INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Whistled vowel identification by French listeners

Anaïs Tran Ngoc[1], *Julien Meyer*[2], *Fanny Meunier*[1]

[1]Université Côte d'Azur, CNRS, BCL, France
[2]Université Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France
anais.tran-ngoc@etu.univ-cotedazur.fr, julien.meyer@gipsa-lab.fr, fanny.meunier@unice.fr

Abstract

In this paper, we analyzed whistled vowel categorization by native French listeners. Whistled speech, a natural, yet modified register of speech, is used here as a tool to investigate perceptual processes in languages. We focused on four whistled vowels: /i, e, a, o/. After a detailed description of the vowels, we built and ran a behavioral experiment in which we asked native French speakers to categorize whistled vowel stimuli in which we introduced intra- and inter- production variations. In addition, half of the participants performed the experiment in person (at the laboratory) while the other half participated online, allowing us to evaluate the impact of the testing set up. Our results confirm that the categorization rate of whistled vowels is above chance. They reveal significant differences in performance for different vowels and suggest an influence of certain acoustic parameters from the whistlers' vowel range on categorization. Moreover, no effect or interaction was found for testing location and circumstances in our data set. This study confirms that whistled stimuli are a useful tool for studying how listeners process modified speech and which parameters impact sound categorization.

Index Terms: vowel categorization, whistled speech, whistled languages, speech perception, acoustic cues

1. Introduction

Whistled speech is a type of natural speech, which transposes spoken speech into whistles (see [1] for a review). At least 40 low-density and remote populations have adapted their local language to this particular speech modality, using it for long distance communication. Notably, whistled speech is intelligible only to trained speakers, and is not directly comprehensible to naive listeners even if they are fluent in the language that is being whistled [2].

Transposition from spoken speech to whistled speech in most non-tonal languages relies on a 'formant-based whistling strategy'[1]. Whistlers make an approximation of the vocal tract articulation used in the spoken form to pronounce the whistled phonemes. In Spanish for example, whistled vowels are emitted at different pitch levels depending on the frequency distribution of the whistler's timbre in the spoken modal speech form (i.e., /i/ has a high pitch, /e/ lower, /a/ even lower, and /o/ the lowest [3]).

Previous studies on whistled speech have proved that naive listeners recognize whistled phonemes using acoustic cues. A first experiment conducted in 2008 showed, using different productions from a single whistler, how naive French listeners were able to categorize whistled Spanish vowels /i, e, a, o/ with a mean level of success corresponding to 55% of correct answers [3]. In 2017, a second experiment using the same stimuli showed that the scores varied per vowel: /a/ and /e/ showed the lowest scores (44.1 and 46.9%, respectively), and /o/ and /i/ were recognized best (50.6 and 78.4% of correct categorizations, respectively, with /i/ being significantly different from the other vowels). The authors also took an interest in the impact of listener experience on vowel recognition, finding that one's native language (Spanish, French or Standard Chinese) impacted whistled vowel categorization, though the results of the French and Spanish participants were not significantly different [3,4].

While previous studies on whistled speech have included some intra-talker variability, very few studies have addressed this variability in whistled speech, despite research showing that inter-talker variability in noise has significant effects on spoken speech perception [5]. In addition, a correlation between certain acoustic phonetic properties and listener comprehension has been observed for non-native listeners [6]: talkers with a larger vowel space were, indeed, easier to understand. An experiment displaying a combination of these conditions (native and non-native listeners with inter-talker variability and presented in slight noise) showed similar results with a significant effect of inter-talker variability on intelligibility [7]. These properties, different for native and non-native listeners, include more energy in the 1-3 kHz range, as well as an enlarged vowel space in the F2 range. Interestingly, the stimuli from these experiments deal with certain constraints which also characterize whistled speech (modified speech forms that are first unintelligible for naive listeners) leading us to investigate the impact of acoustic phonetic inter-talker variations in whistled speech perception.

The present paper extends the previous experiments on whistled speech while considering the impact of slight inter-talker variation with several objectives. First, it aims at (a) testing whistled vowel categorization with new whistled stimuli, to assess whether the previous results can be generalized. It then (b) seeks to introduce inter-individual differences (inter-talker variability) in the productions tested, using stimuli from two different whistlers. It also (c) explores the possibility of a learning effect throughout the different parts of the experiment using a transfer-learning model [8] and finally (d) looks at the impact of the testing set up by comparing data acquired in the lab with data obtained online with participants running the experiment from home. This is particularly relevant to the current quarantine period, which prevents many researchers from conducting experiments in laboratories.

To answer these questions, we constructed a three-part experiment. Part 1 asks participants to respond to stimuli without any previous introduction, part 2 proposes a short learning phase where feedback is given, and finally part 3

http://dx.doi.org/10.21437/Interspeech.2020-2697

consists of the same test as part 1, with stimuli from the other whistler. This allows us to evaluate learning by comparing parts 1 and 3. Finally, to test for potential effects of the experiment set up, half of the subjects participated in the experiment in the lab and the other half participated online from their homes.

2. Experiment

2.1. Method

2.1.1. Stimuli

This experiment was conducted in accordance with the Helsinki agreement. The second author recorded the stimuli in a soundproof room of the Gipsa-Lab (Bedei Platform) with two different expert whistlers, both teachers of whistled speech in the Canary Islands. The whistled Spanish vowels /i/, /e/, /a/ and /o/ were extracted from bisyllabic CVCV whistled words (such as /cada/, /nata/…). In order to retain the same prosody for each vowel chosen as a stimulus for the test, we systematically selected vowels from the second CV syllable, on unaccented syllables only. Moreover, we selected vowels following various consonant attacks (/d/, /k/, /g/, /t/), and, after removing the consonant attack, silence was added to the vowels to create homogenous samples of 500 milliseconds.

The extraction of these whistled vowels from CVCV words causes their duration to vary a great deal. As that duration can be discriminated easily for any difference over 100 milliseconds [9], we chose to use whistled excerpts of sufficiently varying lengths (see Figure 1) to ensure that the overall duration differences between the stimuli could not be used to discern the individual vowels. The vowel stimuli therefore last between 146 ms for a whistled /a/ extracted from a /ta/, to 473 ms for a whistled /o/ extracted from a /go/ (both by whistler A). These durations vary according to the vowel, the whistler producing the stimuli (the recordings of whistler A vary more in duration than those of whistler B) and the consonant attack.

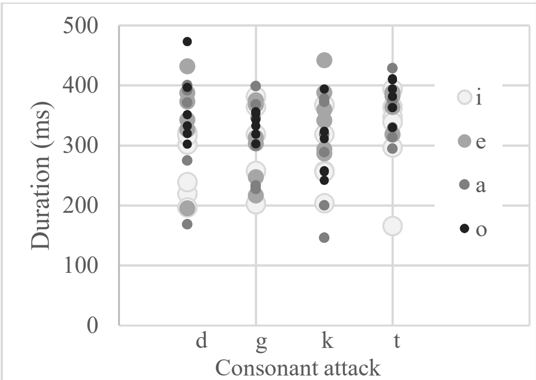

Figure 1: *Whistled vowel duration following consonant attacks.*

In addition, the durations loosely reflect those used in an experiment on vowel length in English [10]. In this experiment, synthesized versions of natural vowels were created with 3 different durations: 272, 144 and 400 ms. Results showed that duration had a small overall effect on vowel identification.

The frequency of the vowel also varies (see Figure 2). This variation is slightly influenced by the consonant attack (as seen in Figure 1) which was removed from the recording and then replaced by a fade-in. However, variation is generally attributed to factors such as the whistlers' physical morphology, vocal range, whistling technique employed for producing the whistle (see [1] for a review), and whistling skills. In line with the previously mentioned experiments [6,7], whistler B, who often teaches Silbo, believes that the further apart the vowel groups are, the easier it is to distinguish and identify them. This also echoes the tendency languages have for maximizing acoustic distances between vowels, often described in linguistic theory [11]. Here, we observed that the frequencies of different vowel positions of whistler A are proportionately less spread out than those of whistler B (Figure 2). This applies to all of the vowels except for [o], where the difference is approximately 50 Hz which can be attributed to morphological variations between the whistlers.

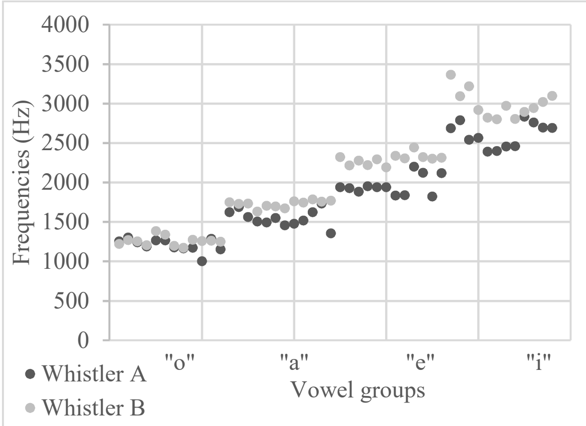

Figure 2: *Distribution of whistled frequencies according to vowel groups per whistler. Whistler A productions are the darker dots.*

Certain vowel groups vary more than others: this is especially the case for high frequency whistled vowels, /i/ and /e/ (see Figure 2 and Table 1). In contrast, /a/ and /o/ are more stable for both whistlers (Figure 2 and Table 1). Overall, not only are the vowel groups of whistler B more distanced from one another, the frequencies are also more stable, reflecting the use of different whistling strategies.

Due to the variation of duration and frequency as well as the importance of relative frequency perception in whistled speech (which relies on modulations of a simple frequency line) participants may need to identify the "range" of the vocalic whistled space of the whistler, which remains proportionately uniform for each individual. The relationship between the vowel frequencies presented in Figure 2 allow us to deduce two linear equations (derived from the values obtained from the linear regression on all the data attributed to each whistler). These equations, based off the average frequency of the vowel group for each whistler, underline the difference in slope and of whistled range, which become more important for the vowels /e/ and /i/. In the linear equations below we considered x to be the position of the vowels, following the order [o, a, e, i] where [o]=1, [a]=2, [e]=3 and [i]=4, and y to be the average vowel frequency. This distribution and vowel order also reflects those of the French or Spanish vowel diagram (or triangle) starting from "back" and "closed" and moving towards "front" and "open". Equation 1 corresponds to whistler A and equation 2 to whistler B.

$$y = 460.9x + 677.01 \tag{1}$$

$$y = 578.37x + 622.36 \tag{2}$$

Despite the difference in slope between the two equations, the general relationship between the vowel groups is around 4/3 of the average frequency of the vowel below it, though slightly lower for whistler A and slightly higher for whistler B.

In this experiment, we maintained the relationship between the whistler and the whistled vowel range by testing the whistlers separately, taking into consideration a possible effect of inter-talker variability and testing whether participants adapt to an individual whistler-specific frequency distribution or a general frequency distribution. In addition, this reflects a more realistic situation concerning the ecological conditions: when whistlers hear each other, they adjust to the other person's range to understand their speech.

2.1.2. Design

We evaluated how naive participants performed on categorizing whistled vowels using one whistler's productions and, after a training section using the vowels of the same whistler, we evaluated how these performances changed when responding to the other whistler's productions. This procedure enabled us to test whether there was an overall learning effect from listening to the first whistler (transfer-learning model), or whether listeners rely on other parameters such as relative frequency perception (similar to the perception of musical notes). This experiment has two versions (one with whistler A first and one with whistler B first) both containing three parts, i.e. part 1 (test), part 2 (training) and part 3 (test).

In part 1, participants listen to 48 whistled vowels, corresponding to 12 versions of each vowel type. These include 3 different recordings of each vowel extracted from the same consonantal context. Part 2, the training session with feedback, comprised 16 vowels, using 4 recordings of each vowel, each corresponding to a different consonant attack. We chose these recordings from the 48 heard in part 1 according to their proximity with the average frequency of that vowel (Table 1).

Whistler	Average frequency of vowels (Hz)			
	"i"	"e"	"a"	"o"
m (A)	2605.02	1958.59	1547.82	1205.61
SD (A)	156.27	123.19	104.59	82.94
m (B)	2995.16	2294.59	1726.85	1256.51
SD (B)	173.98	66.16	44.61	58.79

Table 1: *Average frequency (m) and standard deviation (SD) of vowels according to whistler.*

In part 3, 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). If participants created an abstract representation of the vowel during parts 1 and 2, they should be able to recognize the stimuli from part 3 better than those from part 1.

The online experiment was programmed with PCIbex Farm using headphones, earbuds or speakers at home. The in-person experiment took place in a quiet room in the BCL lab (MSHS, Nice, France), was programmed using PsychoPy, and used Senheiser HD 200 Pro or Senheiser MB360 headphones. All other parameters were identical.

2.1.3. Procedure

Before starting the experiment, we asked participants to indicate the languages they speak and their musical experience. In the online version, participants informed us whether headphones, earbuds or speakers were used, and the corresponding brand. Online, participants were to adjust the volume to a comfortable listening level, in person however, we set the headphones at a fixed comfortable volume.

Part 1: This part presents participants with recordings performed by one of the whistlers. It asks participants to categorize the whistled vowels heard without any training using the arrow keys. The arrow keys are attributed to each vowel following the keyboard layout (both qwerty and azerty), and are presented before and during the experiment (Figure 3).

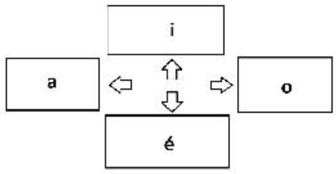

Figure 3: *Arrow keys assigned to each vowel*

Part 2: Participants then complete a short training session with feedback for 4 versions of each whistled vowel. If the participants heard whistler A in part 1, the training used whistler A's recordings. If they heard whistler B in part 1, the training used whistler B's recordings.

Part 3: Finally, 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

Thirty-seven participants were tested for this experiment; they were all native French speakers aged between 19 and 50 years old (m=26.8; SD=8.37). They did not have any language or hearing impairments and did not play any instrument at a high or pre-professional level. Participants gave informed consent before starting the experiment. Seventeen participants completed the experiment in the lab and the other 20 participated online. We recruited the participants online through various social media networks, and in person through the University Côte d'Azur, considering that, once we excluded self-declared speech/hearing impairments, participants did not have any pre-disposed differences in performance.

2.2. Results

In our analyses we took into account the 48 answers given in part 1 and the 48 answers given in part 3 by each participant. Overall, we obtained 53.5 % of correct categorizations out of the 3352 answers given.

We first ran a global repeated measures Anova that included 2 within fixed variables -Vowel type (/a,e,i,o/) and Part (part 1, part 3)- and 2 between subjects fixed variables: Order of presentation (whistler A first, whistler B first) and Experimentation (online, in the lab). We considered the Participant factor to be random.

We first noted that Experimentation (online or in the lab) is never significant (at threshold of .05): neither alone, nor in interaction. We observed a significant effect of Vowel type ($F(3,96)=59.594$; $p<.001$). It appears that /i/ is categorized correctly 86.04% of the time, /o/ 58.95%, /e/ 43.56% and /a/ 38.31%. The interaction Vowel type * Order of presentation *

Phase is also significant (F(3,96)=3.520; p=.02). In order to understand this double interaction we ran two other Anovas, one for each Order of presentation.

When the productions of whistler A are presented in part 1 and the ones of whistler B in part 3 (see results in upper chart of Figure 4), we observed an effect of Vowel type (F(3,51)=44.53; p<.001) as well as a significant interaction between the Vowel type and Part (F(3,51)=3.82; p=.015). We then ran a post hoc test to look at specific comparisons and used a Bonferroni correction (p<.05) in order to perform a multiple comparison test. It appears that specific comparisons between each vowel with itself in parts 1 and 3 are not significant, showing that there is no specific learning for one particular vowel. Within part 1 we observed that /i/ is better categorized than the 3 other vowels /o, e, a/ and that /o/ is different from /a/ which was hardest to categorize correctly. In part 3, significant differences (p<.05) were observed only between /i/ and the 3 other vowels /o, e, a/.

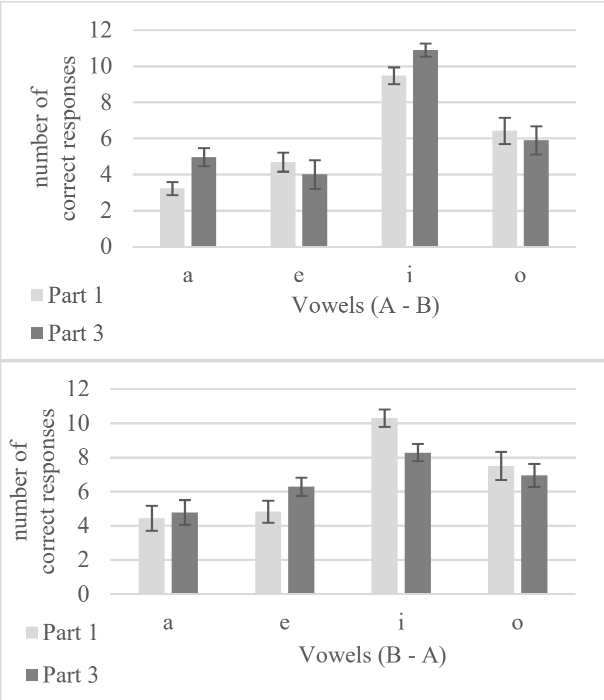

Figure 4 : *Correct whistled vowel categorization when the productions of whistler A are presented in part 1 and the ones of whistler B in part 3 (upper chart) and when (lower chart) the productions of whistler B are presented in part 1 and those of whistler A in part 3.*

When the productions of whistler B are used in part 1 and the ones of whistler A in part 3 (see results in Figure 4, lower chart), we observed an effect of Vowel type (F(3,48)=21.391; p<.001) but no significant interaction between the Vowel type and Part. Running a post hoc test with the Bonferroni correction to look at specific comparisons (p<.05), it appears that /i/ is better categorized than the 3 other vowels /a, e, o/ and that /a/ is different from /o/. The difference between /e/ and /o/ also shows a tendency (p=.07).

3. Discussion

In this experiment, we looked at how whistled vowels are categorized by naive listeners. We aimed at extending previous results to see if they are can be generalized, and introduced inter-individual differences (talker variability) in the productions to see how abstract the representations stored in the brain are, and if certain acoustic phonetic cues allow for better phoneme perception. In addition, we checked for a learning effect throughout the different parts of the experiment. Finally, we explored the effects of online or in-person testing set up.

Overall, whistled vowel categorization was obtained with 53% of correct responses (well over chance 25%), confirming the results obtained previously [3,4]. Having used stimuli from two different whistlers in this experiment, the previous results can be generalized, as they also apply when participants are faced with natural variations of whistled vowels. The vowel specific differences were also replicated [3,4], where /i/ was categorized best and was systematically different from the others vowels, followed by /o/, /e/ and /a/ for which /e/ and /a/ were harder to recognize and were not different from each other. This generalization is also supported by the lack of significant difference found between the results of online and in-lab participants: whistled phonemes are recognized equally well in the two conditions.

The inter-talker variation between whistlers also proved to have an impact, as suggested by the interaction observed in the global analysis. When whistler A (with a smaller vowel space) was presented in part 1, /i/ was better recognized than the other vowels /o, e, a/, and /o/ was distinctive from /a/. Yet this was not the case in part 3 for whistler B where only /i/ was better recognized than the other vowels. When whistler B (with a larger vowel space) was presented first, there was no difference between parts, /i/ being recognized best and /o/ being distinctive from "a" both in parts 1 and 3. In addition, /e/ and /o/ showed a tendency to be different. This suggests that when participants heard whistler B first, the abstract representation of sounds was more easily applicable to whistler A not only for /i/ distinctions, but also for /a/ (and /e/) distinct from /o/. When whistler A was first, these representations only applied to /i/ different from /e/, /a/ and /o/. In line with existing literature [7], our findings suggest that more stable frequencies and larger vowel space facilitate abstract representations of the middle vowels (/e/ and /a/).

Finally, there was no overall learning effect found, though there were some significant differences for specific vowels. This shows that the training portion (part 2) did not systematically help construct an abstract representation of the sounds heard. To better test for the creation of abstract representation, further experiments should be conducted with the same whistler in parts 1 and 3. In addition, to better measure the effect of talker variability, more whistlers should be included in future experiments.

4. Conclusions

In conclusion, naive French listeners recognize whistled vowels between 53 and 55% of the time. These results appear to be robust and generalizable. Our study further showed that the whistler's range and frequency distribution influenced participants' categorization of vowels, and that larger vowel space facilitates the creation of abstract vowel representations.

5. Acknowledgements

We wish to thank the two whistlers for their contribution in whistling words, Jonathan Parente for his help with the vowel experiments in the lab and all the participants for volunteering their time.

6. References

[1] J. Meyer, *Whistled Languages, A Worldwide Inquiry on Human Whistled Speech*, Springer, 2015.

[2] R. Busnel, & A. Classe, *Whistled languages*. Berlin Heideleberg: Springer-Verlag, 1976.

[3] J. Meyer, "Acoustic Strategy and Typology of Whistled Languages; Phonetic Comparison and Perceptual Cues of Whistled Vowels". *Journal of the International Phonetic Association. Cambridge University Press*, vol. 38, no.1, pp. 69-94, 2008

[4] J. Meyer, L. Dentel, and F. Meunier, "Categorization of Natural Spanish Whistled Vowels by Naive Listeners of Spanish, Chinese and French", *Frontiers in Psychology*, 2017.

[5] J. Zaar & T. Dau, "Sources of variability in consonant perception of normal hearing listeners", *The Journal of the Acoustical Society of America*, vol.138, no.3, pp. 1253-1268, 2015.

[6] T. Bent, D. Kewley-Port, S. Fergusson, "Across-talker effects on non-native listeners' vowel perception in noise", *Journal of the Acoustical Society of America*, vol. 128, no.5, pp.3142–3151, 2010.

[7] W. A. Van Dommelen & V. Hazan, "Impact of talker variability on word recognition in non-native listeners". *The Journal of the Acoustical Society of America*. vol. 132, no.3, pp.1690-1699, 2012.

[8] D. Wang & T. F. Zheng, "Transfer Learning for speech and Language Processing", *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA),* 2015

[9] S. A. Gelfand, *Hearing : An Introduction to Psychological and Physiological Acoustics*, CRC Press, 2016.

[10] F. A. Everest, & K. Pohlmann, *Master Handbook of Acoustics*, McGraw Hill Professional, 2009.

[11] J. Hillenbrand, M. Clark, and R. Houde, "Some effects of duration on vowel recognition". *J Acoust. Soc. Am.* vol. 108, no.6, 2000.

[12] A. Martin and S. Peperkamp, "Speech Perception and Phonology", In M. van Oostendorp, C. Ewen, E. Hume & K. Rice (Eds.), *Companion to Phonology. Volume IV.* Hoboken, N.J.: Wiley-Blackwell, pp. 2234-2256, 2011.

[13] A. Tran Ngoc, F. Meunier, and J. Meyer, "Whistled vowel identification by French listeners," in *INTERSPEECH 2020 – 21th Annual Conference of the International Speech Communication Association, September 14-18, Shanghai, China, Proceedings*, 2020, pp.