Speech Communication 159 (2024) 103058

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

The effect of musical expertise on whistled vowel identification

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

[a] Université Côte d'Azur, CNRS, BCL, Nice, France
[b] Université Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France

ARTICLE INFO

Key Words:
Vowel categorization
Whistled speech
Acoustic cues
Musicians
Musical experience
Speech perception

ABSTRACT

In this paper, we looked at the impact of musical experience on whistled vowel categorization by native French speakers. Whistled speech, a natural, yet modified speech type, augments speech amplitude while transposing the signal to a range of fairly high frequencies, i.e. 1 to 4 kHz. The whistled vowels are simple pitches of different heights depending on the vowel position, and generally represent the most stable part of the signal, just as in modal speech. They are modulated by consonant coarticulation(s), resulting in characteristic pitch movements. This change in speech mode can liken the speech signal to musical notes and their modulations; however, the mechanisms used to categorize whistled phonemes rely on abstract phonological knowledge and representation. Here we explore the impact of musical expertise on such a process by focusing on four whistled vowels (/i, e, a, o/) which have been used in previous experiments with non-musicians. We also included inter-speaker production variations, adding variability to the vowel pitches. Our results showed that all participants categorize whistled vowels well over chance, with musicians showing advantages for the middle whistled vowels (/a/ and /e/) as well as for the lower whistled vowel /o/. The whistler variability also affects musicians more than non-musicians and impacts their advantage, notably for the vowels /e/ and /o/. However, we find no specific training advantage for musicians over the whole experiment, but rather training effects for /a/ and /e/ when taking into account all participants. This suggests that though musical experience may help structure the vowel hierarchy when the whistler has a larger range, this advantage cannot be generalized when listening to another whistler. Thus, the transfer of musical knowledge present in this task only influences certain aspects of speech perception.

1. Introduction

Whistled speech is a form of naturally modified speech that transposes spoken (modal) speech into whistles (see Meyer, 2015, for a review). Though whistled speech is intelligible only to trained speakers, previous studies have shown that naive listeners can categorize whistled phonemes correctly and better than chance (see for example Meyer et al., 2017). In most non-tonal languages, the transposition from modal speech to whistled speech relies on a 'formant-based whistling strategy' (Leroy, 1970; Busnel and Classe, 1976; Rialland, 2005; Meyer, 2015), where whistlers make an approximation of the vocal tract articulation in the spoken form to produce the whistled form. This translates to different whistled pitch ranges emitted for each vowel type. This inter-vocalic pitch difference is a simplified whistled reflection of the different spoken frequency distributions characterizing distinct vowels. Vocalic whistled frequencies can be related to specific formant distributions or more generally to different spoken vocalic timbres. In whistled Spanish for example, /i/ has the highest mean pitch value, /e/ is slightly lower, /a/ is even lower and /o/ has the lowest mean frequency, while /u/ is quite low and generally overlaps strongly with /o/ (Meyer, 2008; Diaz, 2008). The intra-vocalic pitch variation observed within the range of whistled frequencies covered by each vowel type depends on several factors including consonant coarticulation, stress, and/or position in the word. There are also some inter-individual variations due to different whistling techniques, different communication distances or to differences in the size of the front oral cavity, specific to each individual (Busnel and Classe, 1976; Diaz, 2008; Meyer, 2015).

Previous behavioral studies on whistled speech have analyzed the four Spanish vowels /i, e, a, o/ extensively, in conditions which verified that the vowel frequency ranges tested did not overlap. They have demonstrated how naive French listeners are able to categorize these whistled vowels correctly well above chance (Meyer, 2005; Meyer et al., 2017; Tran Ngoc et al., 2020). In addition, vowels demonstrated the same categorization hierarchy across studies, despite having been extracted from different whistlers and different types of contexts (words vs. sentences, long distance vs. moderately distanced whistles). Overall,

* Correspondence author.
E-mail address: a.tranngoc.univ@gmail.com (A. Tran Ngoc).

https://doi.org/10.1016/j.specom.2024.103058
Received 6 June 2023; Received in revised form 2 February 2024; Accepted 14 March 2024
Available online 17 March 2024

for naive French listeners, /i/ was always recognized best (with 78 % to 86 % correct answers, depending on the study), followed by /o/ (50 to 59 % correct answers), /e/ (44 to 47 %), and finally /a/ (38 to 44 %). Previous papers have attributed this hierarchy to the frequency distribution of the vowels: one argument was that /e/ and /a/, vowels at intermediate frequency ranges, have two frequency neighbors, while /i/ and /o/ have only one. Another argument has been made based on formant convergence in spoken speech, which also explained the advantage for /i/ (Meyer et al., 2017). The 2017 study integrated listeners with different native languages and showed the impact of listener experience on vowel recognition, where the vowel categorization rate was modulated according to one's native language (Spanish, French or Standard Chinese). The results of the native French and Spanish speakers were not significantly different for correct answers, reflecting the very close vocalic characteristics across languages of the four whistled vowels tested /i, e, a, o/, but they also showed differences in confusions (Meyer, 2008; Meyer et al., 2017), revealing the influence of the different vowel system of each listener's native language. More recently, two studies have focused on whistler variability (Tran Ngoc et al., 2020; Tran Ngoc et al., 2024) by including two different whistlers. They once again confirmed the vowel hierarchy previously observed.

Here, we build on these previous studies, keeping the variability in productions, and go further by looking at how listeners' musical experience affects vowel categorization, given that the whistled speech form may resemble an instrumental mode more than a spoken mode at first listen.

1.1. Musical Experience

Musical experience is shown to have a positive impact on speech perception in a variety of conditions and tasks. This includes improved phonological discrimination compared to non-musicians when listening to L2 sounds (Slevc and Miyake, 2006) or to speech in noisy environments, like in the street or a crowded room (Bidelman and Krishnan, 2010; Strait and Kraus, 2011; Varnet et al., 2015). Such advantages also extend to the discrimination of vowel and consonant sounds (Parbery-Clark et al., 2012), where musicians treat voiced and unvoiced stimuli differently than non-musicians (Ott et al., 2011). Differences in whistled speech perception according to musical experience have also been suggested by Meyer (2008), as the few participants with musical experience included in his whistled vowel perception test showed improved results compared to non-musicians. In addition, another recent study on whistled consonants showed an effect of musical experience and specific instrument training on consonant categorization (Tran Ngoc et al., 2022).

As musical training involves learning to identify and distinguish different pitches, rhythms, and tones (elements that are also present in speech), it would appear that these skills can be transferred to perceiving and processing speech sounds. Among musicians, pitch perception is developed along two different axes, both relevant to speech perception. The first is the auditory perception of frequency, where musicians have been shown to distinguish pitch changes more accurately than non-musicians (Tervaniemi et al., 2016) and with lower frequency discrimination thresholds (Liang et al., 2016). The second is the ability to categorize pitch, encouraged by ear training, often giving rise to abilities such as absolute pitch (giving the name of the note heard, Ross et al., 2005). Such skills have been shown to apply to speech perception in noise (Varnet et al., 2015), and to tone-language perception by non-tone language speakers (Han et al., 2019), where musicians show clear advantages over non-musicians.

Peretz and Coltheart (2003) suggest a mechanistic model for sound processing, where sounds are treated as either music or speech, starting with a common "acoustic analysis" which then feeds-forward either to a music-specific module ("contour analysis"), a language-specific module ("acoustic-to-phonological conversion"), or to a still-not-characterized module ("rhythm and meter"). The initial shared perceptive capacities therefore explain a possible crossover between the two perceptive systems. Moreover, like speech, musical experience also engages various cognitive processes such as attention, memory or executive functions processes (Schellenberg and Weiss, 2013; Parbery-Clark et al., 2009).

Whistled speech blurs the boundaries between the cues typically used to distinguish speech and those used for music, as the whistled vowel pitches reflect aspects of the speaker's timbre. Thus, though musicians may identify the whistled pitches and pitch movement more easily due to improved frequency discrimination, to categorize the vowels correctly, musician participants must integrate such information as part of their phonological representation of the vowels. Swaminathan and Schellenberg (2017), who demonstrated that rhythm training in music did not affect consonant categorization in Zulu clicks for English speakers, suggest that musical competence is only relevant for meaningful cues (see Kraus and Chandrasekaran, 2010). Therefore, the whistled vowels would need to be heard as speech (rather than music) to correctly identify them despite these reduced cues, using what Peretz and Coltheart called the "acoustic-to-phonological conversion". Because of these constraints, we wanted to explore if, and how, musical experience serves as an advantage for whistled vowel categorization.

1.2. Inter-whistler variability

In this study, we integrated inter-whistler variability by using stimuli from two whistlers with different vowel spaces, building on two previous experiments with inter-whistler variability (Tran Ngoc et al., 2020; Tran Ngoc et al., 2024), which tested only non-musician participants. While initial studies on whistled speech have included some intra-talker variability, few studies addressed inter-talker variability in whistled speech, despite research showing that inter-talker variability has significant effects on spoken speech perception presented in adverse conditions (Zaar and Dau, 2015). Indeed, Zaar & Dau (2015) show that the largest perceptual variability was induced by across-talker variability for the same CV items, with different CV confusions according to the speaker. Correlations between certain acoustic phonetic properties and listener comprehension have also been observed for usual modal spoken speech with non-native listeners (Bent et al., 2010), where talkers with a larger vowel space were easier to understand. An experiment displaying a combination of these conditions (native and non-native listeners with inter-talker variability and presented in noise) showed results in the same line, with a significant effect of inter-talker variability on intelligibility (van Dommelen and Hazan, 2012). The acoustic phonetic properties which trigger an effect on speech perception, though they dependent on whether listeners are native or non-native, include more energy in the 1–3 kHz range and an enlarged vowel space in the F2 range. Interestingly, the stimuli from these previous experiments on modal speech deal with constraints which also characterize whistled speech (such as the 1 to 4 kHz range of whistled pitch and the importance of the F2/F3 range in spoken-to-whistled transpositions, see Meyer, 2015) allowing us to investigate the impact of acoustic phonetic inter-talker variations in whistled speech perception. Results from our previous studies testing non-musicians (Tran Ngoc et al., 2020; Tran Ngoc et al., 2024) showed an advantage for the whistler with a larger range, whose whistled vowel productions were easier to categorize, demonstrating an impact of inter-whistler variability.

Here, we add to these conditions, by considering the impact of musical experience on whistled vowel categorization in addition to the inter- and intra- variabilities within the productions of two whistlers. We also explore the possibility of a training effect throughout the different parts of the experiment. To do so, we used a three-part experiment construction, like in one of our previous experiments (Tran Ngoc et al., 2020). Part 1 asks participants to respond to stimuli without any previous introduction, part 2 proposes a very short learning phase where feedback is given, and finally part 3 consists of the same test as part 1, but with stimuli from the other whistler. In one previous study non-musician participants heard the same whistler throughout the experiment (Tran Ngoc et al., 2024) and we did not observe an overall training effect. However, there was a significant improvement of correct categorization rates appearing only for the whistler with the most restricted range and only for the vowel /e/. In the experiment presented here, we included two different whistlers in each list presented to participants, while maintaining the possibility of testing the effect of the whistled vowel range as the different whistlers were also presented separately (one in each part). Thus, we took into consideration a possible effect of inter-talker variability, all while questioning the musician's ability to adapt to an individual whistler-specific frequency distribution.

2. Experiment

2.1. Method

2.1.1. Stimuli

The second author recorded the stimuli in a soundproof room of the Gipsa-Lab (with a H4N-Zoom audio recorder) using the built-in stereo microphone placed 2 m from the whistler to avoid near acoustic effects of the source. Two different expert whistlers were recorded, both teachers of whistled speech in the Canary Islands, and a stereo-to-mono conversion was applied to the recordings (using Matlab). The whistled Spanish vowels targeted, /i/, /e/, /a/ and /o/ were extracted from disyllabic CVCV whistled words. To maintain a similar prosody for each stimulus vowel, we selected vowels from the second unaccented CV syllable. Moreover, to guarantee realistic variability in terms of coarticulation contexts, we selected vowels following various consonant attacks (/d/, /k/, /g/, /t/), where, after removing the consonant attack (only the vocalic nuclei were kept), silence was added to the vowels to create homogenous samples of 500 milliseconds. We selected 48 stimuli per whistler (96 in total) from these recordings, corresponding to 12 versions of each vowel, where 3 different recordings were extracted from the same consonantal context. Due to this limited number of items, the objective was not to test the influence of such contexts, but rather to maintain some aspects of variability due to coarticulation between consonantal and vocalic segments in the experiment. The duration of these 96 whistled vowels also varied, lasting between 146 ms to 473 ms (both by whistler A). The variance for whistler A (6.36 ms) is slightly higher than for whistler B (2.78 ms). When comparing the average duration of each vowel, it appears that /i/ is slightly shorter for both whistlers, as it is in modal speech, though these durations are variable for each of the vowels produced (see Fig. 1). An ANOVA with repeated measures on vowel duration, with Whistler and Vowel type as factors, shows a significant difference between Whistlers ($F_{1,11} = 6.60$, $p = 0.026$), with whistler B's productions being significantly longer than those of whistler A, and a significant difference between Vowels ($F_{3,33} = 3.75$, $p = 0.02$), but no Whistlers*Vowels interaction. The application of post-hoc tests for specific comparisons with Bonferroni corrections only shows /o/ to be significantly longer than /i/ ($p = 0.03$).

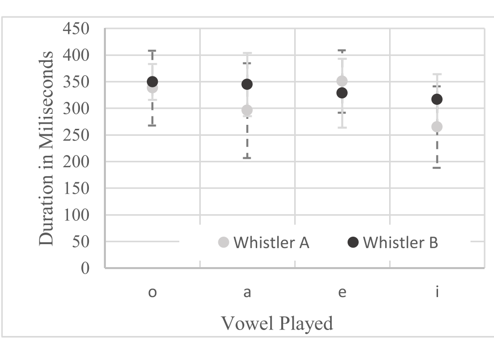

Fig. 1. Distribution of vowel duration average with presentation of the standard deviation according to whistler.

In terms of frequency, the value attributed to each vowel corresponded to the whistled vowel frequency average measured across the duration of each vowel nucleus (thus excluding rapid frequency modulations of consonant attacks). Certain vowel groups vary more than others (see Fig. 2). This is partly due to the fact that some vowels require more strength for whistled production (especially the front /i/ which is most acute, followed by /e/), and to the varying consonantal contexts from which the vowels were extracted, even though we retained only the part of the vowel containing a stable frequency. Typically, coarticulation with coronal consonants push vowel frequencies to higher values, whereas velar consonants pull towards lower values (Meyer et al., 2019; Tran Ngoc et al., 2022). This is especially the case for high frequency whistled vowels, /i/ and /e/ (SD = 256.64 Hz, SD = 196.99 Hz respectively), where we note that /i/ frequencies vary strongly for both whistlers, whereas /e/ variations are more present for whistler A than for whistler B. By contrast, /a/ and /o/ are more stable for both whistlers (SD = 120.6 Hz and SD = 75 Hz respectively). Although there is some variability for each vowel, the main difference between the recordings occurs in the vowel range of the whistlers. Indeed, the vowel frequencies of whistler A are proportionately less spread out than those of whistler B, whereas the vowel groups of whistler B are more distanced from one another and the frequencies are more stable, with the exception of /i/ (see Fig. 2).

2.1.2. Design

The design used is identical to the one in Tran Ngoc et al., 2020. In the first part of the experiment, we evaluated how naive participants performed on whistled vowel categorization when hearing one whistler's productions. In this part, we randomly presented 48 stimuli to the participants, corresponding to 12 versions of each vowel type. These include 3 different recordings of each vowel, extracted from the same consonantal context. After this first part, a training section with feedback ensued, comprising 16 vowels with 4 recordings of each vowel (each corresponding to a different consonant attack), which used the vowel productions of the same whistler as part 1. In a third part, participants listen to the stimuli from the other whistler, which consist of 48 whistled vowels (12 versions of each vowel type, with the same criteria as part 1). This design enabled us to test whether participants with musical experience showed an overall training effect between parts, or whether listeners rely on a single relative pitch scale. Overall two lists, both containing three parts, i.e. part 1 (test), part 2 (training) and part 3 (test), were tested: one with whistler A first and one with whistler B first. Each participant was presented with only one of the lists. The experiment was proposed online and was programmed with PCIbex Farm. Thus, participation took place at home using headphones, earbuds or speakers.

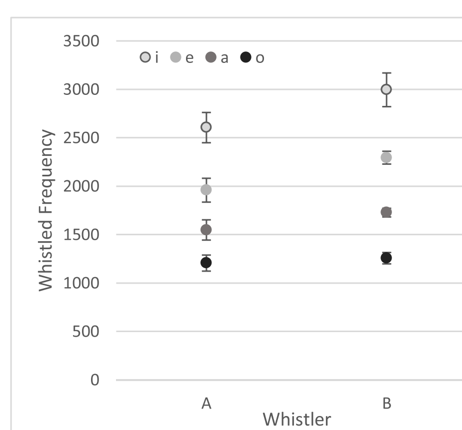

Fig. 2. Variability in mean frequencies of whistled vowel productions for each whistler and each vowel.

2.1.3. Procedure

Before starting the experiment, participants were asked to fill out a questionnaire aiming to collect information on their language and musical backgrounds. We asked participants to indicate their age, the languages they speak and the level of each language (rated on a three-level scale: "beginner", "intermediate", "confirmed"). They were also asked to indicate their musical experience, including the instrument played, the level achieved for each instrument, and their background in said instrument (location or context of lessons and how many years of experience they had). As an indication of each participant's instrument level, we asked participants to choose between 1 – beginner (*Débutant*), 2 – amateur, 3 – confirmed (*Confirmé*), 4 – DEM musical Diploma (*DEM*), 5 – Superior University Diploma (*Diplôme Supérieur*) and 6 – professional musician (*Professionnel*). These levels were chosen specifically to target French classical musicians by including music diplomas such as the DEM.

After the questionnaire was completed, the experiment itself started. Part 1 presents participants with recordings performed by one of the whistlers. It asks participants to categorize the whistled vowels heard without any training, and using the arrow keys. The arrow keys are attributed to each vowel following the keyboard layout, and are visually presented before and during the experiment. After giving participants instructions explaining the task, they are presented with the stimuli in a random order. Each recording is played once, and responses are accepted 500 ms after the start of the recording, to ensure that participants listened to the entire clip before responding. The next recoding is played 200 ms after the participant's response. In the second part (part 2), participants then complete a short training session with feedback for four versions of each whistled vowel (from the same whistler as the one heard in part 1). These stimuli are presented in a random order. The feedback, which was shown as soon as participants responded, consisted of either "No, this was not the correct response" (*Non ce n'était pas la bonne réponse*), or "Congratulations!" (*Bravo!*). The feedback was shown for 1000 ms, before moving on to the following vowel. All participants were given the same form of training regardless of their musical experience. Finally, in part 3, participants are asked to categorize the whistled vowels of the other whistler (if they heard whistler A in parts 1 and 2 now they will hear whistler B and the reverse if they first heard whistler B). Aside from using the other whistler's recordings, this part is identical to part 1.

2.1.4. Participants

This experiment was conducted in accordance with the Helsinki agreement. Sixty-seven participants were included in this study. They were native French speakers who had no language impairments nor any previous knowledge of whistled speech. They were all between 18 and 50 years old ($M = 27.25$, $SD = 7.16$).

We chose to divide the participants into two groups according to the levels of musical expertise, opposing those with a high-level of musical skill (as verified through diplomas), which we called "musicians" (levels 4, 5 and 6) and the low-level musicians or participants without musical experience, which we called "non-musicians" (levels 0, 1, 2 and 3). This echoes previous definitions of the "musician", as a form of knowledge that is skill based, and generally tested in a performance-related capacity. Indeed, citing Hallam (2010), as described by Zhang et al. (2020), we propose that a musician is "someone who has the ability to play a musical instrument". We therefore defined participants' musical level according to participants' main instrument level (often declared first and defined as the instrument on which they have attained the highest level).

The non-musician group included 30 participants, with 22 women and 8 men, who had an average age of 29.5 years old ($SD = 8.79$). In this group, 10 participants had no musical experience whatsoever; however, the non-musician group also included 4 participants who were beginners, 7 amateurs and 9 participants with confirmed musical experience. The musician group included 37 musician participants with 18 women and 19 men, with an average age of 25.37 ($SD = 4.89$). This group contained 18 musicians who had obtained their DEM diploma, 8 musicians with a Superior Diploma and 11 professional musicians.

All of the participants recruited spoke a second language, with a majority speaking English. As the whistled phonemes were based on a Spanish form of whistling and produced by Spanish speakers, we took a special interest in participants who have experience with Spanish. In total 35 out of 67 participants spoke Spanish, where 17 participants had a "beginner" level, 14 participants had an "intermediate" level, and only 4 had a "confirmed" level. Thirty-two participants spoke no Spanish.

Overall, 37 participants heard whistler A first, and 30 participants heard whistler B first. When considering our two musical experience-based groups, 15 participants heard whistler A first and 15 participants heard whistler B first among the non-musicians; In the musician group, 22 participants heard whistler A first, and 15 heard whistler B first.

2.2. Results

In our vowel analysis, we considered the answers for 48 items in part 1 and for 48 items in part 3 for a total of 96 items for each participant. Two musician participants were excluded as they performed outside 2 standard deviations from their group. We therefore considered the data of 65 participants (30 non-musicians and 35 musicians) for a total of 6240 items in this analysis. We find that overall, participants categorized the whistled vowels correctly with 60.83 % of correct responses obtained ($SD = 13.5$), well over chance at 25 %.

Before running our main analyses and as we used Spanish vowel productions (though we underline the similarity between French vowels and the Spanish vowels chosen), we performed an ANOVA on Correct Answers to test whether experience with this language influenced overall vowel categorization rates. We included Spanish Level as a variable. This revealed no significant effect ($F < 1$).

We first ran a generalized linear mixed model (GLMM) on Correct Answers. We included four fixed factors: Musical Experience (musician/non-musician), Part (P1, P3), Whistler (A, B) and Vowel Type (/a/, /e/, /i/, /o/). We included Participant as a random factor.

There is a significant main effect of Vowel (X^2 (3, $N = 65$) = 435.381, $p < 0.001$). It appears that /i/ was categorized best (at 80.8 %), followed by /o/ (at 63.9 %), both of which were much better categorized than /a/ and /e/ (at 50.8 % and 47.8 % respectively). Post-hoc tests (with Bonferroni corrections applied throughout the results section) revealed that all vowels give significantly different performances from each other (ps <0.001) except for /a/ and /e/. We also observed a significant main effect of Whistler, X^2 (1, $N = 65$) = 65.3, $p < 0.001$, where 65.3 % of correct answers ($SD = 17.3$) are obtained for whistler B and 56.3 % (SD = 13.1) for whistler A. There was also a significant effect of Musical Experience (X^2(1, $N = 65$) = 4.772, $p = 0.029$), where musicians obtained 64.6 % ($SD = 13.7$) of correct responses, and non-musicians obtained 56.4 % ($SD = 12$) of correct responses. There were no significant effects for Part ($p > 0.05$) suggesting that there was, overall, no general training effect.

We observed a significant interaction between Musical Experience and Whistler (X^2 (3, $N = 65$) = 6.357, $p = 0.012$). Post-hoc tests reveal that while the difference between the groups is not significant for whistler A, there is a significant difference between musicians and non-musicians for whistler B (71.1 % compared to 58.5 %, $p = 0.023$). The differences between the whistlers are significant for both groups but bigger for the musician group. Indeed, for the musician group, whistler B gives rise to performances that were 13 % higher ($p < 0.001$, where $M =$ 71.13 %, $SD = 16.1$ for B; and $M = 58.09$ %, $SD = 14.15$ for A), while for the non-musician group, performances for whistler B were only 4% higher ($p = 0.001$, $M = 54.31$ %, $M = 58.54$ %, $SD = 16.44$ for B, $SD =$ 11.73 for A).

We also observed three simple interactions in which the factor Vowel type interacted significantly: with Whistler (X^2 (3, $N = 65$) = 21.363, p

< 0.001), Part (X^2 (3, $N = 65$) = 25.310, $p < 0.001$), and Musical Experience (X^2 (3, $N = 65$) = 60.495, $p < 0.001$). These interactions suggest that the effect of each of these factors depends on the vowel played. Moreover, two double interactions reached significance: Musical Experience*Whistler*Vowel (X^2 (3, $N = 65$) = 38.776, $p < 0.001$) and Whistler*Part*Vowel (X^2 (3, $N = 65$) = 15.124, $p = 0.002$). In order to understand these results and interactions, we analyzed the data independently for each vowel (see Fig. 3).

For each vowel, we applied a GLMM with Musical Experience (musician, non-musician), Part (P1, P3), and Whistler (A, B) as fixed factors. We included Participant as a random effect. We note that for each of these analyses $df = 1$ and $N = 65$. All post-hoc analyses used Bonferroni corrections.

For the vowel /i/ there is a significant main effect of Musical Experience ($X^2 = 4.49$, $p = 0.034$) showing a difference between the results of non-musicians and musicians, where non-musicians ($M = 83.12$ %, $SD = 10.85$) perform better than musician participants ($M = 78.8$ %, $SD = 13.15$). There is a significant effect of Whistler ($X^2 = 36.3386$, $p < 0.001$), where performances with whistler B give rise to 86.41 % of correct categorization, and those with whistler A give rise to 75.25 % of correct responses. Thus, the results obtained for whistler B are superior to those obtained by whistler A. However, the significant interaction between Musical Experience*Whistler ($p = 0.008$) reveals that this difference is significant only for non-musician participants ($p < 0.001$) with 91.4 % of correct answers for whistler B and 75.5 % of correct answers for whistler A. We also observed a difference between groups only for whistler B, where musicians obtained only 82.14 % of correct answers.

For /e/, there is a significant effect of Musical Experience ($X^2 = 8.1567$, $p = 0.004$) which shows that musician participants ($M = 53.92$ %, $SD = 21$) perform better than non-musicians ($M = 40.69$ %, $SD = 17.84$). We also find a significant effect of Part ($X^2 = 10.37$, $p = 0.001$) for which performances are higher in part 3 (52.4 %) than in part 1 (43 %). For this vowel, one simple interaction is significant between Musical Experience*Whistler, ($X^2 = 26.0908$, $p < 0.001$) where we find a significant difference between whistlers A (45 %) and B (62.9 %) only for musicians ($p < 0.001$). We also find a significant difference between musicians and non-musicians, though only within the results of whistler B ($p < 0.001$), with 62.9 % of correct responses for musicians, and 36.7 % of correct responses for non-musicians.

For the vowel /a/, we find a main significant effect of Musical Experience ($X^2 = 12.6636$, $p < 0.001$), which shows that musicians ($M = 60.23$ %, $SD = 24.24$) categorize this vowel better than non-musicians do (M = 39.72 %, $SD = 23.30$). We also observe a significant effect between parts ($X^2 = 13.5121$, $p < 0.001$), for which performances are higher in part 3 (54.4 %) than in part 1 (42.4 %), and a significant effect of Whistler ($X^2 = 4.6066$, $p = 0.032$), where, when hearing whistler A, participants obtain 47.9 % of correct responses, compared to 53.56 % for whistler B. No interaction reaches significance.

We observed two significant main effects for the vowel /o/, for Whistler ($X^2 = 34.482$, $p < 0.001$), where responses for whistler B are at 70.5 % and responses for Whistler A are at 57.31 %, and for Part ($X^2 = 6.608$, $p = 0.01$), where participants obtain 65.6 % in part 1 and 62.18 % in part 3. We observe a significant interaction between Musical Experience*Whistler ($X^2 = 19.638$, $p < 0.001$), where the application of a post-hoc test shows a significant difference between whistlers for musician participants ($p < 0.001$), where whistler B responses are at 76.2 % and whistler A responses at 54.8 %. We also find a significant difference between musicians (76.2 %) and non-musicians (63.8 %) for whistler B. We recapitulated the significant differences per vowel in Table 1.

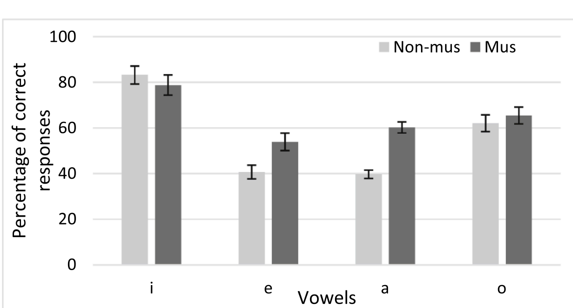

Fig. 3. Average percentage of correct responses for musicians and non-musician participants, with standard error indicated for each vowel.

3. Discussion

In this experiment, we looked at how whistled vowels are categorized by naive listeners depending on their musical expertise. We aimed to extend previous results and look at how musical expertise modulates performance when faced with whistler variability and a short training segment with feedback. First, the data confirmed previous results from other experiments on non-musicians. Overall, whistled vowel categorization was obtained with an average of 60.8 % of correct responses (well over chance at 25 %), and we replicated the hierarchy between the vowel categorization performances: /i/ is better recognized than /o/, which is better recognized that /a/ and /e/ (which were not different from each other), leading to the hierarchy $i > o > a = e$.

Concerning the role of musical expertise, our results are in line with the literature, showing better performances for musicians than for non-musicians. However, the interactions showed that this advantage is in fact more specific than it may seem, as different effects are observed depending on the vowel produced and the whistler, specifying the role of musical experience. Interestingly enough, the effect of musical expertise is clear for the least well-recognized vowels (/e/, /a/, and /o/), where for /e/ and /o/ this advantage is specific to the productions of the whistler with the widest range of frequencies (whistler B). For /a/ the advantage for productions of whistler B over A is general, applying to both musicians and non-musicians. Overall, this suggests that musical expertise mainly affects the categorization of the whistler with the larger vowel frequency range. This effect appears for vowels /e/, /a/ and /o/ that show less pitch variability across different productions. These same vowels are also more stable in whistler B's productions than in whistler A's.

However, for /i/, the best-recognized vowel, the pattern of answers is different. Indeed, only non-musicians show a preference for whistler B, for whom they also show better results than the musician participants (9.26 % difference). This could be due to a higher variability in the whistled frequencies of /i/ (Fig. 2), which may affect musicians more strongly. As musicians are better than non-musicians at distinguishing pitch differences (see Tervaniemi et al., 2016, Liang et al., 2016) this non-musician advantage may be due to an over-reliance on pitch matching, or over-sensitivity to these variations. By contrast, non-musicians' ability to better assimilate the whistled /i/ to modal speech may partly come from more receptivity to secondary acoustic characteristics of vowels, such as duration. Indeed, in our corpus, /i/ is the vowel with the shortest duration, which matches with general spoken intrinsic durations observed in phonetics (House and Fairbanks, 1953;

Table 1
Comparison of post-hoc test results per vowel according to Musicians (M), Non-musicians (Nm), Part (P1,P3) and Whistlers A and B; only significant effects are shown.

Whistled Vowel		/i/	/e/	/a/	/o/
Musical Experience		M < Nm	M > Nm	M > Nm	
Whistler		*A < B*		*A < B*	*A < B*
Part			P1 < P3	P1 < P3	P1 > P3
*Mus*Whist*	*Musicians*		*A < B*		*A < B*
	Non-musicians	*A < B*			
	Whist B	Mus < Nm	Mus > Nm		Mus > Nm

Solé and Ohala, 2010).

To summarize the inter-talker variability effects highlighted above, listening to whistler B, with a larger vowel frequency space than whistler A, leads to better performances than listening to whistler A. This effect exists for every vowel but at different levels: first, for /i/, we find a significant difference for non-musicians only (15.9 % in favor of whistler B). For the two vowels /e/ and /o/, we only observed a significant difference for musicians (with respectively 17.9 % and 12.4 % better results on productions of whistler B), whereas performances for /a/ are significantly higher for all participants (5.66 % in favor of whistler B).

Finally, we found no overall training effect; however, there was a significant increase in performances during the experiment for two specific vowels, generally less well-categorized (/a/ by 12 %, and /e/ by 9.4 %). Interestingly, these training effects were not specific to musicians and were present regardless of the whistler heard. This complements results we highlighted earlier for /e/ and /a/, notably the fact that a musical advantage was found only on whistler B for these two vowels. This reinforces the interpretation that a finer auditory sensitivity to pitch in musicians is partly responsible for the result patterns. These last results on 'training' are in line with those of another study (Tran Ngoc et al., 2024), which included only non-musician participants: findings showed no evidence of a global training effect, even when the same whistler was presented throughout the experiment. However, a significant improvement in correct categorization was observed for the vowel /e/ and only for the whistler with the most restricted range. This suggests that participants quickly improve their recognition of the least well-categorized vowels. Intriguingly, in the present experiment we also observe a slight decrease between parts 1 and 3 (−3.4 %) for /o/, the second best categorized vowel. We wonder whether having whistler A (with a smaller range) in the first part of the experiment may be disturbing, leading listeners to perform worse than during their initial perception. However, given the large difference between the whistlers and the lack of simple interaction between the whistlers and the parts, it would be inappropriate to elaborate on these dynamic effects and further experiments should be conducted to explore this point.

Nonetheless, and in line with existing literature (van Dommelen and Hazan, 2012), our findings suggest that more stable frequencies and a larger vowel space facilitate abstract representations of certain vowels. Previous studies have clearly shown that the task proposed in this experiment triggers whistled vowel categorization with different answer patterns. These have been observed between populations of different language backgrounds tested on the same protocol, but with productions of a different whistler. These findings reflect the influence of different vowel spaces from the listeners' mother tongues (Meyer et al., 2017). As our present task requires the listeners to perform a pitch-to-timbre matching/association between the whistled vocalic pitch heard and the mentally recalled vowel quality, this process is eager to be influenced by acoustic parameters other than pitch. Some examples include parameters defining the formants of each vowel, as well as formant proximities, already known to be important for vowel identity (see Meyer et al., 2017, for a discussion). As musicians in occidental musical traditions are trained to recognize pitches and interpret them as notes (pitch categorization), we wonder whether the stability of the vowels with the lower whistled frequencies, as well as the length of whistler B's pitches could then allow musicians to better construct the relative relationships between pitches (especially for /e/ and /o/). Musician participants may also exploit these cues more efficiently due to their enhanced auditory sensitivity. The higher variability found in /i/ in terms of frequency might hinder such skills, partly explaining the non-musician advantage for that vowel, as they may also rely on different cues than musician participants. This suggests that musicians process the whistled signal differently than non-musicians. However, the musicians' advantages are often specific to whistler B. Thus, though we can consider that musicians are able to use their phonological knowledge to categorize whistled vowels well over chance, they exploit the musical similarities in timbre or frequency discrimination only when the signal is stable and thus possibly more similar to musical notes.

4. Conclusion

In conclusion, naive French listeners with and without a high level of musical experience recognize whistled vowels more than 50 % of the time. These results appear to be robust and generalizable. Having a high-level of musical experience proves to be beneficial for the vowels whistled with lower frequencies (/e,a,o/), though this is specific to one of the two whistlers for /e,o/. This advantage was not uniform, and depended on the stimuli according to specific acoustic conditions. We evidence the impact of the whistler heard, where the whistler with the larger range was categorized better than the other whistler, which is apparent for both musician participants and non-musician participants, though more often specific to musician participants. This musical advantage, observed mainly for one whistler, may be due to the stability of the whistled pitches, allowing for better exploitation of pitch-based skills (such as relative interval definition). It also further underlines the influence of a whistler's range and frequency distribution on vowel categorization, showing that a larger vowel space facilitates the creation of abstract vowel representations for both musicians and non-musicians.

CRediT authorship contribution statement

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

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.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank David for the whistled word recordings, and to the participants for volunteering their time.

Funding

This work is supported by the doctoral grant attributed to Anaïs Tran Ngoc by the Université Côte d'Azur, and the framework of the project GeoTeleLing with the support of the CNRS 80Prime interdisciplinary program.

References

Bent, T., Kewley-Port, D., Fergusson, S., 2010. Across-talker effects on non-native listeners' vowel perception in noise. J. Acoust. Soc. Am. 128 (5), 3142–3151.. https://doi.org/10.1121/1.3493428.

Bidelman, G.M., Krishnan, A., 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 1355, 112–125. https://doi.org/10.1016/j.brainres.2010.07.100.

Busnel, R.G., Classe, A., 1976. Whistled Languages. Springer-Verlag, Berlin Heideleberg.

Diaz, D. 2008. El lenguaje silbado en la isla de El Hierro. ↱ Santa Cruz de Tenerife, Excmo, Cabildo Insular de El Hierro.

Hallam, S., 2010. 21st century conceptions of musical ability. Psychol. Music. 38 (3), 308–330. https://doi.org/10.1177/0305735609351922.

House, A.S., Fairbanks, G., 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. Acoust. Soc. Am. 25, 105–113. https://doi.org/10.1121/1.1906982.
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.
Leroy, C., 1970. Étude de phonétique comparative de la langue turque sifflée et parlée. Revue de Phonétique Appliquée 14/15, 119–161.
Liang, C., Earl, B., Thompson, I., Whitaker, K., Cahn, S., Xiang, J., Fu, Q.J., Zhang, F., 2016. Musicians are better than non-musicians in frequency change detection: behavioral and electrophysiological evidence. Front. Neurosci. 25 (10), 464. https://doi.org/10.3389/fnins.2016.00464.
Meyer, J., 2005. Description typologique et intelligibilité des langues sifflées: Approche linguistique et bioacoustique. Ph.D. dissertation. Lyon: Université Lyon 2.
Meyer, J., 2008. Typology and acoustic strategies of whistled languages: Phonetic comparison and perceptual cues of whistled vowels. J. Int. Phon. Assoc. 38 (1), 69–94. https://doi.org/10.1017/S0025100308003277.
Meyer, J., 2015. Whistled Languages, A Worldwide Inquiry on Human Whistled Speech. Springer.
Meyer, J., Dentel, L., Meunier, F., 2017. Categorization of natural whistled vowels by naïve listeners of different language background. Front. Psychol. 8 https://doi.org/10.3389/fpsyg.2017.00025.
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. INTERSPEECH 2019 –20th Annual Confer. Int. Speech Commun. Assoc. 2296–2299. https://doi.org/10.21437/Interspeech.2019-2251.
Ott, C.G.M., Langer, N., Oechslin, M.S., Meyer, M., Jäncke, L., 2011. Processing of voiced and unvoiced acoustic stimuli in musicians. Front. Psychol. 2, 195. https://doi.org/10.3389/fpsyg.2011.00195.
Peretz, I., Coltheart, M., 2003. Modularity of music processing. Nat. Neurosci. 6 (7) https://doi.org/10.1038/nn1083.
Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009. Musician enhancement for speech-in-noise. Ear. Hear. 30 (6), 653–661. https://doi.org/10.1097/AUD.0b013e3181b412e9.
Parbery-Clark, A., Tierney, A., Strait, D.L., Kraus, N., 2012. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience 111–119. https://doi.org/10.1016/j.neuroscience.2012.05.042.
Rialland, A., 2005. Phonological and phonetic aspects of whistled languages. Phonology. 22 (2), 237–271.
Ross, D.A., Gore, J.C., Marks, L.E., 2005. Absolute pitch: music and beyond. E&B 7 (4), 578–601. https://doi.org/10.1016/j.yebeh.2005.05.019.
Schellenberg, E.G., Weiss, M.W., 2013. Music and cognitive abilities. In: Deutsch, D. (Ed.), The Psychology of Music, 3rd ed. Elsevier, Amsterdam, pp. 499–550. https://doi.org/10.1016/B978-0-12-381460-9.00012-2.
Slevc, L.R., Miyake, A., 2006. Individual differences in second-language proficiency: does musical ability matter? Psychol. Sci. 17 (8), 675–681. https://doi.org/10.1111/j.1467-9280.2006.01765.x.
Solé, M.J., Ohala, J.J., 2010. What is and what is not under the control of the speaker. Intrinsic vowel duration. In: Fougeron, C., Kühnert, B., D'Imperio, M., Vallée, N. (Eds.), Papers in Laboratory Phonology 10. de Gruyter, Berlin. https://doi.org/10.1515/9783110224917.5.607.
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. https://doi.org/10.3389/fpsyg.2011.00113.
Swaminathan, S., Schellenberg, E.G., 2017. Musical competence and phoneme perception in a foreign language. Psychon. Bull. Rev. 24 (6), 1929–1934. https://doi.org/10.3758/s13423-017-1244-5.
Tervaniemi, M., Janhunen, L., Kruck, S., Putkinen, V., Huotilainen, M., 2016. Auditory profiles of classical, jazz, and rock musicians: genre-specific sensitivity to musical sound features. Front. Psychol. 6, 1900. https://doi.org/10.3389/fpsyg.2015.01900.
Tran Ngoc, A., Meyer, J., Meunier, F, 2024. Whistled Phoneme categorization: the effect of vowel space range. In: Botinis, A. (Ed.), Proc. 14th Exling Conference. Athens, Greece, 115-118.
Tran Ngoc, A., Meyer, J, Meunier, F, 2020. Whistled vowel identification by French speakers. In: *INTERSPEECH 2020*, 21st *Annual Conference of the International Speech Communication Association,* 1605-1609. https://doi.org/10.21437/Interspeech.2020-2697.
Tran Ngoc, A., Meunier, F., Meyer, J., 2022. Testing perceptual flexibility in speech through the categorization of whistled Spanish consonants by French speakers. JASA Exp. Lett. 2, 405–413. https://doi.org/10.1121/10.0013900, 095201.
van Dommelen, W.A., Hazan, V., 2012. Impact of talker variability on word recognition in non-native listeners. J. Acoust. Soc. Am. 132, 1690–1699. https://doi.org/10.1121/1.4739447.
Varnet, L., Wang, T., Peter, C., Meunier, F., Hoen, M., 2015. How musical expertise shapes speech perception: evidence from auditory classification images. Sci. Rep. 5 (1) https://doi.org/10.1038/srep14489.
Zaar, J., Dau, T., 2015. Sources of variability in consonant perception of normal hearing listeners. J. Acoust. Soc. Am. 138 (3), 1253–1268. https://doi.org/10.1121/1.4928142.
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.