INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Categorization of natural Spanish whistled vowels by naïve Spanish listeners

Julien Meyer [1,2], *Laure Dentel* [3], *Fanny Meunier* [2,4]

[1] Gipsa-Lab, Département Parole & Cognition, CNRS UMR5216, Grenoble, France
[2] Laboratoire sur le Langage, le Cerveau et la Cognition, CNRS UMR5304, Bron, France
[3] The World Whistles Research Association, Paris, France
[4] Laboratoire Bases, Corpus, Langage, CNRS UMR7320, Nice, France

julien.meyer@gipsa-lab.fr, ldentel.lab@gmail.com, fanny.meunier@unice.fr

Abstract

Whistled speech in a non tonal language consists of the natural emulation of vocalic and consonantal qualities in a simple modulated whistled signal. This special speech register represents a natural telecommunication system that enables high levels of sentence intelligibility by trained speakers. It is not directly intelligible to naïve listeners. Yet, it is easily learned by speakers of the language that is being whistled, as attested by current efforts of revitalization of whistled Spanish in the Canary Islands. To understand better the relation between whistled and spoken speech perception, we looked here at how Spanish native speakers knowing nothing about whistled speech categorized four Spanish whistled vowels. The results show that naïve participants were able to categorize these vowels, although not as accurately as a native whistler.

Index Terms: whistled speech, vowel categorization, speech recognition, Spanish, Silbo.

1. Introduction

Whistled speech is a natural speech register that enables distant communication by transposing spoken languages into whistles (for a review, see [1]). Such a strong reduction of the frequency spectrum of the voice is the reason why whistled speech is language specific, relying on the selection by the whistlers of salient key features of a given language. In this paper we will focus on a non-tonal language, i.e. Spanish (but see [1] and [2] for reviews on whistling in tonal languages). In this type of languages whistlers approximate the oral part of the vocal tract articulation used in spoken form in what is called ‘formant-based whistling’ [1, 2]. This provokes a whistled adaptation of vowel and consonant qualities. For example, in Spanish or Greek and Turkish, whistled vowels are emitted at different pitch levels depending on the frequency distribution of their timbre (i.e, /i/ has a high pitch, /e/ lower, /a/ even lower [3]). Interestingly, the phonetic details that are selected during whistled speech in non-tonal languages are sufficient for trained whistlers – but not for untrained ones - to recognize non-stereotyped sentences as well as to achieve a reasonable degree of word and syllable recognition ([1] for a review).

In the present study we explore how naïve participants categorized whistled vowels. Our results provide new insights into the first steps of whistled speech learning and shed the light on a situation which is quite common nowadays in the Canary Islands where whistled Spanish of La Gomera, which was declared a UNESCO protected oral patrimony in 2009, is taught at school and to volunteer adults for revitalization [1].

There are still few studies that have been conducted on whistled speech recognition or learning. One, using fMRI showed that the brain areas traditionally associated with language are activated in well trained listeners but not in untrained ones [4], while another one, using behavioral technique, showed that the traditionally reported hemispheric lateralization of speech processing is challenged by whistlers’ behavior, as they showed more lateralization in syllable recognition when listening to spoken speech than whistled speech [5]. A third recent study using behavioral measurement showed how naïve French listeners were able to categorize whistled Spanish vowels /i, e, a, o/ quite similarly as does a Spanish trained traditional whistler, even if the whistler does so more accurately [3]. This evidence demonstrated that the cognitive linguistic categorizations used to recognize spoken vowels are easily associated with tonal frequencies by native speakers of a non-tonal language, without any training.

In the present paper we pursue this approach by extending the experiment of [3] to native speakers of the language of the stimuli. Therefore, we use again vowels /i, e, a, o/ of Spanish Silbo - the whistled version of Spanish of La Gomera Island [6] - and we look at how Spanish native speakers knowing nothing about whistled speech categorized these whistled vowels in a simple and intuitive task. We compared these patterns to a reference, i.e. the pattern observed for a native whistler of La Gomera (reported in [3]).

2. Methods

2.1. Stimuli

The four tested vowels from the Spanish whistled language of La Gomera (Silbo) were: /i/, /e/, /a/, /o/. In whistled Spanish, the four vowels /i/, /e/, /a/ and /o/ show a similar frequency scale pattern with four intervals, statistically different, following a decreasing order of mean frequencies; this pattern holds across whistlers [7]. The experimental material consisted of 80 vowels, all extracted from the recording of 20 long semi-spontaneous sentences whistled relatively slowly in a single session by the same whistler in controlled conditions (same whistling technique during the entire session, constant distance from the recorder and from the interlocutor, and background noise between 40 and 50 dBA). These 80 vowels (20 /i/, 20 /e/, 20 /a/ and 20 /o/) were extracted from the stimuli used in [3] that were recorded in

http://dx.doi.org/10.21437/Interspeech.2016-379

2003 by the first and the second authors. The sounds played as stimuli ranged from 1 kHz to 3.7 kHz and concern only the vowel nucleus without the consonant modulations (remember that in whistled speech, vowel nuclei are typically whistled as rather steady in frequency and are modulated at their extremity by the consonant articulation). The selected vowels were chosen inside a confidence interval of 5% around the mean value of the frequencies of each vocalic interval and therefore, the vowel frequency bands of the experiments do not overlap (figure 1). The amplitudes of the stimuli were normalized to equal maximum values. Moreover, the durations at which the vowel nuclei were originally whistled in the sentences was kept for the stimuli. They ranged from 85ms to 1s, with 71% of the vowels below 400ms (corresponding to vowels non lengthened by prosodic effects). The 29% remaining vowels corresponded to vowels lengthened by prosodic effects (as described in the literature on whistled Spanish (e.g. [6], [3]). Altogether, we selected stimuli that represented the variability of pronunciation of the whistled vowels in spontaneous whistled speech.

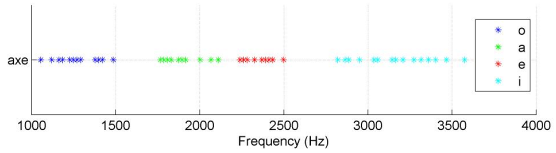

Figure 1: *Frequency distribution of played vowels of the experiments*

2.2. Design and Procedure

The experiment included 3 phases: the first phase was the training to the task and the third was exactly the same, allowing to evaluate the training effect between the beginning and the end of the experiment. They were composed of the same 40 whistle sounds (10 of each vowel, coming from 4 recordings of each vowel), presented in a fixed order. The Phase 2 was the core of the test with a larger and more diverse set of stimuli. It was composed of 64 whistle sounds randomly presented (16 of each vowel, coming from 16 different recording per vowel). They were different from the sounds used in Phases 1 and 3. Overall each participant processed 144 stimuli. The vowels were presented on their own without any context.

The task was the same as the ones used by [3]: listening of a whistle sound followed by a four-alternative forced choice (4-AFC). The participants listened to a whistled vowel and immediately afterwards selected the vowel type that he/she estimated was the closest to the one heard by clicking on one of the four buttons corresponding to the French letters «a», «é», «i», «o».

The test was programmed in Flash-Actionscript, and presented to subjects on a computer in a quiet room using high-quality Sennheiser headphones (HD 449). To start, the participant listened to one whistled Silbo sentence to discover the type of stimuli he/she will have to process. This allowed the check of the volume comfort for the participant at around 70 dB. Then the Phase 1 started, and the participant listened one by one to each stimulus and had to categorize them as "a", "é", "i", or "o" by clicking on the display (written "a", "e", "i", "o" for convenience in the rest of the paper). Clicking started the next trial. Phase 2 and Phase 3 followed directly. Only one listening was possible per stimulus and there was no feedback. Overall the experiment lasted 20 minutes. Sole the type of answer was recorded ("a", "e", "i", or "o").

2.3. Participants

Twenty volunteers (12 women and 8 men) native Spanish speakers living in Lyon (France) that were students at the University of Lyon or teachers at the Instituto Cervantes of Lyon, aged 19-34 years, took part in the experiment. None of the participants had known hearing loss. The present and subsequent experiments were conducted in accordance with the Declaration of Helsinki.

3. Results

We propose several different analyses of correct answers, of confusion matrices of answers (“o”, “a”, “e”, “i”) as a function of the played vowels (/o/, /a/, /e/, /i/), and of correct answers + confusions together by visualizing the answers of the participants as a function of the acoustic frequency of the whistle of each played vowel (see Figure 3). First and in order to have a point of reference we present the whistled vowel identification pattern for a single native whistler of La Gomera as observed previously in [3] for the Phase 2 of the present protocol. Next, we present the answers on the three phases for Spanish naïve participants.

We focus here on the influence of frequency values ofwhistled vowels on the answers as it was the only varying parameter of the stimuli that was found to affect them. Indeed, no influence of duration on answers was found for any vowel type (no significant difference appear when comparing the pattern of answers of naïve participants split into the duration categories described earlier: ‘prosodically lengthened vowels’ vs. ‘non prosodically lengthened vowels’ (X2(9)=0.26, n.s.)).

3.1. Reference Spanish whistler of La Gomera

Table 1 shows the performance on whistled vowel identification by a single native whistler of La Gomera. He reached 87.5% of correct answers confirming that a native whistler practicing nearly daily Spanish whistled speech identifies accurately the four whistled vowels. The vowels where significantly categorized differently [X2(9)=136.97, p<.0001]. The agreement of the answers with the vowel categories was different from chance and not accidental as it was ‘substantial’ - and nearly ‘almost perfect’ - according to Cohen’s Kappa statistics (k=0.79, z=10.94, p<.0001) which give a quantitative measure of the magnitude of such an agreement while being adjusted for agreement due to random chance alone [8], [9].

Table 1: *Confusion matrix for the answers of a native whistler (in %). In Tables 1-2 values in italics correspond to correct answers and values in bold correspond to confusions with neighbouring-frequency vowels).*

		Answered vowels			
		« o »	« a »	« e »	« i »
Played Vowels	/o/	*87.50*	**12.50**	0	0
	/a/	**6.25**	*75*	**18.75**	0
	/e/	0	**6.25**	*87.50*	**6.25**
	/i/	0	0	**0**	*100*

The distribution of the confusion matrix showed that /i/ is perfectly identified, and that the most difficult vowel to recognize for this whistler was the /a/ still reaching 75% of correct identification but categorized as « e » in 18.75%.

3.2. Spanish naïve listeners experiment

3.2.1. Phase 1 and 3: the training effect

We ran an Anova by participants including ‘Vowel played’ (/a/, /e/, /i/, /o/) and ‘Phase’ (1 and 3) as within factors. Analyses have been done on raw data of correct answers however for clarity percentages will be used in the text.

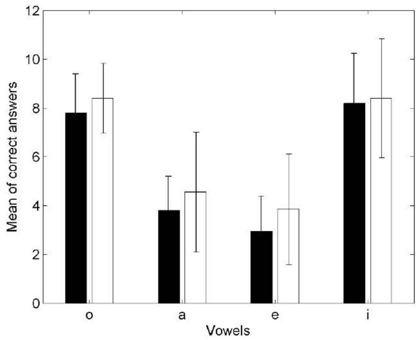

Figure 2: *Correct answers (raw scores) of Spanish speakers as a function of vowels and for Phases 1 and 3.*

The mean level of success corresponding to correct answers was 56.9% in Phase 1 and 60.3% in Phase 3. Considering the protocol and the task with 4 possible answers, these results are largely above chance (25%) for all vowels on both phases and they show that the vowels are categorized differently on both phases ([Phase 1: $X2(9)=676.68$, $p<.0001$], [Phase 3: $X2(9)=824.27$, $p<.0001$]).

There is a significant effect between Phase 1 and Phase 3 (cf. figure 2) showing a training effect of an increase of 3.4 % ($F(1,19) = 5.22$, $p<.05$). It also appears that the mean rates of correct answers varied largely as a function of the vowels played (/a/, /e/, /i/, /o/). For both Phases /e/ is the vowel with the lowest score (29.5% in Phase 1, 38.5% in Phase 3), close to /a/ (with 38% in Phase 1 and 45.5% in Phase 3). The performance is much higher for /o/ (78% in Phase 1 and 84% in Phase 3) and /i/ (82% in Phase 1 and 84% in Phase 3). This effect is significant on both Phases ($F(3,57)=84.29$, $p<.0001$). The interaction between Vowel and Phase is not significant ($F(3,57)=0.55$, n.s) showing that training does not benefit differently to vowel categorization.

3.2.2. Phase 2

In table 2 we find the distribution of answers in a typical confusion matrix showing the answered vowels as a function of the played vowels of Phase 2. Considering the protocol and the task with 4 possible answers, these results show a different categorization for the 4 vowels [$X2(9)=833.97$, $p<.0001$]. The agreement of the answers with the vowel categories was still different from chance and not accidental as it was almost ‘moderate’ according to Cohen’s kappa (k) statistics ($k=0.37$, $z=22.62$, $p<.0001$).

Table 2: *Confusion matrix for the answers of 20 untrained Spanish native speakers (in %).*

		Answered vowels			
		« o »	« a »	« e »	« i »
Played Vowels	/o/	*62.5*	**20.63**	14.06	2.81
	/a/	**18.13**	*35.31*	**36.87**	9.69
	/e/	4.69	**25.31**	*35.31*	**34.69**
	/i/	1.56	4.38	**16.25**	*77.81*

Correct answers

The mean level of success corresponding to correct answers was 53%. An Anova has been performed on correct responses recorded during the Phase 2 with ‘Vowel played’ (/a/, /e/, /i/, /o/) as a within factor. In accordance with what was observed previously in Phase 1 and 3, the scores varied significantly depending on vowels ($F(3,57)=22.42$, $p<.0001$). Again /a/ and /e/ gave the lowest scores (35.1%), while /o/ and /i/ are best recognized with 62.5% and 78 % of correct categorizations respectively. Post hoc multiple t-tests with Bonferroni correction ($p<.05$) confirmed that /a/ and /e/ are less well recognized than /i/ and /o/, with no difference within each group of vowels. As we will see in the next paragraphs, the analysis of the confusions allows us to better understand this pattern of answers.

Confusions

First we looked at confusions between vowel types. Next, in order to determine the influence of the individual frequency of each played vowel on the pattern of answers, the answers were presented as a function of the frequency distribution of the whistled vowels, with estimated curves of the answers that appear averaged by polynomial interpolations of the second order (figure 3).

The values in bold of Table 2 correspond to the confusions with vowels which are frequency neighbors. We ran a multiple t-tests with Bonferroni correction ($p<.05$) on all answers to explore vowel effects on confusion patterns. We first observed that the vowel /o/, that represents the extreme category of lower whistled frequencies, is significantly different from its unique direct whistled vowel neighbor /a/, despite the levels of confusions on both sides of the matrix (/o/ is answered “a” in 20.63% of the cases of confusions of /o/ and /a/ is answered “o” in 18,13% of the cases of confusions of /a/). By contrast, intermediate vowels /a/ and /e/ are largely taken for one another up to the point of not being significantly different in both directions (/a/ answered “e” and /e/ answered “a”). Finally, confusions between /e/ and /i/ show an asymmetry depending of the vowel that is played: /i/ is answered “e” in 16.25% but “e” and “i” answers remain statistically different when /i/ is played, whereas /e/ is much more often answered “i” (34.69%) and the test shows no statistical difference between “e” and “i” answers when /e/ is played.

Finally, it appears that the extreme vowels (/i/ and /o/) follow an effect of distance: the more the frequencies of the vowels are far from each other’s, the less they are mistaken for others. For example, for /o/, the rates of confusions are as follows “a” (20,63%) > “e” (14,06%) > “i” (2,81%).

To provide a more detailed view of the perceptual results, the collected data were also presented on Figure 3 with the details of the influence of the frequency of the played vowels on the answers. On this figure, the estimated curves of the answers appear averaged by polynomial interpolations of the second order. For Spanish naïve participants, the pattern of answers shown by this figure as a function of the frequency distribution of the stimuli shows that the maximum of estimated curves is always within 5% of variance of the range of variation of the vowels and therefore within the range of stimuli distribution. Moreover, extreme frequency vowels /i/ and /o/ show concave estimated curves reflecting lower degrees of confusions with neighboring vowels than correct answers. Together, these criteria show that the listeners categorize the vowels accurately in accordance with the vowel production, even for /a/ and /e/ despite their high inter confusions. Therefore, this graphical representation is accurate to show that naïve participants are accurate in identifying whistled vowels of their own language.

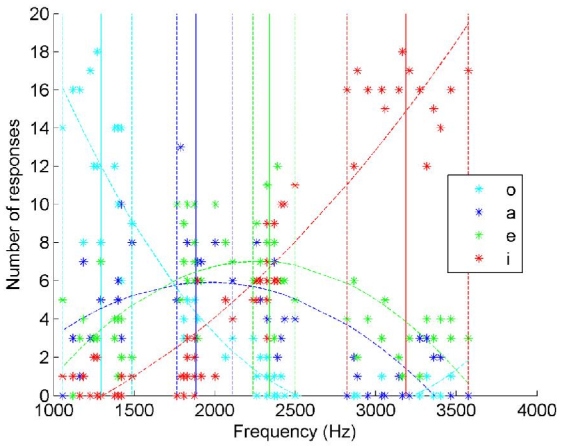

Figure 3: *Distribution of the answers of 20 Spanish non whistlers as a function of the frequencies of the Spanish whistled vowels.*

4. Discussion

In this experiment we asked naïve Spanish speaker participants to categorize whistled vowels of Spanish Silbo, the whistled version of Spanish of La Gomera Island. The experiment was made of 3 phases. The first and the third were similar and used repeated whistled vowels, while phase 2 used different and unique productions of whistled vowels.

As expected naïve Spanish speakers categorized whistled vowels much better than chance, reaching 53 % of correct categorization (phase 2). This score appears however to be much less than the trained whistler who scored 87.5% of correct responses. This is in line with the training effect observed with our naïve participants (comparing phases 1 and 3): after only 15 minutes of training and with no feedback they showed an increase of 3.4 % of correct categorizations. These results showed that naïve participants can categorize quite efficiently whistled vowels (learning is rapid and easy), which suggests that the recognition of whistled segments relies on perceptual properties that may already be exploited in the comprehension of spoken segments.

Looking in more details it appears that the pattern of categorization is very similar between naïve Spanish participants and the trained whistler. Expertise appears to be of particular interest for "middle" frequencies vowels /a/ and /e/, but also in a less strong way for extreme frequency vowels /i/ and /o/. For naïve participants /a/ can be equally categorized "a" or "e" and /e/ as "e" or "i", showing that these whistled vowels are particularly difficult. Expertise seems to play an important role in this disambiguation as the trained whistler doesn't show such confusions. What appears striking is that, except for /i/ that is perfectly categorized, the trained whistler makes confusions only with adjacent vowels, but never with further ones, contrarily to naïve participants that can make any type of confusions (even /i/ is taken for "o" in 1.56%).

The results of the naïve Spanish speakers are very similar to the ones obtained by French speakers, in [3], who obtained 55% of correct responses (compared to 53% for Spanish). This suggests that there is no gain to have the exact same language as L1 and the whistling and that proximity between French and Spanish vowels is enough to allow French to reach scores similar to Spanish speaker ones. Further analysis should compare more carefully confusion patterns between French and Spanish. Moreover, further experiment should explore categorization by speaker of a tonal language to see how it modulates categorization patterns.

5. Conclusions

We found that whistled vowels of Spanish Silbo can be categorized better than chance by naïve Spanish speakers and that learning takes place rapidly, suggesting that the system relies on perceptual properties already present in spoken speech and that may be already exploited in the comprehension of spoken sequences. This study also shows that the natural practice of whistled speech is an adequate scientific object to further examine speech perception.

6. Acknowledgements

The authors would like to thank J. Polge and L. Duchamp and the participants for their collaboration during testing. This research was financially supported by the 7th research program of the European Community FP7 2007-2013 (Marie Curie IIF Grant 630076).

7. References

[1] J. Meyer, *Whistled Languages - A Worldwide Inquiry about Human Whistled Speech*. Berlin: Springer, 2015.

[2] A. Rialland, "Phonological and phonetic aspects of whistled languages", Phonology, vol. 22, no. 2, pp. 237–271, 2005.

[3] J. Meyer, "Typology and acoustic strategies of whistled languages: Phonetic comparison and perceptual cues of whistled vowels", Journal of the International Phonetic Association, vol. 38, no. 1, pp. 69–94, 2008.

[4] M. Carreiras, J. Lopez, F. Rivero and D. Corina, "Linguistic perception: neural processing of a whistled language", Nature, vol. 433, no. 7021, pp. 31–32, 2005.

[5] O. Güntürkün, M. Güntürkün, C. Hahn, "Whistled Turkish alters language asymmetries", *Current Biology*, vol. 25, no.16, 2015.

[6] A. Classe, "The whistled language of La Gomera", *Scientific American*, vol. 196, pp. 111-124, 1957.

[7] J. Meyer, Description Typologique et intelligibilité des langues sifflées: approches linguistique et bioacoustique, Université Lyon2 : Thèse de doctorat, 2005.

[8] Cohen, J., "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, vol. 20, pp. 37-46. 1960.

[9] Y. Bishop, S. Fienberg and P. Holland, Discrete *Multivariate Analysis – Theory and Practice*, Cambridge, Massachusetts: MIT Press, 1975.