was given (either "Good Job" – *Bravo*, or "No this was not the correct response" – *Non ce n'était pas la bonne réponse*) after each categorization. Finally, in part 3, participants were once again asked to categorize 48 whistled consonant stimuli using the arrow keys without any feedback. All of the stimuli were presented in a random order.

2.1.4. Participants

This experiment includes 66 participants, 40 women and 26 men with an average age of 27.2 years old ($SD = 7.03$). All of the participants were native French speakers, with no language disorders, nor any previous knowledge of whistled speech. 36 participants (out of the 66) have a strong musical background, including flutists, pianists, violinists and singers who had achieved a « DEM Musical Diploma » (level 4) at the very least. Among all 66 participants, 10 had no musical experience whatsoever, 4 participants were beginners, 7 amateurs, 9 with confirmed musical experience, 17 participants declared having level 4 (DEM diploma), 8 participants declared having achieved level 5 (Superior University Diploma) and 11 were professional musicians (level 6).

2.2. Results

2.2.1. Differentiating musical transfer according to level of musical experience

We analyzed the data from parts 1 and 3, excluding the results from part 2 (with the feedback), which had too few responses. As a result, we analyzed 88 responses for each of the 66 participants, for a total of 5808 data points. The participants generally achieved an average of 69.1 % correct responses ($SD = 15.6$), with chance at 25 %, demonstrating a high level of correct whistled consonant categorization. When examining the percentage of correct responses based on musical experience level (L), it appears clearly that overall consonant performance increases according to the level of musical experience (except for L6, the highest level/professional musicians): L0 – 47,4 %, L1 – 62,8 %, L2 – 66,7 %, L3 – 68,9 %, L4 – 77,9 %, L5 – 78,4 %, L6 – 76 % (see Fig. 2). These differences highlight gaps between levels, with the two largest gaps being between levels 0 and 1 (15.4 %) and between levels 3 and 4 (9 %). Thus participants with no musical experience (L0) can be distinguished from those with some musical experience (levels 1–3), and they in turn can be differentiated from participants with high levels of musical experience (levels 4–6). Interestingly, this differentiates the type of musical training received, as levels 1–3 include a more diversified musical background (including local music school and being self-taught) compared to the conservatory training which dominates in levels 4–6. For subsequent analyses, the factor Musical experience is categorized in 3 groups of participants (no experience – L0, low-level experience – L1–3, and high-level of experience – L4–6).

We ran a generalized linear mixed model (GLMM) on correct answers (0, 1) with Consonant played (/k/, /p/, /s/, /t/), Part (P1, P3), and Musical experience (None, Low, High) as fixed factors, and Participant as a random effect. All post-hoc analyses were conducted using
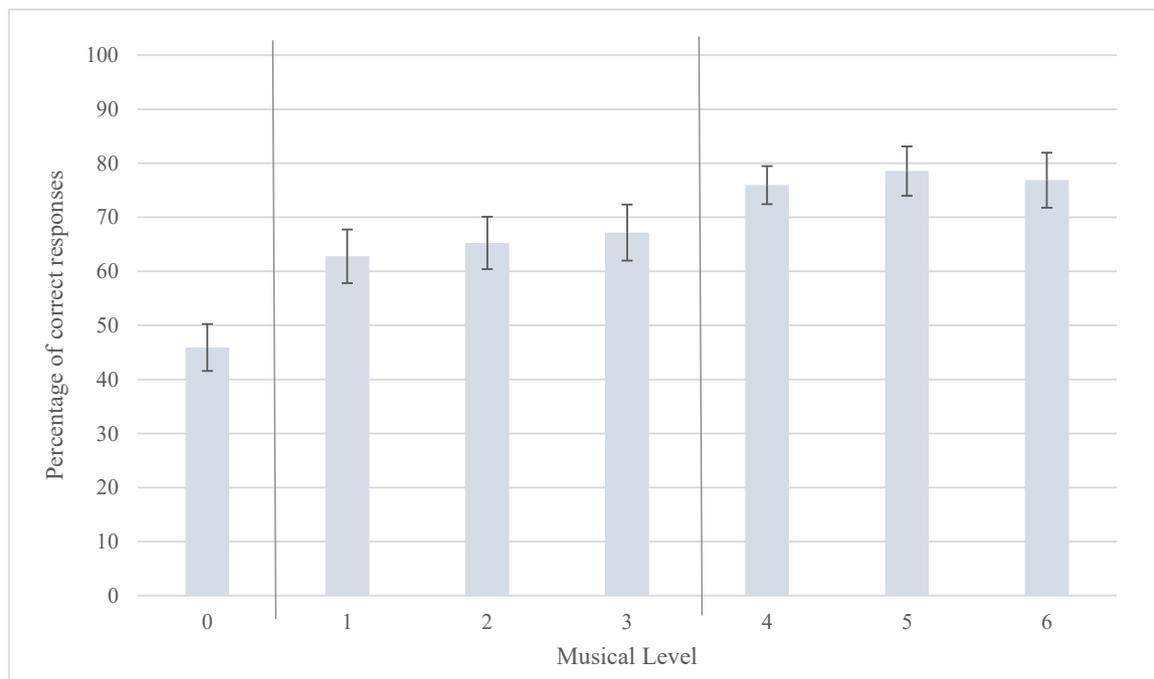
Fig. 2. Percentage of correct consonant responses according to Level with Standard Error shown.

Bonferroni corrections. Significant effects were found for all 3 main factors. There was a significant effect of Musical experience (X^2 (2, $N = 66$) = 31.55, $p < .001$), with post hoc tests showing that participants with a high-level musical experience, who had 77.5 % correct responses ($SD = 14.38$), performed better than those with low-level experience, who had 65.6 % correct responses ($SD = 13.12$). In turn, participants with low-level experience outperformed those with no musical experience, who had 45.9 % correct responses ($SD = 13.74$). There was also a significant effect of Part (X^2(1, $N = 66$) = 6.01, $p = .014$), showing that overall performance was higher in part 3 than in part 1. Finally, Consonant played also reached significance (X^2 (3, $N = 66$) = 289.71, $p < .001$), with the following order of performance: /s > t > k = p/.

We also found three significant interactions, i.e. Part*Consonant, Consonant*Musical experience and Part*Musical experience. The significant interaction between Part*Consonant, (X^2 (3, $N = 66$) = 11.89, $p = .008$), through the post-hoc testing, describes the progression of individual consonant recognition. Specific comparisons reveal that the learning effect between part 1 and part 3 is significant only for the consonant /t/, where participants obtain 69.4 % correct responses in part 1 and 79.4 % in part 3 ($p = .004$). This learning effect is reflected in the hierarchies observed in each part: in part 1, /s > t = k> p/ (ps <0.05) and in part 3, /s = t > k > p/ (ps<0.001).

The significant interaction between Part*Musical experience (X^2 (2, $N = 67$) = 20.56, $p < .001$) provides insight into the effect of musical experience. Post-hoc tests reveal that in both parts, participants with high-level musical experience performed significantly better than participants with no musical experience: part 1 (73.5 % vs. 49.2 %, $p < 0.001$) and part 3 (80.7 % vs. 43.1 %, $p = .004$). In part 3, the effect of musical experience is further nuanced, as a significant difference is observed between participants with low-level musical experience (68.5 %) and those with no experience (43.1 %, $p = 0.004$), as well as between participants with high-level and low-level experience ($p = 0.026$). The only significant learning effect observed (P1 < P3) applies to the high-level musical experience group ($p < 0.001$).

The interaction Consonant*Musical experience is also significant, (X^2 (6, $N = 66$) = 62.62, $p < .001$), with significant differences observed between the Musical experience groups for only 3 consonants: /k/, /p/ and /s/. For /k/, we find significant differences between High (83.3 %,

$SD = 18.85$), Low (60 %, $SD = 20.95$), and None (38.63 %, $SD = 20.35$), with participants in the high-level experience group performing significantly better than the other two groups (ps <0.001). For /p/, a significant difference is observed only between High (60.22 %, $SD = 25.83$) and None (31.36 %, $SD = 21.7$, $p = .009$). Finally, for /s/, significant differences are observed between None (52.27 %, $SD = 29.9$) and each of the other groups (High, 86.36 %, $SD = 14.25$, $p < 0.001$; Low, 85.68 %, $SD = 14.16$, $p < 0.001$), with participants in both the High and Low musical experience groups showing very similar and high categorization rates. These results suggest that while a high-level musical experience confers advantages for 3 of the 4 consonants, even a low-level of musical experience has an effect, differentiating participant's results from those with no musical experience for /s/ (see Fig. 3).

These differences also reveal different consonant hierarchies across the groups. Participants with high-level musical experience show significant differences where /s > t > p/ (ps < .05) and /k > p/ ($p < .001$), resulting in the following hierarchy: /s (=k) > t (=k) > p/. Participants with low-level experience show significant differences where /s > t > k > p/ ($p < 0.01$). Finally participants with no musical experience show that /t > k > p/ and that /s > p/ (ps <0.001), leading to the following hierarchy: /t (=s) > k (=s) > p/. This especially highlights the shift in participants' perception of /s/ based on their level of musical experience.

In view of the various differences found between high-level musicians and other groups, we further explored the nature of the transfer occurring here, by reconsidering the high-level musicians through the instruments they play.

2.2.2. Instrument specialization

To understand how musical experience differs according to the instrument played, we targeted 4 instruments: the violin (9 participants), the piano (7 participants), the flute (8 participants) and voice (7 participants). We excluded other high-level musician participants who did not play these target instruments, thus reducing the number of high-level musician participants to 31. We also included all 30 participants with a low-level (L1 to L3) or no musical experience (L0) in this analysis, therefore amounting to a total of 61 participants, with 5368 data components. Though participants with low-level musical experience play
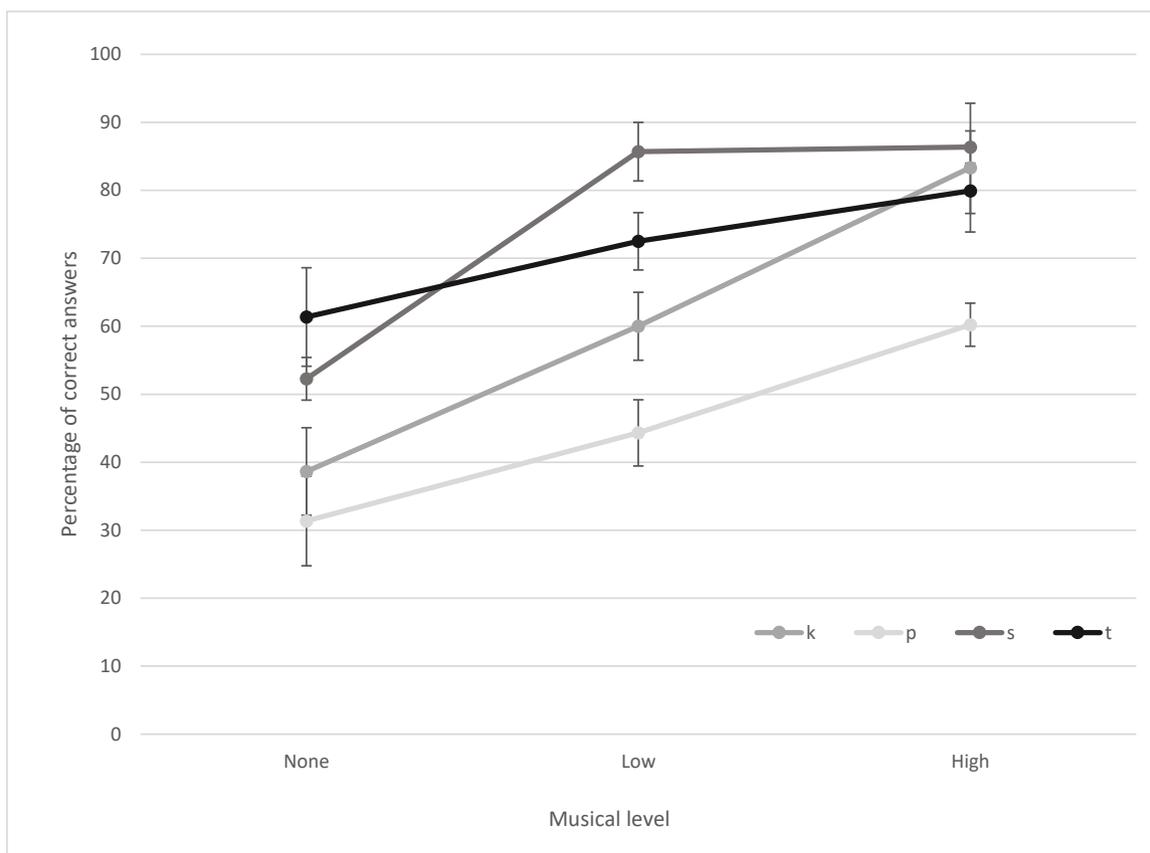
Fig. 3. Percentage of correct responses obtained per consonant for participants with None, Low and High Levels of musical experience.

and specialize in an instrument, we consider that their instrumental skills may not be sufficient for instrument-specific differences. Therefore, we maintained low-level musicians as a distinct group. We applied a GLMM to Correct Answers (0,1) with Instrument specialization (None, Low, Flute, Violin, Piano, Voice) and Part (P1, P3) as fixed factors, with Participants as a random effect. Once again, all post-hoc tests were conducted with the Bonferroni correction.

We find significant main effects of Instrument (X^2 (5, $N = 61$) = 49.98, $p < .001$) and Part (X^2 (1, $N = 61$) = 8.59, $p = 0.003$). The descriptive analyses of the Instrument showed that the highest average performance is obtained by flutists (87.3 % of correct answers, $SD = 10.29$), followed by singers (76.6 %, $SD = 13.17$), violinists (75 %, $SD = 16.86$), and pianists (74.3 %, $SD = 12.24$); while participants with a low-level of musical experience obtained 65.6 % ($SD = 27.48$), and participants with no musical experience 45.9 % ($SD = 26.26$). When applying post-hoc tests, we observe significant differences between participants with no musical experience and every instrument: Flute > None ($p < .001$), Violin > None ($p < .001$), Piano > None ($p = .004$) and Voice > None ($p < .001$), in addition to Low > None ($p = .027$) as found previously. We also observe a significant difference between flutists and participants with a low-level of musical experience (Flute > Low, $p < .001$). Overall, these results suggest that flutists performed best in categorizing consonants and represent the only group of instrumentalists who significantly differ from those with only a low-level musical experience.

However, we also find a significant interaction between Instrument and Part (X^2 (5, $N = 61$) = 17.38, $p = 0.004$). In applying a post-hoc test to this interaction, instrumentalists show significant differences with participants with no musical experience in part 1 and part 3. Interestingly, in part 1, only the Flute is significantly different from None ($p < 0.001$), though the Violin shows a tendency for difference with None ($p = 0.064$). The Flute is also different from Low (Flute > Low, $p < 0.001$).

In part 3, every instrument group performs significantly better than participants with no musical experience (None, all $ps < 0.01$), as does the Low level musician group ($p = 0.005$). As in part 1, Flutists are the only participants that perform significantly better than participants with a low-level of musical experience (Flute > Low, $p = .004$). The singers are the only ones to show a tendency for a learning effect between parts 1 and 3 ($p = 0.073$), where P3 (81.5 %) > P1 (70.7 %).

3. Discussion

3.1. Using fine-grained musical levels and skills to explore music-to-speech transfers

In this study, we took an interest in the role of musical expertise on the perception of consonants while seeking to understand the transfer of skills between music and speech. We choose to use whistled speech, a special phonotype encoding co-articulations between vowels and consonants into a highly melodic phonetic form, akin to a musical melody. Though this speech form only provides insight into the reduced speech cues conserved in whistled speech, it has the advantage of preserving the characteristics of natural spoken language. As a result, we assume that the phoneme categorization task, while not overly easy, still relies on the same type of processing as modal speech (Meyer et al., 2017). Indeed, participants categorize the 4 whistled consonants we tested well over chance, at 69.1 % (with chance at 25 %), with participants scoring well on part 1, which introduces the stimuli without any previous training, and part 3 which includes more variability in the productions. Thus, in line with Meyer et al. (2017), we suggest that listeners -both musicians and non-musicians- processed the stimuli as speech, relying on the cues of their native language, here French. This is all the more justified as the reference to spoken speech was made explicit through the categorization task focusing on phonemes, and this was reinforced by the feedback in

part 2. This further supports our use of whistled speech as a tool to gain insight into specific aspects of speech perception. As we will discuss below, our results also support the decision to categorize musician experience according to different levels – classified here between 0 (for non-musicians) to 6 (for professional musicians) – and further analyzed in 3 groups: participants with no musical experience, those with little experience and those with high levels of musical experience. It can be noted that the high-level musicians included here corresponded to a strict completion of conservatory-based studies, whereas the low-level musicians had globally received a more sparing or irregular musical education. Differentiating these 3 groups and the instruments played by the expert musicians provides fine-grained insight into the effect of musical experience on speech perception.

3.2. A gradual advantage for musicians

The differences in results according to the general level of musical experience reveal that even a small amount of musical experience provides an advantage. Indeed, we found that participants with only a low level of musical experience (levels 1, 2 and 3) reach better scores than participants with no musical experience (19.7 % advantage). Interestingly, participants with high levels of musical experience (levels 4, 5 and 6) showed a 31.6 % difference with participants who had no musical experience, and an 11.9 % difference with participants who had a low level of musical experience. This gradual improvement suggests an incremental transfer, reflecting continuous improvements in musical skills as suggested by Smit et al. (2023). In this view, transfers between music and speech when listening to whistled consonants occur with even a small amount of musical training and continue to increase with higher levels of musical training.

The differences between the three groups of participants (non-musicians, low-level musicians, high-level musicians) were also revealed while comparing results across the different parts of the experiment. High-level musicians showed a significant advantage over participants with no musical expertise in both parts 1 and 3 (24.3 % difference in part 1, and 37.6 % in part 3). In part 3, they also outperformed low-level musicians by 12.2 %, while in turn these low-level musicians showed significant differences with non-musicians in part 3 (but not in part 1). This suggests that musical expertise enhances the ability to learn rapidly throughout the experiment. Supporting this, high-level musicians demonstrated a clear learning effect of 14.5 % between part 1 and part 3. Overall, these findings suggest that musical experience positively affects the ability to process modified speech phonemes and to adapt throughout the experiment, even when more variability is introduced (such as in part 3). Furthermore, these differences suggest a change in one's capacity for adaptation, which also seems to increase according to musical skill as only the high-level musicians show a learning effect.

These differences between groups are further specified when comparing results found for different consonants. Although all participants categorized whistled consonants well over chance, with consonant hierarchies generally in agreement with those found in previous studies (see Tran Ngoc et al., 2022), high-level musicians showed an advantage for 3 of the 4 consonants compared to participants with no musical experience: /k/, with 44.67 % difference, /p/, with 28.69 % difference and /s/ with 33.79 % difference. Low-level musicians showed an advantage for only one consonant (/s/) compared to participants with no musical experience (24.89 % difference). Interestingly, by separating participants into three groups (non-musicians, low-level musicians and high-level musicians), the advantage for /s/ already noted in previous experiments (Tran Ngoc et al., 2022) only appears once participants have a bit of musical experience. This difference in results may be due to the fact that no differentiation was made between non-musicians and low-level musicians in previous studies. For /t/, the consonant with the highest recognition rate among non-musicians, we observe a specific learning effect when all participants are taken into account, though there are no differences between the music level groups. This suggests

that the combination of vowel pitch rise and stop –cues which distinguish /t/ from the three other consonants - are already quite clear for participants with no musical experience and their perception improves rapidly throughout the experiment. In contrast, the distinct perception of pitch change combined with a dip (like /s/) rather than a clear stop highly improves with musical training. This is surprising as we noted that, descriptively, /s/ reaches slightly higher frequencies than /t/, thus amplifying the pitch change. Though more generally, fricatives have been shown to be more robust in noise (Alwan et al., 2011; though the participants' musical experience is not indicated), a recent study comparing the categorization of stops and fricatives in babble noise by musicians and non-musicians found no interaction between consonant type and musical experience, suggesting that neither musicians nor non-musicians had a particular advantage for consonant type (Sajjadi et al., 2021). It is thus surprising that consonant hierarchies should change between non-musicians and low-level musicians. However, it could be that, in whistled speech, the emulation of the fricative noise in the "dip" creates a more delicate shift in the sound quality of the signal (and thus its perceived 'timbre') which is recognized better with musical experience. Moreover, the fact that the advantage shown by high-level musicians is strongest for /k/, suggests that these participants may use the amplitude rise time cue for consonant categorization more than other participant groups. This suggestion is supported by the learning effect found for /t/, as, like /k/, /t/ is also characterized by a short amplitude rise time. The advantage shown for this specific acoustic cue could also explain the improved performance for /p/ found for participants with increased levels of musical experience, as /p/ opposes /k/ only in terms of articulatory cues reflected in amplitude dynamics. Another explanation for this could be a better awareness of articulatory cues compared to participants with little or no musical experience. These differences show that consonant cues are recognized and exploited differently according to musical experience, where only a high-level of musical experience helps participants to focus on acoustic details akin to timbre and the articulatory related cues.

Thus, despite the differences between higher-level musicians and low-level musicians, these results highlight how playing even a small amount of music affects consonant categorization and adaptation to new productions. By considering whistled speech as a reduced form of modal speech, the effect of musical experience on the naïve listener's ability to use whistled speech cues suggests that they should be able to access and use these cues in modal speech (though their role may be redundant). We suggest that these results are ecologically meaningful, highlighting a transfer from musical experience to other perceptive mechanisms (both auditory and cognitive). As such, a number of implications can be deduced from these findings, the most notable being the benefits of music education at any level. In furthering this reflection, we wonder if the difference in results between the low-level musicians and high-level musicians may also lie in the role of institutionalized musical training, as such conditions are ideal for emphasizing precision and repetition which are key to music to speech transfers (Patel, 2012).

More generally, these results are somewhat consistent with previous studies investigating the effects of musical experience on various aspects of whistled speech. In Tran Ngoc et al. (2024a), we examined vowel processing in participants naïve to whistled language with varying levels of musical experience. We observed an advantage for musicians, particularly for /e/ and /a/. However, unlike what we observe here for consonants, there was no significant difference between non-musicians (L0) and low-level musicians (L1,2,3), leading us to contrast only two groups (L0,1,2,3 vs. L4,5,6) in the vowel analyses. In the present findings on consonants, we observe a more nuanced advantage compared to vowels, which appears to emerge progressively with increased musical experience. Moreover, no learning differences were found between the two participant groups over the course of the experiment for vowels, though such an effect is observed for high-level musicians in consonant processing. Overall, these results suggest that the benefits of musical training are more pronounced for consonants than for vowels,

highlighting the variability and granularity of transferable musical skills.

Another recent study on whistled word identification also demonstrated a musical advantage (Tran Ngoc et al., 2024b): participants exhibited gradual improvements in word identification as a function of musical experience, quite similarly to what is observed for consonants here, underscoring the importance of assessing musical experience on a continuous scale and leading to the same three groups. Interestingly, the differences between the three participant groups are larger in the consonant categorization task than in the word identification task detailed in Tran Ngoc et al. (2024b): 19.7 % versus 10.74 % between non-musicians and low-level musicians, and 11.9 % versus 5 % between low-level and high-level musicians. These results suggest that type of improvements associated with musical expertise are task-specific, with certain tasks yielding greater advantages for musicians. Such differences may reflect the nature of musical skills that are transferred to speech processing.

3.3. Instrument specificity and identifying transfer mode

When including instrument specialization in our analyses and opposing 4 instruments from different families, each with their own skill set, we further detailed the transferred skills according to the differences observed between each instrument group. Flutist participants show the strongest advantage compared to participants with no musical experience (41.4 % difference), a significant difference with low-level musicians (21.7 %), and advantages over other musical instruments (with a minimum of 10.7 % difference); a result in line with previous findings. Indeed, Tran Ngoc, et al. (2024b) show that flutists perform more similarly to expert whistlers on whistled word identification than other instrumentalists. The flutists' advantages can be explained in several ways (see Tran Ngoc et al., 2024b). First, the similarity in sound quality between whistled speech and the flute may help flutists to identify the essential acoustic cues. This capacity for enhanced sound categorization according to timbre reflects previously demonstrated timbre-based advantages shown in other contexts (notably for cortical representations of tone, or other neural activity, see for example Margulis et al., 2009; Shahin et al., 2008). Second, the similarities in production linked to the instrument-specific motor skills required for playing, notably the use of consonant articulation in flute attacks (see Dickey and Lasocki 2020), could give flutist participants a more expansive awareness of these consonant productions, and the possible variations present. This point warrants further investigation to determine whether this perceptual capacity extends to other wind instruments with timbres less similar to whistled speech. Indeed, many skills used in flute playing such as continuous pitch control (through air pressure modulation), breath control, and tonging may have an impact, but are common to wind instruments more generally. Investigating differences in performance among wind instrument players would help clarify whether the observed improvements for flutists stem from acoustic processing or production mechanisms.

Additionally, we observe a tendency for an overall learning effect in singers' performances (10.8 % difference between parts) and for a difference between violinists and non-musicians, further suggesting that each instrument is associated with a distinct categorization profile and perceptual behavior. While flutists may benefit from both timbre and articulation similarities with whistled speech, the relative contribution of these traits appears to depend on the musician's specific expertise, with other instrumentalists enhancing their performance through different skill sets. Thus, it is challenging to generalize musical advantages as a homogeneous effect, as different forms of expertise confer distinct perceptual benefits. These findings align with previous instrument-based comparisons in Tran Ngoc et al. (2024b), where performance in the whistled word categorization task varied depending on instrumental background, with some musicians exhibiting processing patterns more or less similar to those of whistlers.

The differences observed here between instrumentalists' results show that the transfer between music and speech varies according to the instrument profile, suggesting that specific changes in perception take place according to the instrument played, rather than through domain general modifications, such as cognitive improvements, due to musical training. This supports the bottom-up hypothesis, implying that high-level musician participants rely on the instrument-specific, low-level perceptual-cognitive skills for improved performances. Indeed, though commonly trained musically skills such as memory, attention or executive functions may also be more efficient for musicians, for example allowing for participants to improve their performances between parts, we suggest that the high-level advantages found in our consonant categorization task do not stem from such cognitive efficiency.

3.4. Limitations

Despite these findings, one of the main limitations of this study is the fact that whistled speech reduces modal speech to only a small number of cues. Though these allow us to target certain aspects of the consonant production, such results may be difficult to reproduce with modal speech. Nonetheless, as demonstrated here, these speech cues remain applicable in the context of speech perception, and provide a deeper understanding of modified speech perception. We suggest that a more generalized application of this experiment using other whistled speech cues, or other forms of modified speech with different categories of acoustic cues, would help to broaden these findings and their application. Another limitation lies in the definition of musical experience and therefore of the application of our hypothesis. As highlighted in this study, the complexity of musical experience is difficult to measure, and even through self-evaluation, the nuances of one's musical knowledge, especially when taught outside the context of a conservatoire, may be misrepresented. Indeed, our hypothesis relies on the differentiation between acoustic and articulatory cues (bottom-up) and cognitive processes (top-down) as highlighted through instrument-specific differences. Yet, unlike the high-level instrumentalists, the low-level musicians in our study cannot be clearly differentiated through this dichotomy (specific and general music skills). Thus, though our findings apply to high-level musicians, it is unclear if the dominating role of bottom-up processes also applies to low-level musicians. However, proposing a continuous measure of musical skill is a first step towards investigating this complex form of knowledge and should allow future studies to expand upon these findings when considering musical experience.

4. Conclusions

This study on whistled consonant categorization provides insight into the transfer process between musical experience and speech perception. Beyond investigating whether musical experience confers an advantage in whistled consonant perception, we also tested two different transfer hypotheses, one based on auditory perception (bottom-up) and the other on cognitive skills (top-down). The results obtained, which underscore the importance of considering musical experience on a graduated scale, confirm that musical experience, even at a low level, offers perceptual benefits. However, higher-level musicians exhibit a stronger advantage. Notably, differences between groups become particularly evident in the final part of the experiment, highlighting the fact that high-level musicians are able to improve over the short duration of the experiment. Moreover, our results reveal that the level of musical experience impacts the four consonants of interest differently. When further refining our analysis by considering instrument specialization, we observed performance variations depending on the type of instrument played. These findings support the bottom-up hypothesis, suggesting that the transfer between speech and music primarily occurs at the low-level perceptual level, shaped by instrument-specific auditory and motor skills (listening and production) rather

than being driven by enhanced memory or executive functions.

Author statement

I hereby certify that all of the authors participating in this article have seen the final version of the manuscript submitted today the 24th of July 2025. This article is original work and is not for consideration elsewhere.

CRediT authorship contribution statement

Anaïs Tran Ngoc: Writing – review & editing, Conceptualization, Investigation, Methodology, Formal analysis, Writing – original draft, Funding acquisition. **Julien Meyer:** Validation, Investigation, Resources, Writing – review & editing, Funding acquisition, Supervision, Conceptualization. **Fanny Meunier:** Writing – review & editing, Formal analysis, Conceptualization, Supervision, Validation, Methodology, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Data availability

Data will be made available on request.

References

Bailey, P.J., Summerfield, A.Q., Dorman, M., 1977. On the Identification of Sine-Wave Analogues of Certain Speech Sounds. Haskins Laboratories, pp. 1–24. SR - 51/52.

Barbaroux, M., Noreña, A., Rasamimanana, M., Castet, Ec., Besson, M., 2020. From psychoacoustics to brain waves: a longitudinal approach to novel word learning. J. Cogn. Neurosci. 33 (1), 8–27.

Barbaroux, M., 2019. Pratique Musicale et Effets de Transfert : De La Perception à La Cognition. Thesis.

Bermudez, P., Lerch, J.P., Evans, A.C., Zatorre, A.C., 2009. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. Cereb. Cortex. 19, 1583–1596. https://doi.org/10.1093/cercor/bhn196.

Besson, M., Chobert, J., Marie, C., 2011. Transfer of training between music and speech: common processing, attention, and memory. Front Psychol. 2. https://doi.org/10.3389/fpsyg.2011.00094.

Best, C.T., Morrongiello, B., Robson, R., 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. Percept. Psychophys. 29, 191–211.

Best, C.T., Studdert-Kennedy, M., Manuel, S., RubinSpitz, J., 1989. Discovering acoustic coherence in acoustic patterns. Percep. Psychophys. 45, 237–250.

Bhide, A., Power, A., Goswami, U., 2013. A rhythmic musical intervention for poor readers: a comparison of efficacy with a letter-based intervention. Mind, Brain, Educ 7 (2), 113–123.

Bidelman, G., Yoo, J., 2020. Musicians show improved speech segregation in competitive, multi-talker cocktail party scenarios. Front Psychol 11, 1927. https://doi.org/10.3389/fpsyg.2020.01927.

Busnel, R.G., 1970. Recherches expérimentales sur la langue sifflée de Kusköy. Revue. De Phonétique Appliquée 41–57. *14/15.*

Choi, U.S., Sung, Y.W., Hong, S., Chung, J.Y., Ogawa, S., 2015. Structural and functional plasticity specific to musical training with wind instruments. Front Hum. Neurosci. https://doi.org/10.3389/fnhum.2015.00597.

Cooper, P.K., 2020. It's all in your head: a meta-analysis on the effects of music training on cognitive measures in school-children. Int. J. Music Educ 38 (3), 321–336. https://doi.org/10.1177/0255761419881495.

Criscuolo, A., Bonetti, L., Särkämö, T., Kliuchko, M., Brattico, E., 2019. On the association between musical training, intelligence and executive functions in adulthood. Front. Psychol. 10, 1704. https://doi.org/10.3389/fpsyg.2019.01704.

Degé, F., Kubicek, C., Schwarzer, G., 2011. Music lessons and intelligence: a relation mediated by executive functions. Music Percept 29 (2), 195–201.

Degé, F., 2021. Music lessons and cognitive abilities in children: how far transfer could Be possible. Front. Psychol. 11, 557807. https://doi.org/10.3389/fpsyg.2020.557807.

Diaz, D., 2017. *El Lenguaje Silbado En La Isla de El Hierro* (segunda edicion Ampliada). Le Canarien ediciones, La Orotava, Tenerife.

Dickey, B., Lasocki, B., 2020. In: Root, D. (Ed.), Tonguing, Grove Music Dictionary Online.

Dittinger, E., Chobert, J., Ziegler, J.C., Besson, M., 2017. Fast brain plasticity during word learning in musically-trained children. Front Hum Neurosci. 11, 233. https://doi.org/10.3389/fnhum.2017.00233.

Frischen, U., Schwarzer, G., Degé, F., 2021. Music training and executive functions in adults and children: what role do hot executive functions play? Z. Erziehungswiss. 25, 551–578.

Gaser, C., Schlaug, G., 2003. Brain structures differ between musicians and non-musicians. J. Neurosci. 23, 9240–9245. https://doi.org/10.1523/JNEUROSCI.23-27-09240.2003.

Johnson, N., Shiju, A.M., Parmar, A., Prabhu, P., 2021. Evaluation of auditory stream segregation in musicians and nonmusicians. Int. Arch Otorhinolaryngol. 25 (1), e77–e80. https://doi.org/10.1055/s-0040-1709116.

Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. Nat. Rev. Neurosci 11, 599–605. https://doi.org/10.1038/nrn2882.

Kyrtsoudi, M., Sidiras, C., Papadelis, G., Iliadou, V.M., 2023. Auditory processing in musicians, a cross-sectional study, as a basis for Auditory training optimization. Healthcare 11 (14). https://doi.org/10.3390/healthcare11142027.

Leroy, C., 1970. Etude de phonétique comparative de la langue turque sifflée et parlée. Rev. Phonétique. Appliquée. 119–161. *14/15.*

Margulis, E.H., Mlsna, L.M., Uppunda, A.K., Parrish, T.B., Wong, P.C., 2009. Selective neurophysiologic responses to music in instrumentalists with different listening biographies. Hum. Brain Mapp 30 (1), 267–275. https://doi.org/10.1002/hbm.20503.

Meyer, J., Dentel, L., Meunier, F., 2017. Categorization of natural whistled vowels by naïve listeners of different language backgrounds. Front. Psychol 8 (25), 1–12.

Meyer, J., Dentel, L., Gerber, S., Ridouane, R., 2019. A perceptual study of CV syllables in both spoken and whistled speech: a Tashlhiyt Berber perspective. In: Proceedings of Interspeech 2019. Fr: Int. Speech Commun. Assoc, Baixas, pp. 2295–2299.

Meyer, J., Rolland, V., Socas, T., Díaz Reyes, D., 2024. A sentence comprehension test with whistled spanish experts. In: Botinis, A. (Ed.), Proc. 14th Exling Conference. Athens, pp. 65–68. Greece.

Meyer, J., 2015. Whistled Languages, A Worldwide Inquiry on Human Speech. Springer, Berlin.

Meyer, J., 2021. Environmental and linguistic typology of whistled languages. Annu. Rev. Linguist. 7, 493–510. https://doi.org/10.1146/annurev-linguistics-011619-030444.

Moles, A., 1970. Etude sociolinguistique de la langue sifflée de Kusköy. Revue De Phonétique Appliquée. 78–118. *14/15.*

Tran Ngoc, A., Meyer, J., Meunier, F., 2020a. Categorization of whistled consonants by naive French speakers. In: INTERSPEECH 2020 –21th Annual Conference of the International Speech Communication Association, pp. 1600–1604.

Tran Ngoc, A., Meyer, J., Meunier, F., 2020b. Whistled vowel identification by French listeners. In: INTERSPEECH 2020 –21th Annual Conference of the International Speech Communication Association, pp. 1605–1609.

Tran Ngoc, A., Meunier, F., Meyer, J., 2022. Testing perceptual flexibility in speech through the categorization of whistled spanish consonants by French speakers. JASA Lett. https://doi.org/10.1121/10.0013900.

Nisha, K.V., Parmar, A., Shivaiah, C., Prabhu, P., 2023. Differential advantages of musical backgrounds on binaural integration and interaction skills in instrumentalists, vocalists, and non-musicians. J. Ontol. 18 (4), 185–192.

Pantev, C., Herholz, S., 2011. Plasticity of the Human auditory cortex related to musical training. Neurosci. Biobehavior. Rev 35 (10), 2140–2154.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12 (1), 169–174. https://doi.org/10.1097/00001756-200101220-00041.

Patel, A., 2012. Why would musical encoding benefit the neural encoding of speech? The OPERA hypothesis. Front. Psychol 2, 142.

Rialland, A., 2005. Phonological and phonetic aspects of whistled languages. Phonology 22 (2), 237–271.

Sala, G., Gobet, F., 2020. Cognitive and academic benefits of music training with children: a multilevel meta-analysis. Mem. Cognit 48 (8), 1429–1441. https://doi.org/10.3758/s13421-020-01060-2.

Serniclaes, W., Sprenger-Charolles, L., Carré, R., Démonet, J.F., 2001. Perceptual categorization of speech sounds in dyslexics. J. Speech, Language Hearing Res 44, 384–399.

Shadle, C.H., 1983. Experiments on the acoustics of whistling. Phys. Teach. 21, 148–154. https://doi.org/10.1119/1.2341241.

Shahin, A.J., Roberts, L.E., Chau, W., Trainor, L.J., Miller, L.M., 2008. Music training leads to the development of timbre-specific gamma band activity. NeuroImage 41 (1), 113–122. https://doi.org/10.1016/j.neuroimage.2008.01.067.

Sharp, A., Houde, M.S., Maheu, M., Ibrahim, I., Champoux, F., 2019. Improved tactile frequency discrimination in musicians. Exp. Brain Res 237 (6), 1575–1580. https://doi.org/10.1007/s00221-019-05532-z.

Smit, E., Rathcke, T., Keller, P., 2023. Tuning the musical mind: next steps in solving the puzzle of the cognitive transfer of musical training to language and back. Music Sci 6. https://doi.org/10.1177/20592043231175251.

Strait, D.L., Kraus, N., 2011. Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. Front Psychol 2 (113), 1–10.

Tervaniemi, M., Just, V., Koelsch, S., Widman, A., Schröger, E., 2005. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. Exp. Brain Res 161, 1–10.

Tierney, A., Bergeson-Dana, T., Pisoni, D., 2008. Effects of early musical experience on auditory sequence memory. Empir. Musicol. Rev 3 (4), 178–186.

Tran Ngoc, A., Meyer, J., Meunier, F., 2024a. The effect of musical expertise on whistled vowel identification. Speech Commun 159, 103058. https://doi.org/10.1016/j.specom.2024.103058.

Tran Ngoc, A., Meyer, J., Meunier, F., 2024b. Musical experience and speech processing: the case of whistled words. Cogn. Sci 48, e70032. https://doi.org/10.1111/cogs.70032.

Varnet, L., Wang, T., Peter, C., Meunier, F., Hoen, M., 2015. How musical expertise shapes speech perception: evidence from auditory classification images. Sci. Rep. [Paper #SREP-15-17927A]. hal-01132995v1.

Yurgil, K.A., Velasquez, M.A., Winston, J.L., Reichman, N.B., Colombo, P.J., 2020. Music training, working memory, and neural oscillations: a review. Front. Psychol. 11, 266. https://doi.org/10.3389/fpsyg.2020.00266.

Zhang, J.D., Susino, M., McPherson, G.E., Schubert, E., 2020. The definition of a musician in music psychology: a literature review and the six-year rule. Psychol. Music. 48 (3), 389–409. https://doi.org/10.1177/0305735618804038.