COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 48 (2024) e70032

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70032

Musical Experience and Speech Processing: The Case of Whistled Words

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

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

Received 16 October 2023; received in revised form 31 August 2024; accepted 4 December 2024

Abstract

In this paper, we explore the effect of musical expertise on whistled word perception by naive listeners. In whistled words of nontonal languages, vowels are transposed to relatively stable pitches, while consonants are translated into pitch movements or interruptions. Previous behavioral studies have demonstrated that naive listeners can categorize isolated consonants, vowels, and words well over chance. Here, we take an interest in the effect of musical experience on words while focusing on specific phonemes within the context of the word. We consider the role of phoneme position and type and compare the way in which these whistled consonants and vowels contribute to word recognition. Musical experience shows a significant and increasing advantage according to the musical level achieved, which, when further specified according to vowels and consonants, shows stronger advantages for vowels over consonants for all participants with musical experience, and advantages for high-level musicians over nonmusicians for both consonants and vowels. By specifying high-level musician skill according to one's musical instrument expertise (piano, violin, flute, or singing), and comparing these instrument groups to expert users of whistled speech, we observe instrument-specific profiles in the answer patterns. The differentiation of such profiles underlines a resounding advantage for expert whistlers, as well as the role of instrument specificity when considering skills transferred from music to speech. These profiles also highlight differences in phoneme correspondence rates due to the context of the word, especially impacting "acute" consonants (/s/ and /t/), and highlighting the robustness of /i/ and /o/.

Keywords: Speech perception; Knowledge transfer; Whistled speech; Musical experience

Correspondence should be sent to Anaïs Tran Ngoc, Université Côte d'Azur, CNRS, BCL Lab, CNRS, 25 Av. François Mitterrand, 06300 Nice, France. E-mail: a.tranngoc.univ@gmail.com

1. Introduction

Musical experience affects speech processing in various ways and on various levels (see review by Besson, Chobert, & Marie, 2011). This includes better performance in processing at phonological levels, notably in foreign speech, both in phonological production (Milovanov, Pietilä, Tervaniemi, & Esquef, 2010) and phonological perception (Slevc & Miyake, 2006). Such advantages have also been shown to extend to modified speech conditions such as speech in noise (Bidelman & Krishnan, 2010; Varnet, Wang, Peter, Meunier, & Hoen, 2015), due to a better representation of the target acoustic stimuli (Parbery-Clark, Skoe, & Kraus, 2009) or improved attention skills (Strait & Kraus, 2011). Such findings underline similarities in processing between speech and music (see review by Sammler & Elmer, 2020), as well as common structural aspects.

A promising new path combining language processes and musical expertise considers using musical surrogate languages to understand shared processing mechanisms (McPherson & Winter, 2022). Along the same line, recent studies testing language processing by musicians have been applied to whistled speech (Tran Ngoc, Meunier, & Meyer, 2023; Tran Ngoc, Meyer, & Meunier, 2022a). This practice, used to acoustically transpose spoken dialogs rather than as a type of musical production, reduces the vocal spoken signal to a simple modulated whistled line akin to a musical melody. Whistled speech has evolved in a large diversity of languages worldwide in mountainous and densely forested regions, enabling true distance communication (Meyer, 2015). The physical characteristics of whistles are well adapted to the acoustic constraints of the environment, as they focus on a narrow range of frequencies (1000–4000 Hz) that favor sound propagation and that are higher than most prevalent natural background noises (usually low-frequency content). Moreover, these frequencies are optimal for human audibility and sound discrimination (Meyer, 2021). The transposition of the linguistic segments—typically vowels and consonants—by speakers of nontonal languages (such as Spanish, Turkish, Tamazight, and Greek) is one of the most interesting aspects of whistled speech, proposing alternative insights into how the acoustic realization of phonemes can be drastically reduced without hindering recognition from listeners. While mastering this speech form does require training, studies have shown that even without extensive training, naive listeners can successfully categorize phonemes in both whistled consonants and whistled vowels. Recent findings demonstrate how naive listeners can already categorize phonemes in this modified form correctly and well over chance (Tran Ngoc, Meunier, & Meyer, 2022b; Tran Ngoc, Meyer, & Meunier, 2020b). Furthermore, categorization tasks highlighted differences in performance that depended on the type of consonant or vowel heard (among the four consonants/vowels of interest).

In nontonal languages, whistled speech produces different pitch categories according to the spoken vowel timbre, thus transposing each of the vowels of modal speech to a specific whistled frequency range (which is also relative to the speaker and the whistling technique). In whistled Spanish, the language tested in our previous studies, the whistled vowel pitches can be ordered from highest to lowest in the following manner: /i/, /e/, /a/, and /o/, with /u/ generally overlapping with /o/ and /a/ (Busnel & Classe, 1976; Meyer, 2015; Rialland, 2005). These pitches often mimic the second or third formant in modal speech, without maintaining

the precise frequency (Meyer, 2015). Whistled consonants modulate/change these pitches according to their corresponding spoken articulation. For example, when produced in a vowel-consonant-vowel (VCV) context with the vowel /a/, articulation causes a large and rapid pitch change toward/from high acoustic loci for consonants /s/ and /t/. These are called "acute" consonants in some studies (Diaz, 2017; Trujillo, 2006). By contrast, articulation causes only a minor pitch change for /k/ and /p/, therefore, referred to as "grave" consonants because they target lower acoustic loci. Thus, this distinction is based on acoustic loci in whistled speech (high vs. low) that mimic those of the spoken word. We also observe an opposition between "continuous" or "semi-continuous" consonants (like /s/) and interrupted consonants (/p/, /t/, k/) (Tran Ngoc et al., 2022). Linguists have used these oppositions to characterize, categorize, and regroup whistled consonants (Diaz, 2017; Rialland, 2005; Trujillo, 2006).

The ability for phoneme categorization in whistled speech by naive listeners has also been extended to words: Tran Ngoc et al. (2023) showed that whistled words could be recognized well over chance (which was at 20%), with 45.6% of correct responses obtained. However, when compared to phoneme categorization rates, participants did not show significant improvements in word recognition. This contrasts with the performances of expert whistlers, where previous experiments have shown word categorization to be closer to 60–75% in whistled Turkish (Busnel, 1970), with an increase of 20–30% of correct answers compared to VCV or CV tokens (Meyer, 2015). Rather, Tran Ngoc et al. (2023) highlight differences for consonant and vowel recognition rates in words, where vowels were much better recognized. The vowel hierarchies (or order of best-recognized vowels) deduced from these results were generally consistent with the hierarchies found with isolated vowels.

In line with these experiments, we propose using whistled speech as a tool to understand speech perception because this type of speech induces a different perception of fully intelligible words or sentences. This change has sometimes been interpreted as an example of "perceptual insight" or of a pop-out in a top-down perceptual process (Meyer, Dentel, & Meunier, 2017), where higher-level knowledge and expectations apply to sounds that can potentially be heard as speech, much like what happens in artificial Sine Wave Speech (see Davis & Johnsrude, 2007; Remez, Rubin, Pisoni, & Carell, 1981). Here, we choose to study the whistled word, thus reprising previous considerations concerning the role of phonemes in words (Benki, 2003; Delle Luche et al., 2014; Tran Ngoc et al., 2023) and whistled word perception as a top-down/bottom-up process. We ask how musical experience will affect the recognition and the relationship between phonemes and words.

The beneficial transfer of musical training toward speech perception can also be considered from both a bottom-up and a top-down approach. First, participating in music engages the musicians in specific auditory skills such as identifying and distinguishing pitches, rhythms, and timbres—skills also required in speech. Thus, by enhancing the listener's ability to discern such acoustic cues, "bottom-up" or "cascade" transfers occur from music to speech (Kraus & Chandrasekaran, 2010), improving, for example, word learning (Barbaroux, Noreña, Rasamimanana, Castet, & Besson, 2020) and the perception of speech in noise (Chandrasekaran & Kraus, 2010). In addition to lower-level enhancements, music and the development of musical skills also engage in cognitive processes like attention, memory, and executive functions (Parbery-Clark et al., 2009; Schellenberg & Weiss, 2013). Several studies

suggest that such skills can combine in a "top-down" transfer to speech perception (Dittinger et al., 2017) and can lead to improved performances in other tasks requiring cognitive abilities in further reaching domains (far transfer, Degé, 2020).

Various models include both fluxes, bottom-up and top-down. Peretz and Coltheart (2003), for example, propose a mechanistic model for sound processing that supports a bottom-up form of transfer, where sounds are initially treated according to a "common acoustic analysis" before feeding into either a music-specific module ("contour analysis"), a language-specific module ("acoustic-to-phonological conversion"), or an as-yet-uncharacterized module ("rhythm and meter"). The shared perceptual capacities and the initially common acoustic analysis could explain a potential crossover between the two perceptual systems, leading to certain advantages. In addition, as each musical instrument has specific acoustic properties, such transfers could vary depending on the instrument played. However, this differentiation between speech and music processes contradicts models which support the multidimensional or "top-down" approach, where transfers are due to interactions between common processing in speech and music rather than separate processes (Besson, Dittinger, & Barbaroux, 2018). In the OPERA hypothesis, Patel (2012) also suggests a way in which transfers occur, highlighting the importance of the overlap between neural networks and the precision of cues used for low-level processing, as well as elements of higher-level processing including the attention given to the task and the emotional value of music that enhanced neural plasticity (Zaatar, Alhakim, Enayeh, & Tamer, 2023). This suggests an interactive process using both top-down and bottom-up mechanisms to stimulate a skill transfer.

The unique form of whistled speech also allows us to consider the role of articulatory and acoustic cues in speech processing (bottom-up cues), as well as elements of top-down processing through the "perceptual insight" provided by the modified speech form. Furthermore, considering both acoustic and articulatory cues in speech perception addresses a crucial issue that opposed Motor theory with Acoustic theory of speech perception. In Motor theory (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967 and Liberman & Mattingly, 1985), speech perception is based on the matching of articulatory gestures to one's own articulatory representation of sound, thus relying on a knowledge of speech production for perception. However, according to Acoustic theories, speech perception would use acoustic cues as tools for speech perception without considering production (Fant, 1960). In whistled speech, though the articulatory cues found in modal speech are used in production, the acoustic realization of these forms reduces the complex signal of modal speech. This, therefore, modifies the relationship between articulatory and acoustic cues found in modal speech.

Here, we target the effect of musical experience and different types of instrumental specialization more specifically, comparing naive listeners to participants with a knowledge of whistled speech. These forms of experience add complexity to the participants' relationships with acoustic and articulatory cues as well as their cognitive processes. This leads to several possible analyses, enabling us to explore the relationships between cues with different focuses/insights. Indeed, the participants who have an expert level of experience with whistled speech (expert whistlers) have a full knowledge of whistled articulatory and acoustic cues, which is similar to their knowledge of modal speech, even though they may be unfamiliar with the whistled words heard. This contrasts with instrumental knowledge,

where musical skills (including instrument-specific acoustic and production training) differ from whistled speech skills, though they have been shown to transfer. Moreover, on musical instruments, the production of the sound is very different from speech production: percussive instruments (such as piano) rely mostly on speed, string instruments use bow pressure and speed, and wind instruments, as a slight exception, often use spoken articulation in their productions. Thus, by exploring and comparing instrument specialization, we contrast different types of musical articulation and acoustic training, further differentiating forms of experience. However, musician participants, even if unfamiliar with the whistled speech mode, are native French speakers. As words are whistled in French, they are, therefore, very familiar with the words used in the experiment, though only in the modal spoken form. These forms of experience enhance the divide between the different roles played by acoustic and articulatory elements in whistled speech perception, and, by exploring these different elements, provide insight into theoretical speech perception models.

In this study, we focus on the same whistled French words as Tran Ngoc et al. (2023) and consider the effect of musical experience on word recognition and phoneme correspondence. In doing so, we assume that the natural, yet modified, whistled speech form represents a relevant tool to investigate perceptual processes in language, and more specifically the impact of expertise—such as musical experience—on speech perception. We then measured speech processing according to musical instrument specialization (for violin, piano, flute, and voice) to detect differences in perception according to specific instrumental production and perception skills, and compared this form of expertise to that of expert whistlers.

We ask the following questions: What is the effect of musical experience on whistled word perception? What is the contribution of specific vowels and consonants to word recognition? If an advantage exists between participants with musical experience compared to nonmusicians, how do expert musicians' skills compare to those of highly trained whistlers? Are musical advantages for whistled words specific to individual instrument specializations? Finally, how do these results fit into theoretical approaches to speech perception?

In this paper, we address these questions through the presentation of a single behavioral experiment studying whistled word categorization. We include two different analyses: one which considers participants according to three groups of musical skill level (None, Low-level, and High-level), and another which compares only the high-level musicians according to targeted instrument specialization, and a group of expert whistlers, all teachers of whistled speech.

2. Experiment

2.1. Method

2.1.1. Stimuli

We selected 24 French words for this categorization task, chosen to include vowels and consonants from previous phoneme experiments, thus enabling us to compare results.

The words selected were disyllabic nouns with a CVCV(C) structure, noted as C1 V1 C2 V2 (C3). These words included only the vowels of interest [i], [e], [a], and [o], which were

equally represented in each vowel position, each appearing six times as the V1 and six times as the V2, providing two occurrences of each V1-V2 combination (a-o, a-e, a-i, o-a, o-e, o-i, e-a, e-o, e-i, and i-a, i-o, i-e). We also selected words that included the four consonants used in previous experiments, [k], [p], [s], and [t]: each appeared both at the start of the word (C1 position), for at least four words, and in the second consonant position, for three words (C2 position). To ensure that words were known by all participants, we controlled their frequency of occurrence in an adult lexicon (Lexique by New & Pallier, 2023). The frequency of occurrence out of 1 million words averages 55.31 ($SD = 180.25$). The completed word list (see Supplementary Appendix, Table 1) fulfills these criteria, though to do so, several other consonants were also present ([b, d, f, ʃ, m] in the initial C1 position and [ʃ, n, l, m, g, ʁ, d, z] in the C2 position).

In adhering to these criteria, we used four recordings of each word: the target consonants /k/, /s/, and /t/ appear 16 times, and /p/ appears 20 times in C1, and each of the target consonants (/k/, /p/, /t/, and /s/) appear 12 times in C2. A single whistler, fluent in whistled Spanish and sufficiently knowledgeable in French to properly pronounce the words, was recorded on a Zoom H1 by the second author. As testing whistled word categorization by naive listeners has never been done before, and as vocalic whistled frequencies also depend on the speakers and their whistling technique (see Meyer, 2015, Tran Ngoc et al., 2020b), we chose to include the natural production variation of only one whistler. It should also be noted that the target vowels and consonants included in this test have similar pronunciations in both French and Spanish, ensuring that difficulties in pronunciation that may appear in second language production will not affect the target sounds. The recordings nonetheless consisted of a spoken version of the word (used to control the pronunciation) followed by the whistled version which was repeated four times.

In the transformation from usual modal spoken speech to whistled speech, the salient characteristics of the spoken word are reflected in the whistled pitches and amplitude modulations. Indeed, as previously observed (Tran Ngoc et al., 2023), the duration of the word in whistled speech (i.e., its elongation) is not correlated with the word duration in modal speech, though it is in agreement with French prosody. Also, while each whistled vowel is produced within a certain pitch range, in the context of the word, the position of the vowel affects the variability of the pitch range (or the stability of the vowel)—where the V2 vowels /e/, /a/, and /o/ are more stable than the corresponding V1 vowels. It also appears that the V1 /o/ is much higher than the V2 /o/ (see Tran Ngoc et al., 2023 for a more detailed description). For consonants, it appears that the consonant cues described in the VCV format (see Rialland, 2005; Tran Ngoc et al., 2022b; Trujillo, 2006) are drastically modified in the context of words, because of elements of co-articulation.

In the C1 position, distinctions between "continuous" consonants and "interrupted" consonants become superfluous, lacking the preceding vowel. Thus, the C1 consonant is better characterized by pitch change (see, e.g., /siʁo/), though we could also consider amplitude rise (e.g., in /tapi/) or articulation points (bilabial—/p/ or velar—/k/, and dental—/t/). In the C2 position, descriptions remain more consistent with previous studies, once again including the opposition between continuous (or semi-continuous /s/) and interrupted (/k/, /t/, /p/). However, the "acute"/"grave" opposition is also affected by the vowel context, modifying the size of the pitch change of acute consonants (e.g., in /kasis/ and /pase/, see Figure 1).

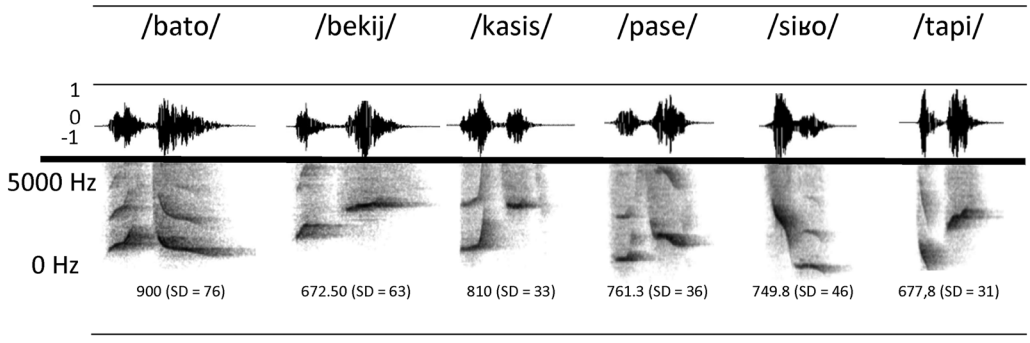

Fig. 1. Wavelength and spectrogram of whistled words representing /k, p, s, and t/ in C1 and in C2

2.1.2. Design

The design for this experiment is identical to that of Tran Ngoc et al. (2023) and aims to evaluate participants' recognition performance for the 24 whistled words. Four different recordings of each word were used, therefore, including natural production variation. For each word heard, participants were proposed five-word options. Given the novelty of this word categorization task, we chose to propose a limited amount of word options to the participants, also maintaining continuity with the previous vowel-focused study. These options included four-filler words from our list of 24 words, which were selected randomly (using https://www.random.org/lists/). We constructed two answer lists which were randomly attributed to each participant. The experiment was conducted online and was programmed using PCIbex Farm. Thus, participation took place at home using headphones, earbuds, or speakers.

2.1.3. Procedure

Before starting the experiment, participants answered a short questionnaire indicating their native language, other languages spoken, their age, and gender. A detailed description of their musical experience was also requested, including the instrument played, a self-evaluated musical "level," and their background in the said instrument (number of years played/context). We asked participants to choose between six different musical levels: 1—beginner (*Débutant*), 2—amateur, 3—confirmed (*Confirmé*), 4—DEM musical Diploma (*DEM*), 5—Superior University Diploma (*Diplôme Supérieur*), and 6—professional musician (*Professionnel*). If participants had no musical experience, they were asked to leave this section blank.

Once the questionnaire was completed, the experiment format was presented to participants by showing an example of a whistled word, /pate/ ("pâté"), as well as the drop-down answer menu. When the experiment began, participants heard a randomly selected whistled word from the list and had to pick the corresponding word among the five choices suggested (which included the correct answer) from the drop-down menu. They were then asked to validate their answer, and the next whistled word was played immediately afterward. Thus, participants first heard the word and then viewed the possible responses. This was done for the 96 target words.

2.1.4. Participants

Ninety-three participants were included. They were all French speakers and had no language impairments or hearing problems. The participant group included 53 women and 40 men, who were between 18 and 50 years old, with an average age of 27.52 years old ($SD = 6.18$). All participants spoke at least one other language with an average proficiency level of 2.39 ($SD = 0.66$) out of 3, where participants rated their proficiency as either 1 (beginner), 2 (intermediate) to 3 (confirmed). Within this group, 18 participants had no musical experience whatsoever, 2 participants declared having a "beginner" level, 11 participants declared being "amateur" musicians, 22 participants declared being "confirmed" musicians, 16 participants had obtained the "DEM" or "Musical Studies Diploma," 6 participants had obtained the "Superior Musical Diploma," and 18 participants were professional musicians. Their average ages per group were as follows: 26.5 years old ($SD = 4.81$) in the nonmusician group, 24.5 years old ($SD = 2.12$) among beginners, 27.36 years old ($SD = 7.59$) among amateurs, 27.05 years old ($SD = 6.26$) among confirmed musicians, 26.5 years old ($SD = 4.66$) among DEM-level musicians, 23.66 years old ($SD = 2.87$) among those with the superior diploma, and 31.77 years old ($SD = 7.19$) among professionals.

In addition, and in order to have a control group, we asked seven expert whistlers with low levels of musical experience to complete the task. Four participants had no musical experience, one was a beginner and two were amateur musicians. This whistler group consisted of native Spanish speakers with a basic knowledge of French, who had no language or hearing problems. Their average age was 41.12 years old ($SD = 6.62$).

This experiment was conducted in accordance with the Helsinki Agreement.

2.2. Results

In a first analysis, we included only 93 naive French speakers. Overall, participants categorized whistled words correctly with an average of 57.4% of correct responses obtained ($SD = 15.53$), well over chance at 20%. When considering the percentage of correct responses obtained per word according to musical experience (Levels 0, 1, 2, 3, 4, 5, and 6), we observe a strong increase in overall correct answers according to Level, going from 46.58% at Level 0 ($SD = 10.68$), to a maximum of 71.70% at Level 5 ($SD = 14.33$). This progressive increase shows a significant positive correlation between level and correct answer rate per word (Pearson's correlation $r(91) = .36$, $p < .001$), suggesting that there is an advantage for participants with musical experience, see Fig. 2. We then further investigate the effect of musical experience at the phoneme level.

To consider the effect of musical experience, we differentiate participants with no musical experience (Level 0), called "None," from participants with some musical experience (Levels 1, 2, and 3), defined through self-evaluation, called "Low," and participants with a high level of musical experience—defined through a musical diploma. This final group (called "High") thus regroups Levels 4, 5, and 6. We observe that even the group None obtained correct word answers above chance, with 46.58% of correct answers ($SD = 10.9$). Participants with a low level of musical experience (Low) recognized words with 58.08% of correct answers ($SD = 14.04$), and participants with a high level of musical experience (High)

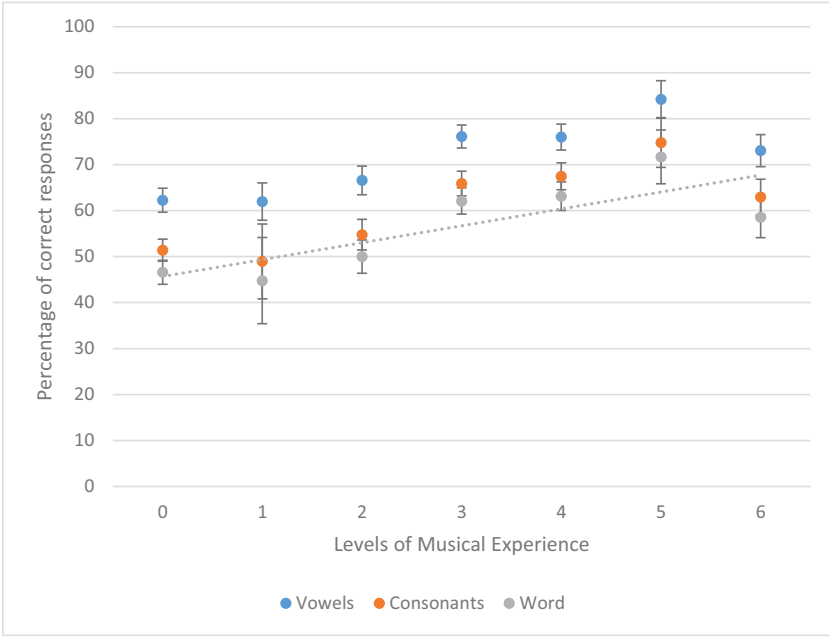

Fig. 2. Percentage of correct word responses and phoneme correspondences (all phonemes) with standard error, obtained according to levels of musical experience, with trend for word responses.

obtained 62.37% ($SD = 15.5$). The amount of correct answers obtained, therefore, increases according to musical experience.

We then considered both correct and incorrect answers, and the correspondence between phonemes in played and answered words according to position (1 or 2). We focused on the phonemes of interest (/a,e,o,i, k,p,s,t/). There were 96 vowels played in V1, and 96 vowels played in V2 for each of the 93 participants. Therefore, we considered 17,856 elements of vowel data points. For the consonants, those of interest (/k,p,s,t/) appear a total of 68 times each in C1 and 48 times each in C2 for each of the 93 participants, amounting to consonant 10,788 data points. Thus, in total, we considered 28,644 data points. We applied a Generalized Linear Mixed Model (GLMM) to phoneme correspondence, with Phoneme Type (Consonant, Vowel), Position (1, 2), and Musical Experience (None, Low, High) as fixed factors. We included Participants and Words as random effects.

We find a significant main effect of Phoneme Type ($X^2(1, N = 93) = 41.019$, $p < .001$) showing that correspondence rate is higher for vowels (at 68.4%) than for consonants (at 64.1%). We also observe a significant effect of Musical Experience $X^2(2, N = 93) = 15.095$, $p < .001$) and a significant interaction between Phoneme type*Musical Experience ($X^2(2, N = 93)$, $p = .039$). Post-hoc tests (using Bonferonni correction) reveal that the difference between vowel correspondence rates (V) and consonant correspondence rates (C) is significant only for the two groups of musicians: low-level musicians (V = 69.1% vs. C = 63.1%) and high-level musicians (V = 72.3% vs. C = 68.8%; $ps < .001$). This is not the case for the

group of nonmusicians. We observe significant differences between the high-level musician group (High) and None for both vowels (72.3% vs. 58.3%; $p < .001$) and consonants (68.8% vs. 55.6%; $p = .006$). For vowels, we also observe a tendency for difference between Low and None (69.1% vs. 58.3%; $p = .054$). Overall, we did not find an effect of Position (ps $> .05$).

These results suggest that musical advantages for vowel and consonant recognition within the word increase with experience, with a stronger advantage for high-level musicians over nonmusicians, than for low-level musicians over nonmusicians. In light of these results, we wish to further explore the effect of a high-level of musical experience by defining high-level musicians' knowledge according to their instrument specialization.

In a second analysis, we focused only on participants with a "high-level" of musical experience, to explore the impact of instrument specialization more precisely. We targeted four instrument groups among the high-level musicians for which we had more than five individuals: violin, piano, flute, and voice; and retained only the high-level musician participants who played these instruments, reducing the total number of participants from the previous analysis. Among these high-level musicians, six were singers (Voice), seven were flutists (Flute), eight were pianists (Piano), and seven were violinists (Violin). In addition, and in order to better characterize the performance of instrumentalists, we included the performances of a group of seven expert whistlers with low levels of musical experience who have a fluent knowledge of whistled speech in Spanish (Silbo). This amounts to a total of 35 participants.

In this second analysis, as in the first one, we considered the eight phonemes included in the words: /a,e,i,o,k,p,s,t/. We did not include the factor Position as this showed no significant effect in our first analysis. We applied a GLMM to phoneme correspondence with Phoneme (/a,e,i,o,k,p,s,t/) and Group (Violin, Piano, Voice, Flute, Whistler) as fixed factors. We included Participants and Words as random effects.

We find a significant main effect of Phoneme ($X^2(7, N = 35) = 79.6, p < .001$), where the correct correspondence rates of the vowels are at 87.8% for /i/, 80.3%, for /o/, 76% for /e/, and 73.8% for /a/. The consonant correspondence rates were at 74.7% for /k/, 74.1% for /t/, 73.8% for /p/, and 67.4% for /s/. We also observe a significant main effect of Group ($X^2(4, N = 35) = 39.4, p < .001$). When considering the performance of each of the groups, we observe that overall flutists obtain 75% correct phoneme correspondences, singers 73.4%, violinists 70.9%, and pianists 70.7%. The whistlers show a much higher performance rate with 94.7% correct correspondences obtained. The Phoneme*Group interaction is significant ($X^2(28, N = 35) = 66.9, p < .001$), and we applied post-hoc tests to specific comparisons of this interaction, using the Bonferroni correction. We observe significantly different profiles for each of the groups present, underlined by differences with the whistlers, see Figs. 3 and 4.

These differences also highlight the presence of some phonological hierarchies according to the group. Compared to pianists, the whistlers perform significantly better for every phoneme except for /o/ (/i/, $p < .05$; /e/, $p < .001$; /a/, $p < .001$; /k/, $p = .01$, /s/, $p < .001$; /t/, $p = .002$). There are no significant differences between phonemes among the pianist group. When compared to singers and violinists, we observe significant advantages for whistlers for two vowels and two consonants: /a/ (ps $< .001$), /e/ (ps $< .05$), /s/ (ps $< .001$), and /t/ (ps $< .05$). There were no significant differences between the phonemes for singers, while, for violinists, we observed significant differences among the vowels, where /i/ is better rec-

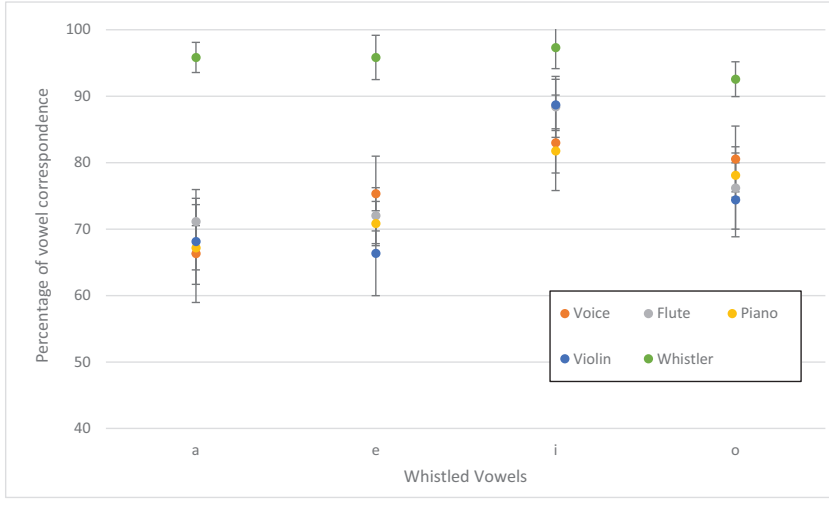

Fig. 3. Whistled vowel correspondence for each instrument group, shown with standard error.

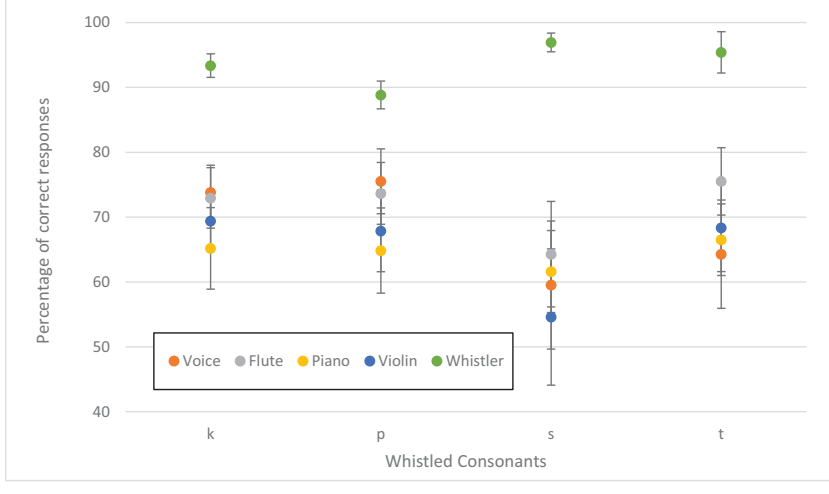

Fig. 4. Whistled consonant correspondence for each instrument group, shown with standard error.

ognized than each of the three other vowels: /e/ ($p < .001$), /a/ ($p < .001$), and / o/ ($p = .004$). Finally, for flutists, whistlers show a significant advantage only for two vowels and one consonant: /a/ ($p = .031$), /e/ ($p = .004$), and /s/ ($p < .001$). We also observe significant differences among the vowels within the flutist group, where /i/ is better recognized than /e/ and /a/ ($ps < .05$) and tends to be better recognized than /o/ ($p = .065$). Whistlers show no significant differences in performance between or among phoneme correspondence rates.

These differences underline how each of the high-level instrumentalists is significantly different from the whistlers for /a/, /e/, and /s/. Flutists are the only instrument group to show differences solely for these three phonemes. For each of the other instrumentalists, whistlers show significantly more advantages: on /t/ for singers and violinists and on /t/, /k/, and /i/ for pianists, see Figs. 3 and 4.

3. Discussion

In this study, we considered the impact of musical experience on whistled word recognition among naive listeners. Participants perform well over chance (at 20%), with a categorization rate of 57.4%, and we observe an increase in the percentage of correct responses obtained according to musical level, supported by a significant positive correlation between musical level and correct responses. This led us to contrast three groups of participants according to musical experience while considering phoneme correspondence between played and answered words.

Groups show a gradual improvement in response rates. We find that participants with a low level of musical experience perform better than those with no musical experience (10.74% increase), and those with a high level of musical experience perform better than those with a low level of musical experience by 5%. When comparing these three groups, we also took into consideration the differences between vowels and consonants, extending the analysis applied in Tran Ngoc et al. (2023), by this time including levels of musical experience. These results, consistent with those obtained previously, that is, the advantage for vowels over consonants, specify the effect of musical experience according to vowels and consonants. Indeed, low-level musicians show a 10.8% advantage for vowels over nonmusicians; and high-level musicians show an advantage over nonmusicians of 14% for vowels and 13.2% for consonants. This suggests a continuous improvement in musical skills according to experience and thus an incremental increase in skill transfer according to musical level, as suggested by Smit, Rathcke, and Keller (2023).

We further explored high-level musicians' behavior by specifying their knowledge according to musical instrument expertise and comparing it with expert whistlers. This comparison underlined several important differences. First, there was a large performance gap between whistlers and expert musicians, where knowledge of whistled speech (even with a limited amount of experience with French) produced results with almost 100% accuracy (an average of 94.7% of correct phoneme correspondences). Such a performance surpasses any musically related transfer toward whistled word perception, where the highest performing musical instrument group (flutists) obtained a lower score by 19.7%. This suggests that though musical experience may create some perceptive advantages, more targeted training (such as learning to whistle speech and recognize phonemes in whistles) has a much stronger effect. Such results were to be expected from the literature (see for a review, Neves, Correia, Castro, Martins, & Lima, 2022), which shows a higher effect of task-specific training than of musical transfer.

Second, the differences observed between high-level musicians and whistlers highlighted different profiles according to instrument specialization, characterized by differences in phoneme correspondence rates. Each instrument group showed a different response profile compared to whistlers: flutists showed differences for only three phonemes (two vowels and one consonant), violinists and singers for four phonemes (two vowels and two consonants), and pianists for six phonemes (three vowels and three consonants). These profiles suggest that behavior varied according to instrument specialization, where flutists were most similar to whistlers, and pianists were the least similar. Most notably, in this word recognition task, the flutists behave similarly to whistlers for three consonants (/k/, /t/, and /p/). This may reflect instrument-specific similarities in terms of articulation or timbre that do not exist in other instruments. For example, the use of tongued plosives in flute playing with consonants such as /t/ and /k/ resembles that of whistlers, thus highlighting similar production mechanisms. Singers also produce such articulatory consonant movements; however, the acoustic sound quality of sung productions are further away from those of whistled speech. Indeed, neither the violin nor the piano emulates speech-like sounds while playing; however, the musical articulations on the violin produce transitions that are closer to whistled speech, contrary to those of the piano which are limited due to instrumental constraints, and thus furthest away from whistled speech (Bresin & Battel, 2000).

The difference between phoneme correspondence rates also shows specificities linked to instrument specialization. When considering the vowels, we underline the advantages for /i/: This vowel, systematically categorized better than other vowels in previous studies (Tran Ngoc et al., 2023; Tran Ngoc, Meyer, & Meunier, 2020a), shows significant differences with the other vowels (/a/, /e/, and /o/) for violinists and flutists (though only a tendency for /o/). Interestingly, difficulties for /o/ in the context of the word (as shown in Tran Ngoc et al., 2023) also seem present for expert whistlers, as their performances do not differ from those of musician participants. This could be due to the large variability of production of /o/ as documented in the Method section.

When considering the whistled consonant correspondence in the word, all of the high-level musician participants show significant differences with whistlers for /s/, and almost all instrumentalists show a difference for /t/ (except flutists). In addition, all of the instrument groups performed equivalently to whistlers for /p/, and generally behaved similarly for /k/ (except pianists). Interestingly, this consonant hierarchy is reversed compared to the one observed in the VCV context used in previous studies, in which the highest categorization rates were observed for /s/ and /t/ and the lowest categorization rates for /p/ (see Tran Ngoc et al., 2022b). Thus, when we consider the whistled cues characterizing these consonants, consonant correspondence in the word underlines a clear opposition between "acute" (/s/ and /t/) and "grave" (/k/ and /p/) consonants, where "acute" consonant correspondence is more difficult. The difference observed between the word context and the previously studied VCV format may be due to word-specific influences such as varying vowel co-articulation (rather than maintaining the same vowel /a/ in both positions), producing inconsistent acoustic cues most notably among the acute consonants.

While these results depend on only a small number of participants in each instrument group, thus only beginning to explore the effect of musical instrument specialization, these

findings suggest how musical expertise is specific to the type of instrument training received, including knowledge of certain timbres and articulation/production mechanisms. We observe how acoustic improvements through musical experience show an effect on whistled speech perception generally, suggesting that these acoustic cues are important for whistled speech perception. This, therefore, supports acoustic theories of speech perception. However, the two groups who also use a knowledge of articulatory elements in the production of whistled timbres, flutists and whistlers, show greater advantages. This suggests that whistled speech perception also uses articulatory cues in addition to acoustic elements. Thus, instead of favoring the Motor Theory or Auditory/Acoustic theories, we suggest that these findings provide support for theories that establish a connection between articulation and acoustics, such as the PACT theory (Schwartz, Basirat, Ménard, & Sato, 2012). Indeed, in such an approach, the perceptive unit is characterized by both its articulatory gestures and its acoustic role. This would explain how improved knowledge of both articulatory and acoustic elements in whistled speech (as is the case for flutists and whistlers), therefore, produce the largest advantages. In addition, expert whistlers' performance over native French-speaking musicians and non-musicians suggests a stronger effect of whistled transposition cues rather than higher-level word knowledge. These results, therefore, suggest the presence of bottom-up transfers toward speech perception.

The results observed also highlight behavioral differences between phoneme categorization and word categorization, as well as between vowels and consonants. This is first shown by the difference between whistled vowel and consonant correspondence rates, where vowel rates are higher than those of consonants. Consonant correspondence rates are, however, much closer to those of the whistled word (see Fig. 2). In previous findings, we observed increased categorization rates for words with larger vowel intervals, Tran Ngoc et al. (2023), consequently suggesting a top-down approach to word perception, as identifying the interval considers the relationship between vowels at a word level rather than as individual phonemes. The higher vowel correspondence rates observed here, which echo findings in modal speech (see Fogerty & Humes, 2010 and Delle Luche et al., 2014), may also rely on the construction of these relationships within the word. Yet, through our findings, we also highlight an important bottom-up musical transfer according to acoustic/articulatory cues. Furthermore, the similarity between consonant correspondence and word categorization rates could also reflect a bottom-up approach, where the ability to categorize consonants directly affects word choice. Thus, the different treatment of vowels and consonants in the word may suggest that both top-down and bottom-up approaches come into play, echoing suggestions by Patel (2012), Dittinger et al. (2017), and Besson et al. (2018).

4. Conclusion

In conclusion, musical experience is shown to impact disyllabic whistled word categorization through an incremental increase in categorization rates, where high-level musician participants show a larger difference in behavior than participants with no musical experience. Interestingly, when comparing high-level musical experience according to instrument

specialization with that of expert whistlers, each instrument shows a different profile. These differences generally highlight a very stable response rate for the vowels /i/ and /o/, and a strong impact of the whistled word context on the consonants, where the hierarchy of consonant correspondence is reversed compared to the hierarchy found in the previously tested VCV form. Overall, our results clearly show an advantage of musical experience on whistled speech processing for all musicians, with specific profiles depending on the instrument played. This allows us to gain further insight into perceptive approaches used by naive listeners to categorize whistled speech.

Acknowledgments

This work was supported by the doctoral grant of the Université Côte d'Azur attributed to Anaïs Tran Ngoc, and the framework of the project GEOTELELING with the support of the CNRS 80Prime interdisciplinary program. We would like to thank all the participants of this experiment, with a special thank you to the expert musicians (from various conservatoires in France) as well as our experts in whistled Spanish – all volunteer teachers for the Asociación Cultural y de Investigación de lenguajes silbados Yo Silbo which currently revitalizes this practice in Canary Islands. Finally, we sincerely thank David Diaz Reyes for having produced the whistled French words used in this experiment.

References

Barbaroux, M., Noreña, A., Rasamimanana, M., Castet, Ec., & Besson, M. (2020). From psychoacoustics to brain waves: A longitudinal approach to novel word learning. *Journal of Cognitive Neuroscience*, *33*(1), 8–27.

Benki, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica*, *60*, 129–157.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, *2*.

Besson, M., Dittinger, E., & Barbaroux, M. (2018). How music training influences language processing: Evidence against informational encapsulation. *L'année Psychologique*, *3*(118), 273–288.

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112–125.

Bresin, R., & Battel, G. U. (2000). Articulation Strategies in Expressive Piano Performance Analysis of Legato, Staccato, and Repeated Notes in Performances of the Andante Movement of Mozart's Sonata in G Major (K 545). *Journal of New Music Research*, *29*(3), 211–224.

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

Busnel, R. G., & Classe, A. (1976). *Whistled languages*. Berlin: Springer-Verlag.

Chandrasekaran, B., & Kraus, N. (2010). Music, noise exclusion and learning. *Music Perception*, *27*, 297–306.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*, 132–147.

Degé, F. (2020). Music lessons and cognitive abilities in children: How far transfer could be possible. *Frontiers in Psychology*, *11*.

Delle Luche, C., Poltrock, S., Goslin, J., New, B., Floccia, C., & Nazz, T. (2014). Differential processing of consonants and vowels in auditory modality: A cross-linguistic study. *Journal of Memory and Language*, *72*, 1–5.

Diaz, D. (2017). *El lenguaje silbado en la Isla de El Hierro* (The Silbo Language on the Island of El Hierro), 2nd Edition. Tenerife, Le Canarien ediciones, La Orotava, Spain.

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast Brain Plasticity during Word Learning in Musically-Trained Children. *Frontiers in Human Neuroscience*, 11.

Fant, G. (1960). *Acoustic theory of speech production*. The Netherlands, Mouton: The Hague.

Fogerty, D., & Humes, L. E. (2010). Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors. *Journal of the Acoustical Society of America*, *128*, 3114–3125.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews: Neuroscience*, *11*, 599–605.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

McPherson, L., & Winter, Y. (2022). Editorial: Surrogate languages and the grammar of language-based music. *Frontiers in Communication*, *7*, 838286.

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. *Frontiers in Psychology*, *8*, 25.

Meyer, J. (2021). Environmental and linguistic typology of whistled languages. *Annual Review of Linguistics*, *7*, 493–510.

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P. A. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, *20*(1), 56–60.

New, B., & Pallier, C. (2023). Lexique. Retrieved from http://www.lexique.org/

Neves, L., Correia, A. I., Castro, L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience & Biobehavioral Reviews*, *140*.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience*, *29*(45), 14100–14107.

Patel, A. (2012). Why would musical encoding benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*, 947–950.

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

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

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (pp. 499–550). Elsevier Academic Press.

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336–354.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681.

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 & Science*, *6*.

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. *Frontiers in Psychology*, *2*, 113.

Tran Ngoc, A., Meyer, J., & Meunier, F. (2020a). Categorization of whistled consonants by French speakers. In *INTERSPEECH 2020 –21th Annual Conference of the International Speech Communication Association, September 14–18, Shanghai, China, Proceedings* (pp. 1600–1604).

Tran Ngoc, A., Meyer, J., & Meunier, F. (2020b). Whistled vowel identification by French speakers. In *INTERSPEECH 2020 –21th Annual Conference of the International Speech Communication Association, September 14–18, Shanghai, China, Proceedings* (pp. 1605–1609).

Tran Ngoc, A., Meyer, J., & Meunier, F. (2022a). Bénéfices de la pratique musicale sur la catégorisation de la parole sifflée: Analyse des processus de transferts. In *Proceedings XXXIVe Journées d'Études sur la Parole – JEP 2022* (pp. 405–413).

Tran Ngoc, A., Meunier, F., & Meyer, J. (2022b). Testing perceptual flexibility in speech through the categorization of whistled Spanish consonants by French speakers. *JASA Express Letters*, *2*, 095201.

Tran Ngoc, A., Meunier, F., & Meyer, J. (2023). Effect of whistled vowels on whistled word categorization for naive listeners. In *Proceedings INTERSPEECH 2023* (pp. 3063–3067).

Trujillo, R. (2006). *El silbo gomero. Análisis lingüístico*. Interinsular Canaria.

Varnet, L., Wang, T., Peter, C., Meunier, F., & Hoen, M. (2015). How musical expertise shapes speech perception: Evidence from auditory classification images. *Scientific Reports*, *5*, 14489.

Zaatar, T. M., Alhakim, K., Enayeh, M., & Tamer, R. (2023). The transformative power of music: Insights into neuroplasticity, health, and disease. *Brain, Behavior, and Immunity - Health*, *35*, 100716.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: Whistled words chosen and tested and target consonant.