Identification of Natural Whistled Vowels by Non-Whistlers

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

[1] Laboratoire Dynamique Du Langage (DDL-CNRS), Lyon, France
[2] Laboratori d'Aplicacions Bioacústiques, Universitat Politècnica de Catalunya, Barcelona, Spain
[3] Centro Politécnico Superior, Universidad de Zaragoza, Spain

julien.meyer@univ-lyon2.fr, fanny.meunier@univ-lyon2.fr, laure.dentel@theworldwhistles.org

Abstract

Whistled speech consists of a phonetic emulation of the sounds produced in spoken voice. This style of speech is the result of the adaptation of the human productive and perceptive intelligence to a language behavior. In the typology of whistled forms of languages, Spanish is among the languages for which the whistled strategy emulates primarily segmental acoustic cues of vowels and consonants. The present study tests the perception of four Spanish whistled vowels by French non-whistlers. The results show that French non-whistlers were able to categorize these vowels without any learning, although not as accurately as native whistlers.

Index Terms: speech perception, vowel identification, whistled speech, whistled languages, speech model.

1. Introduction

Whistled speech of languages represents a little studied style of speech shaped for distant dialogues [1]. It consists of an emulation of the spoken voice based on modulated whistles and thus called 'whistled language'. In the typology of whistled languages, Spanish whistled strategy emulates primarily segmental acoustic cues of vowels and consonants [2]. In order to deepen the understanding of the perception of such natural whistled vowels and their links to the vowels of the spoken speech, two variants of a same experiment were developed. We observed the whistled vowel categorisation performances for subjects who knew nothing about whistled languages. The performances of a native whistler of Spanish whistled language (Silbo) are also provided for reference. Participants had to recognize the four vowels /i, e, a, o/ in a simple and intuitive task. They did not receive any feedback on their performances or any information concerning the distribution of the whistled vowels before the end of the test. The first experiment tested the vowels presented on their own, without any context (Experiment A) while the second experiment tested the vowels presented in the context of a sentence (Experiment B).

2. Method

2.1. Participants

The tested subjects were 40 students of 19 to 29 years old who were French native speakers. 20 persons have performed Experiment A (vowels on their own), and the 20 others Experiment B (sentence context). The students' normal hearing thresholds were tested by audiogram. The participants had never listened to whistled speech before or to an explanation of the distribution of the vowels in non tonal whistled speech.

2.2. Stimuli

The tested four vowels from the Spanish whistled language of La Gomera (called Silbo) are: /i/, /e/, /a/, /o/. These vowels also exist in French with similar or close pronunciations [3]. Another reason for this choice of four whistled vowels was that they have the same kind of frequency distribution in Greek and Turkish [2, 4]. Given the structure of French, one can reasonably expect that whistled vowels of French would bear the same scale. The experimental material consisted of 84 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 84 vowels (21 /i/, 21 /e/, 21 /a/ and 21 /o/) were chosen by taking into account statistical criteria based on the analysis of whistled vowels in Silbo [2]. First, the mean frequency of a whistled vowel is much easier and precise to measure than spoken formants as each occurrence of a whistled vowel is a rather stable and simple pitch. Next, each vowel is whistled in a specific interval of frequency values corresponding to the articulatory variability of the concerned vowel type. Silbo vocalic system is the result of a whistled emulation of the Spanish dialect spoken in the Canary Islands in which /o/ and /u/ are often assimilated [5]. The spoken vowels (i, e, a, o, u) are therefore whistled in five frequency intervals, two of which largely overlap. Four intervals are statistically different: /i/, /e/, /a/ and /o, u/ in a decreasing order of mean frequencies. Moreover, we excluded final vowels of sentences from the vowels presented in our experiments as they are often marked by an energy decrease. Finally, the selected vowels were chosen inside a confidence interval of 5% around the mean value of the frequencies of each vocalic interval. In this sense, the vowel frequency bands of the experiments do not overlap

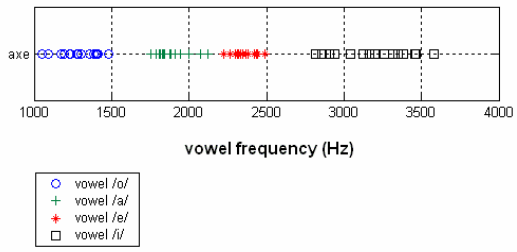

Figure 1: Frequency distribution of the played whistled vowels of the experiments.

The sounds played in the Experiment A concerned only the vowel nucleus without the consonant modulations, whereas the stimuli of the corpus of Experiment B kept the part of whistled sentence preceding the vowel during two to three seconds. This second experiment aimed at testing the effect of the acoustical context on the subject and to get rid of some bias that might appear because of presenting nearly pure tones one after another. As a consequence, this second corpus consisted of 84 whistled sentences ending by a vowel. For both variants, among the 84 sounds, 20 (5 /i/, 5 /e/, 5 /a/, 5 /o/) were dedicated to a training phase and 64 (16 /i/, 16 /e/, 16 /a/, 16 /o/) to the test itself.

2.3. Design and procedure

For each experiment, participants listened to whistled vowel and right afterward 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 task was therefore a four alternative forced-choice (4-AFC). The interface, programmed in Flash-Actionscript, controlled the presentation of the sounds: first the 20 sounds of the training phase in an ordered list containing all the possible successive combinations of vowels; then, the successive 64 sounds of the test in a non-recurrent random algorithm. The subjects where tested in a quiet room with high quality Sennheiser headphones.

3. Results

A specific program was developed to summarize the answers in confusion matrices (tables 1, 2, 3, 4) and to present them graphically by reintegrating some information regarding for example the frequency distribution of played vowels (figure 2 and 3).

3.1. Reference performances of a whistler

In table 1 we present the performances of whistled vowel identification of a native whistler of La Gomera (Experiment A on isolated vowel, which represents the most difficult task). These performances confirm with a high level of correct answers (87,5 %), that a native whistler practicing nearly daily the Spanish whistled speech identifies accurately the 4 four whistled vowels [$X^2(9)=136.97$, $p<.0001$] [4]. The variability of pronunciation of the vowels in spontaneous speech and the distribution of the played vowels (figure 1) explain the few confusion errors.

Table 1. Confusion matrix for the answers of a native whistler for isolated vowels (in %)

	Answered vowels				
Played vowels		/o/	/a/	/e/	/i/
	/o/	*87.5*	**12.50**	0.00	0.00
	/a/	**6.25**	*75.00*	**18.75**	0.00
	/e/	0.00	**6.25**	*87.5*	**6.25**
	/i/	0.00	0.00	**0.00**	*100.00*

3.2. Identification of isolated vowels (Experiment A)

The mean level of success corresponding to correct answers was of 55%. Considering the protocol and the task these results are largely above chance (25 %) [$X^2(9)=900.39$, $p<.0001$)]. But the mean rates of correct answers varied largely in function of the vowels. Moreover, most of the confusions can be qualified of logical in the sense that a vowel was generally confused with its frequency neighbouring vowels (83% of the cases of confusion: bold letters in table 2).

Table 2. Confusion matrix for the answers of 20 subjects for isolated vowels (in %).

	Answered vowels				
Played vowels		/o/	/a/	/e/	/i/
	/o/	*50.63*	**40.31**	7.50	1.56
	/a/	**13.44**	*44.06*	**31.56**	10.94
	/e/	5.94	**22.19**	*46.88*	**25.00**
	/i/	0.00	4.38	**17.19**	*78.44*

In order to determine the influence of the individual frequency of each played vowel on the pattern of answers of the subjects, the results of the answers have also been presented in function of the frequency distribution of the whistled vowels presented during the experiment (figure 2). In this figure, appear also the estimated curves of the answers of the subject, averaged by polynomial interpolations of the second order.

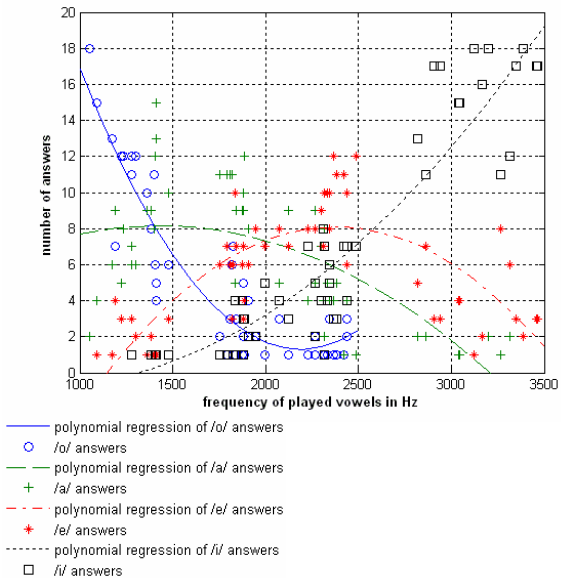

Figure 2: Intuitive perception of the isolated Spanish whistled vowels by 20 French subjects (distribution of the answers in function of the played frequencies).

3.2.1. Individual variability

Two participants have very high performances with 73,5 % of correct answers. Then a group of six persons has more than 40 correct answers for 64 sounds (62,5%). Four other persons follow them with more than 58% of correct answers. This means that half of the participants have good performances to the test. The ten other participants all have performances above 37% but below 54% of correct answers.

Generally speaking, the less efficient participants still had a confusion matrix with logical confusions. Their relatively low performances were often due to a systematic confusion between two whistled vowels neighbour at the level of the frequencies. The variability of the performances depended on the concerned vowel: (i) for /i/ most of the participants had very good success results as 16 of them obtained a score over

75% and two even 100% of correct answers. The less efficient participant reached a rate of 56% of correct answers. (ii) For /o/, six persons identified more than 62,5% of the vowels. All the others often thought the /o/ is an /a/. (iii) The /a/ was the least well-identified letter, often mis-categorised as an /e/ or sometimes as an /o/. (iv) The /e/ was confused equally with its whistled neighbours /a/ and /i/.

The lower performances for /a/ and /e/ can be partly explained by the fact that they both have two perceptive neighbours in terms of pitch, a situation which multiplies the possibilities of confusion in comparison to the more isolated vowels /i/ and /o/. In spite of this situation, the most efficient participants categorized them successfully as different vowels throught the pitch they perceived. Finally, the more frequent confusions were the following: the /o/ was often thought of as an /a/ and the /a/ and the /e / were reciprocally often mistaken one another, partly because of the proximity of their intervals of whistling.

3.2.2. Musicians vs. non-musicians

Among the subjects of this experiment, six were musicians. The results of this group were significantly different from the non-musicians [F (1,18)=6.71, p<.02]. One can therefore say that the musicians had more success to the task of the test than the non-musicians (64% of correct answers versus 51%).

3.2.3. Discussion

All the analysie detailed above support the fact that the French subjects were able to categorise the whistled vowels « a », « é », « i », « o », however they were not as accurate as a whistler from la Gomera Island [p<.001]. Nonetheless, the tendencies of the curves of correct answers show that the French-speaking subjects in general have good performances to the task.

These results occur despite presenting isolated vowels without any sound context except the one of the preceding listened vowel. Indeed, some participant performances revealed some effect of one vowel on the following answer. For example, if for a whistled /e/ an answer /a/ was given and if the following played vowel was an /a/ the participants had the tendency to mistake it with /o/. Consequently, one can observe a cascading effect of logical confusions that stops when there is a significant frequency jump. This confirms that non-whistlers subjects perceptively floor their vowel prototypes in a distribution that depends on the frequency. In these conditions, it is not surprising to note that the musicians have better performances because they are more used to associating an isolated pitch to a sound reference culturally marked. Despite randomisation of item presentation this cascading effect is difficult to control with only four types of vowels,. For this reasonm, we developed Experiment B.

3.3. Identification of vowels with preceding sentence context (Experiment B)

This second experiment aimed at testing the effect of the context on vowel perception. Particularly we hypothesized that by an approach nearest to the ecological conditions of listening of whistlers -who do not perceive the vowels isolated but integrated in a sound flux – one could observe a suppression of the cascading effect of confusions.

The results show the same general tendencies as for Experiment A, with slightly better performances for the identification task: 60,2% [X^2(9)=1201.63, p<.0001]. The whistled vowels /o/ and /i/ were even better identified than in the Experiment A (respectively 73,13% and 87,81%) whereas the vowels /a/ and /e/ were slightly less well identified (see table 3 for percentages and figure 3 for estimated curves of answers).

Table 3. Confusion matrix for the answers of 20 subjects for isolated vowels (in %).

	Answered vowels				
Played vowels		/o/	/a/	/e/	/i/
	/o/	*73.13*	**23.13**	2.81	0.94
	/a/	**10.94**	*39.06*	**39.38**	10.63
	/e/	5.00	**19.38**	*40.94*	**34.69**
	/i/	0.31	1.56	**10.31**	*87.81*

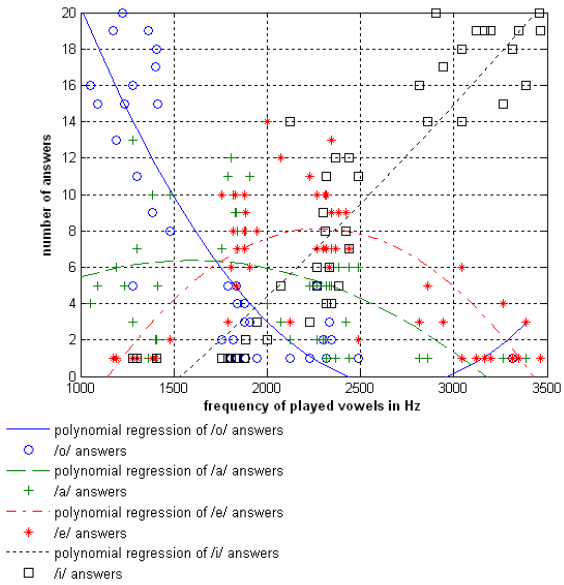

Figure 3: Distribution of the answers in function of the frequencies of the whistled vowels. Intuitive perception of the Spanish whistled vowels by 20 French subjects (vowels with preceding context).

3.3.1. Confusions and inter-individual variations

Eight persons had a success score above 62,5%. 3 other even have more than 73%. The best subject, a saxophonist who has learned the musical pitches through a particular technique of localisation of the sounds by association to a part of his body reached an overall score of 75%. In contrast, the least efficient participant obtained a score of 46% of identification, which was more than in Experiment A. For the confusions, the better scores for /o/ showed that it is less thought as an /a/, whereas /a/ was still often confused with /e/. Finally, and this is new, /e/ was often thought of as an /i/ with strong differences between participants. It was often at the level of the identification of /a/ and /e/ that differences were found between subjects with high scores and subjects with lower scores.

3.3.2. No difference between musicians and non-musicians

Again, there were six musicians among participants (despite the fact that the participants of the two experiments were distinct). An analysis of variance similar to the one performed for Experiment A showed that this time the results of the

musicians were not significantly different from those of non-musicians [F (1,18)=6.71, n.s.]. The context effect eased the choice of the non-musicians without affecting the performances of the musicians.

3.3.3. Limited effect of training

Because of the elimination of the confusions specific to Experiment A, due to successive presentation of isolated vowels, it is relevant in Experiment B to compare the performances of the test to the performances of the training phase in order to see if there is a learning effect. One can note that again answer distribution is far from chance [$X^2(9)=1113.47$, $p<.0001$] and the tendencies described in the test were already at play in the training (table 4). These results were obtained in a first contact to whistled vowels with only 20 occurrences of vowels. As a consequence, it supports the fact that the subjects rely on categorizations that are already at play in their linguistic use.

Table 4. Confusion matrix for the answers of 20 subjects in training phase, listening to whistled vowels in context (%)

	Answered vowels			
Played vowels	/o/	/a/	/e/	/i/
/o/	*59.00*	**22.00**	13.00	6.00
/a/	**20.00**	*32.00*	**36.00**	12.00
/e/	6.00	**16.00**	*49.00*	**29.00**
/i/	2.00	5.00	**9.00**	*84.00*

4. Discussion

The results obtained for the two experiments of identification show that participants whose native language has the same types of vowels as the Spanish /i, e, a, and o/, succeed in categorizing the whistled emulations of these vowels without any cues on the phenomenon of whistled languages and even if they listened to such whistled sounds for the first time. The fact that this ability is already stable during the training phase shows that the tested subjects were already familiar with a perceptual representation of the vocalic space in a frequency scale. This suggests that such a representation must play an important role in the process of identification of the spoken French vowels /i, e, a, and o/.

To explain these results, one must consider which parts of the frequency spectrum of the spoken voice carry perceptual cues explaining the frequency distribution of whistled vowels in non tonal languages. It appears to be a key of vowel identification, even for languages with larger vocalic inventories like Turkish [4]. Some results obtained on the basis of production support that the second formant (F2) plays an important role: Brusis [6] was the first to notice that F2 often shows frequency shapes similar to the transposed whistled signal. On this basis, Rialland [7] proposed that only F2 is transposed in Silbo. But F2 may well be only one of the parameters of a more complex phenomenon. By using a protocol based on perception, our experiments draw the attention on this option. For example, a distribution of vowels in frequency scales is characteristic of other perceptual studies based either on the notion of *perceptual integration* between close formants [8], or on the notion of *« effective upper formant »* (F2') [9]. According to Stevens [10] these notions highlight strong effects at play in the classification of vowels. For example, Schwartz and Escudier [11] showed that a greater formant convergence explains better performances of vowel identification and a better stability of vowels in short-term memory. Concerning Spanish, the human ear is sensible to the convergence of F3 and F4 for /i/, F2 and F3 for /e/ and F2 and F1 for both /a/ and /o/. The fact that /a/ and /o/ share the same formant convergence would participate to the fact that /i/ is identified with better performances than /o/ (/i/ and /o/ being with only one neighbour in whistled frequency).

5. Conclusions

Our results confirm that the whistlers rely on a perceptual reality at play in spoken speech to transpose the vowels in whistled frequencies. For research on Silbo it confirms the observation of Classe [5] that at least 4 whistled vowels can be phonetically distinguished by whistlers of La Gomera and therefore invalidates another theory often denounced by local whistlers that stated that only two vowels are perceived in Silbo [12]. Finally our experiments prove that whistled speech is a good model to further investigate the processes of perceptual selection in the complex distribution of vowel formants.

6. Acknowledgements

The authors would like to thank the whistlers and the participants for their collaboration. This research was supported by the CNRS and the Fyssen Foundation.

7. References

[1] Busnel, R-G. and Classe, A. Whistled languages. Springer Verlag. Berlin. 1976.

[2] Meyer, J., Description typologique et intelligibilité des langues sifflées : approche linguistique et bioacoustique. Ph. D Thesis. University Lyon2. 2005. http://www.lemondesiffle.free.fr/whistledLanguages.htm

[3] Calliope, La parole et son traitement automatique. Masson, Paris. 1989.

[4] Meyer, J., "Whistled Turkish: statistical analysis of vowel distribution and consonant modulations". ICPhS XVI Proc. 2007.

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

[6] Brusis, T., "Über die phonetische Structur der Pfeifsprache Silbo Gomero dargestellt an sonagraphischen Untersuchungen". Zeitschrift für Laryngologie, 52, 292-300. 1973.

[7] Rialland, A., "A New Perspective on Silbo Gomero", ICPhS 15 Proc., 2131-2134. Barcelona. 2003.

[8] Chistovitch, L. A. and Lublinskaja, V. V., "The center of gravity effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel like stimuli". Hearing research, 1, 185-195. 1979.

[9] Carlson R., Granström B. and Fant, G., "Some studies concerning perception of isolated vowels", STL-QPSR 2-3/1970, 19-35.

[10] Stevens K. N., Acoustic Phonetics. Cambridge: MIT Press. 1998.

[11] Schwartz, J. L. and Escudier, P., "A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception". Speech Com. 8, 235-259.1989.

[12] Trujillo, R., El Silbo Gomero: analisis linguistico. Ed. I. Canaria-I. Andres Bello, Tenerife. 1978.