Acoustic strategy, phonetic comparison and perceptive cues of whistled languages

Julien Meyer

Acoustic strategy, phonetic comparison and perceptive cues of whistled languages

Julien Meyer

Laboratori d'Aplicacions Bioacustiques (LAB), Universitat Polytecnica de Catalunya(UPC), Avda. Rambla Exposició s/n, 08800 Vilanova i la Geltrú, Barcelona, Espanya

julien.meyer@lab.upc.edu

Abstract

Whistled speech is a complementary natural style of speech to be found in more than 30 languages of the world. This phenomenon also called "whistled language" enables distant communications in the background noise of rural environments. Whistling is used as a sound source instead of the vibrations of the vocal chords. The resulting acoustic signal is characterized by a unique functional narrow band of frequencies encoding the words. Such a strong reduction of the frequency spectrum of the original spoken voice explains why whistled speech is language-specific, relying on selected salient key features of a given language. However, for a fluent whistler, a spoken sentence transposed into whistles remains highly intelligible. That's why whistled languages represent valuable sources of information to phoneticians. This study is based on new original data collected in 7 different cultural communities or gathered during specific perceptive experiments. Whistling is first found to prolong the strategy at play in shouted voice in order to increase the possible distance of communication. Then an analysis of the linguistic behaviours of various types of whistled speech is presented, with an emphasis on non-tonal languages. Finally, whistled vowels of non-tonal languages are shown to be the reflection of the perception of formant proximities. (198 words)

1. Introduction: a style of speech concerning a large diversity of languages

Its users treat whistled speech as an integral part of a local language as it fulfils the same aim of communication as spoken speech while encoding the same syntax and vocabulary. Its function is to enable dialogs at middle or long distance in conditions where the normal or the shouted voice, covered by the ambient noise, wouldn't be intelligible. The linguistic information is adjusted and concentrated into a phonetic whistle thanks to a natural oral acoustic modification of the voice that is shown in this study to be similar but more radical to what occurs in shouting. The whistled signal encodes selected key traits of the given language through modulations in amplitude and frequency. It is sufficient for the brain of trained whistlers to recognize non-stereotyped sentences. For example non-words could be recognized in 70% of the cases in Busnel (1970) and sentences are intelligible with regular performances of 90 % (Busnel 1970, Meyer 2005). Contrary to a "language surrogate" whistled speech does not create a substitute for language with proper rules of syntax or even a proper repertoire of codes, and contrary to Morse code it doesn't rely on an intermediary code, like the written alphabet. In 1976, Busnel and Classe explained: "*when a Gomero or a Turk whistles, he is in fact still speaking, but he modifies one aspect of his linguistic activity in such a way that major acoustic modifications are imposed upon the medium*". All the whistlers interviewed for the present paper underlined that they whistle exactly what they think in their language and that an equivalent process is at play when they receive a message. They confirmed that: *«at the receiving end, the acoustic signal is mentally converted back into the original verbal image that initiated the chain of events"* (Busnel et Classe, 1976: 107). In brief, whistled speech is a true *style of speech*. The point of view of the pioneers in the study of whistled languages converge also to the definition of the whistled form of a language as a style of speech. Cowan observed in 1948: "*The whistle is obviously based on the spoken language*"and described the high degree of intelligibility and variability of the sentences in whistled Mazatec (Cowan 1948). Later he said about whistled Tepehua: *«The question might be well asked, if whistled Tepehua should not be considered a style of speech (as whisper is for example), rather than a substitute for language"* (Cowan 1972). Busnel and Classe found the classification of whistled languages among "surrogates" as improper: "*Whereas the sign language of deaf-mutes, for instances is truly a surrogate since it is a substitute for normal speech, whistled languages do not replace but rather complement it in certain specific circumstances. Rather than surrogates, they are adjuncts.*" (Busnel et Classe, 1976 : 107). The direct drawback is that any language could be whistled, provided that the ecological and social conditions favor such a linguistic behavior. Indeed, this phenomenon covers a very large diversity of languages and of linguistic families. Tonal languages (Mazateco, Hmong,,..) as well as non tonal languages (Greek, Spanish, Turkish,…) are concerned. The present study has localized whistled languages that extend the known linguistic structures that have been adjusted into whistles (Akha, Yupik, Surui, Gaviao, Mixteco…). Even incipient tonal languages (Chepang) are concerned[1].

[1] It is important to notice that the fieldwork practice consisting in asking to a speaker to whistle the tones of his language in order to ease their identification by a linguist cannot be called a whistled language. Yet, this

In this article a broad approach of the phenomenon of whistled languages is first given by explaining their acoustic strategy and the drawbacks for different types of linguistic systems. Then a comparative description of productive data of whistled speech based on non-tonal languages is developed in an unprecedented statistical analysis. Finally, an original experiment of whistled vowels identification by persons knowing nothing about whistled languages gives new insight of the perceptive relevance of formant convergence.

2. A telecommunication system in continuity with shouted voice

2.1. From spoken to shouted voice…towards whistles

Nearly all the low-density populations, which have developed a whistled language, live in biotopes made of mountains or dense forests. Such ecological middles favor the isolation of the inhabitants during several vital activities of their everyday life: shepherding, hunting, harvesting in the field... On one hand a rough topography increases the occasions to speak in the distance; on another hand, the dense vegetation restrains the visual contact and the sound propagation in a noisier environment. Usually, to increase the distance range of their normal voice or to cover noise, individuals raise its amplitude level in a quasi-unconscious way. This phenomenon, called "Lombard effect", has been described in the beginning of the century (Lombard 1911). During this process the spoken voice progressively passes to the register of shouted voice. But, if the noise or the distance to cover is continuously increased, the shouter will soon tire his vocal chords and reach their biological limit. The effort is all the more strong as the tendency is to prolong the syllables and reduce the flow of speech (Handley and Steer 1949). That is why most of the shouted dialogs are short. For example, in a definite natural mountain environment like a valley in the Vercors (France), the limit distance of intelligibility of the normal spoken voice has been measured to be under 50 m (figure 1 and figure 2) while the limit distance of intelligibility of several shouted voices produced at different amplitude levels could reach up to 200 m (figure 2) (Meyer 2005). At this limit distance of 200 m, the tiring of the vocal chords was reached around 90-100 dBA. The experiment consisted in recording a male shouted voice targeted at reaching a person situated at distances going from 20 to 300 meters. The acoustic strategy at play in shouted speech showed a quasi-linear increase of the frequencies of the shouted formants and a lengthening of the duration of the sentences (figure 2 and figure 3).

fieldwork technique underlines that the real whistled languages represent a high potential of interest for the description of languages because it has highly contributed to develop the modern phonology these last 30 years.

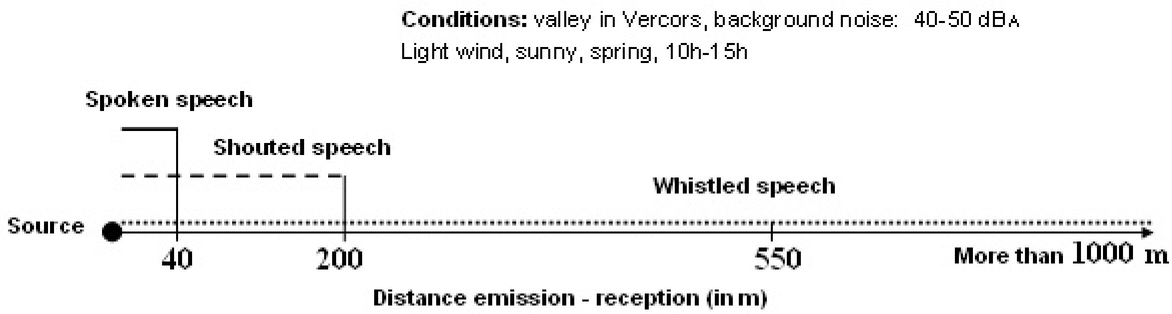

Figure 1 : Typical limit distances of intelligibility of spoken, shouted and whistled speech in the conditions of the experiment.

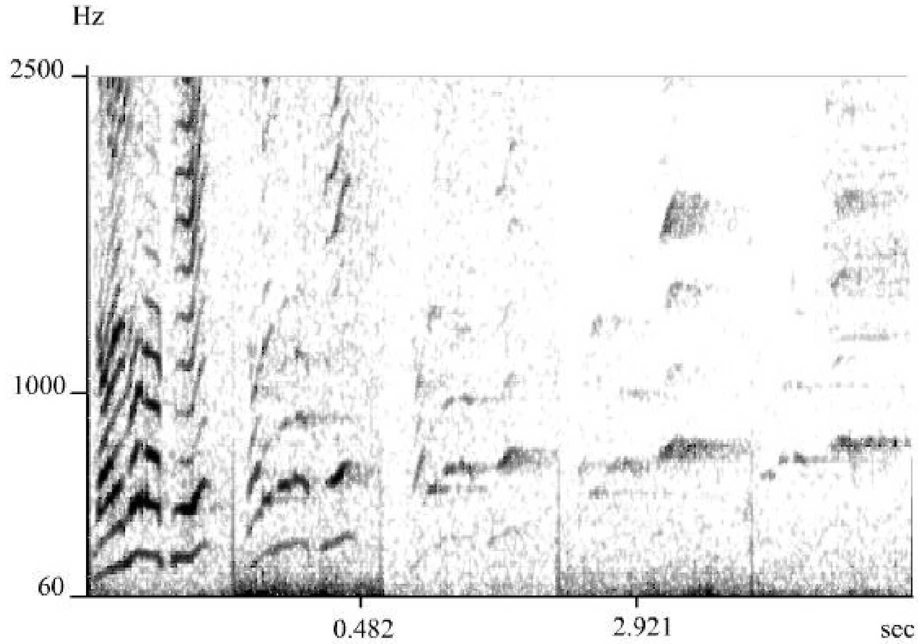

Figure 2 : extracts of a same sentence spoken at 10 m and then shouted at 50, 100, 150, 200 m. One can notice a strong degradation of the harmonics of the voice with the conservation of some of them on which the speaker insists in the situations of distance communications.

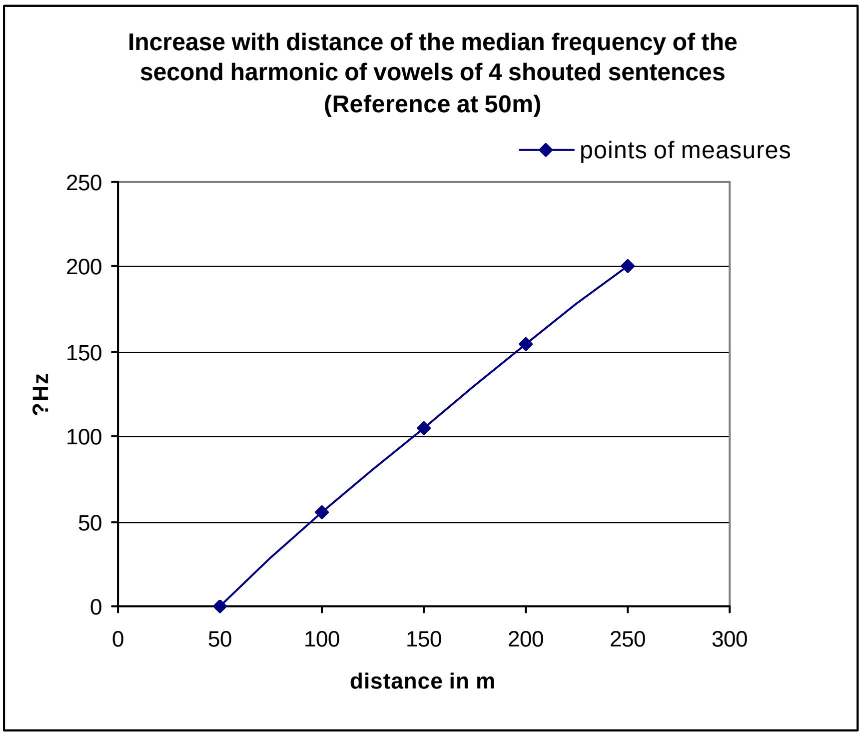

Figure 3

For a comparison, whistled speech is typically produced between 80 and 120 dBA in a band of frequencies going from 1kHz to 4 kHz and its general flow is 10% to 50% slower than normal speech (Meyer & Gautheron 2005, Moles 1970, Meyer 2005). As a consequence, whistling enables to continue the strategy of shouted speech without requiring the vibrations of the vocal chords: it is a natural answer to the constraints observed for shouted speech in the above experiment. Amplitude, frequency and duration, which are the three fundamental parameters of speech, can be more comfortably adapted to the distance of communication and to the ambient noise. Whistled speech is so efficient that full sentences are still intelligible at distances 10 times farther than shouted speech (Busnel & Classe 1976, Meyer 2005).

2.2. Adaptation to sound propagation and to human hearing

A close look at the literature in bioacoustics and psychoacoustics shows that such performances are also possible because the whistled frequencies are adapted to the propagation of sounds and correspond to a privileged static and dynamic behavior of human hearing. In terms of propagation in forests and open habitats, the frequencies going from 1 kHz to 4 kHz are the ones that resist best to reverberation variations and ground attenuation as distance increases (Wiley & Richards 1978, Padgham 2004). In terms of perception, the peripheral ear highlights the whistled frequency domain, for which, on a psycho acoustical level, the audibility and the selectivity of human hearing are also the best (Stevens & Davis 1938). Moreover, until 4000 Hz the ear is at its best performance to develop a precise temporal analysis of an acoustic signal (Green 1985). Additionally, the functional frequencies of whistling are largely above the natural background noise, and these frequencies are concentrated in a narrow band, which has the consequence to reduce the masking effects and to lengthen the possible distance of transmission of the encoded information without risks of degradation. The whistled languages are naturally efficient also because their functional bandwidth is below 500 Hz at a given time, which means that it activates a maximum of 4 perceptive filters of hearing[2], optimizing the signal to noise ratio (SNR) and the clarity of the syllables. Finally, whistled speech defines a true natural telecommunication system spectacularly adapted to the environment of use and to the human ear thanks to an acoustic modification of speech mainly in the frequency domain.

3. Language specific frequency choices imposed by whistled speech

3.1. General productive and perceptive aspects

A phonetic whistle is produced by the compressed air in the cavity of the mouth. The jaws are fixed by the tighten lips, the muscles of the neck and eventually the finger (point 1, figure 4). The movements of the tongue and of the glottis are the principal elements controlling the tuning of the sound to articulate the words (points 2 and 3, figure 4). They enable to regulate the pressure of the air expelled and the variation of the volume of the resonance cavity to produce modulations both in the frequency and amplitude domains.

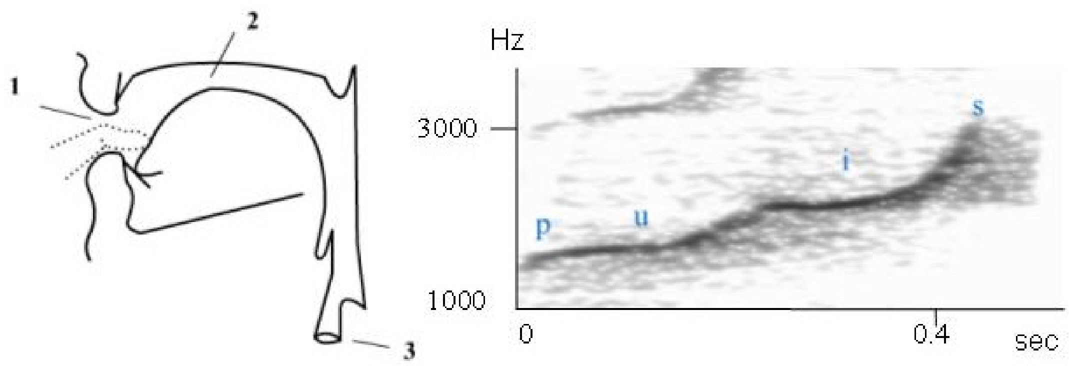

Figure 4 : Position of whistling and example of production of the Greek syllable /puis/

The resulting whistled articulation is a constrained version of the one used for the equivalent spoken form of speech. For non-tonal languages whistlers learn to approximate the form of the mouth of the spoken voice while whistling; this provokes for example an adaptation of the quality of the vowels in a simple frequency. For tonal languages, the control of a transposition of the fundamental frequency of the normal voice is favoured in the resonance areas of the mouth to encode the distinctive phonologic tones carried by vowel nuclei. For both cases, an acute sound is produced in the high front part of the palate and a lower sound comes from a more profound and back part of the mouth. Therefore, the

2 while ERB bands of perception of a whistle are between 120 Hz and 500 Hz; the bandwidth emerging from the background noise has been measured around 400 Hz at short distance (15m) and 150 Hz at 550m.

whistlers make the choice to reproduce definite parts of the frequency spectrum of the voice in function of the phonologic structure of their language.

The literature in psychoacoustics concerning complex sounds like the one of spoken voice provides an explanation for the conformation of whistles to the phonology: human beings perceive spontaneously and simultaneously two qualities of heights (Risset 1968) in a synthetic listening (Helmholtz 1862). One is the perceptive sensation resulting from the complex aspect of the frequency spectrum (called timbre in music); it strongly characterizes the quality of a vowel through the formants. The other is the perceptive sensation resulting from the fundamental frequency, called pitch. In the normal spoken voice these two perceptive variables of frequency can be combined to encode phonetic cues. But a whistled strategy renders the two in a unique frequency, that's why whistlers must adapt the production to the rules of organisation of sounds of their language, selecting the most relevant parts to optimise the intelligibility of the receiver (figure 5 and 6).

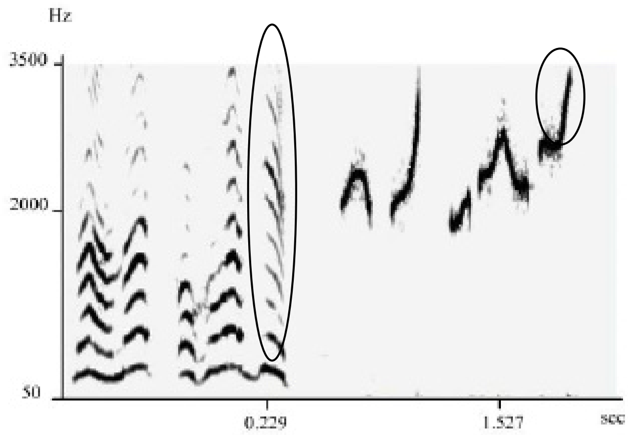

Figure 5: Last /t/ underlined in the spoken and whistled forms of the sentence "mehmet okulagit" in Turkish. The whistles reproduce the spectral distribution of the frequency.

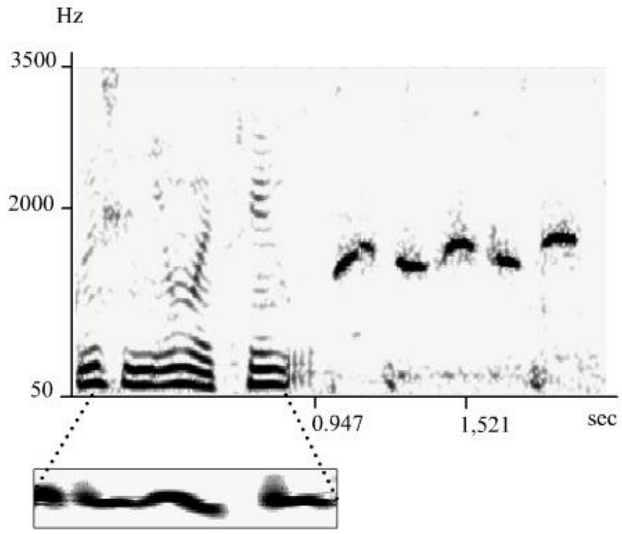

Figure 6: Tonal Mazatec sentence spoken and then whistled. The whistles reproduce mainly Fo

3.2. Typology

The reduction of the frequency space in whistles divides the whistled languages in typological categories. As stated above, the main criterion of distinction depends on the tonal or non-tonal aspect of the concerned languages. The two first papers on whistled languages immediately revealed this gross difference as Cowan described the 4 tone Mazatec whistled form of Mexico (Cowan 1948) and Classe described the Spanish whistled form of the Canary islands (Classe 1956). The papers on Béarnais (Busnel & al 1962), Turkish (Busnel 1970), Hmong (Busnel & al,1989) or Greek (Xirometis & Spyridis 1994) have shown that there is a large variability in each category. Caughley (1976) observed the Chepang whistled language with a behaviour differing from the former ones described. Only recently, his description has been explained in the context of a general typology (Meyer 2005): for each language the whistlers give priority in frequency to a dominant trait that is carried either by the formant distribution (most of the non tonal languages, example figure 5) or by the fundamental frequency (most of the tonal languages, figure 6) of the spoken equivalent, but in the case of an incipient tone language like Chepang the contribution of both is balanced, which explains its intermediate strategy in whistles. As shown later in this paper, this tendency can also be noticed in the rendering of stress in some non-tonal whistled languages like Siberian Yupik, whereas in other cases like Turkish or Spanish the stress only slightly influences whistled frequency and is therefore a secondary whistled feature. Some tonal languages also have an intermediary strategy to emulate the voice in whistles: in the Amazon language called Surui for example the influence of the formant distribution has been described at the level of some whistled consonant types (Meyer 2005). This phonetic behaviour doesn't occur in the majority of the tonal languages that render the consonants in whistles only by prosodic silences. This intermediate category of languages underline that from tonal to non-tonal languages, there is a continuum of variation in the frequency strategies. The analysis of their variability shows that whistled forms are simplified phonetic emulations of the spoken forms rather than phonologic simplifications of the linguistic systems.

4. Comparative description of some non tonal whistled languages

The adaptation of the complex spectral and formant distribution of normal speech into whistles in non-tonal languages is one of the most peculiar and instructive aspects of whistled speech. This phenomenon illustrates extensively the process of transformation of speech from a pluri-dimensional frequency space of spoken voice to a mono dimensional whistled space. In the present study, the detailed results obtained for Greek, Spanish and Turkish whistled vowels and consonants have been taken as a basis. Some complementary analysis coming from Yupik and Chepang extend the insight of non-tonal whistled speech strategies.

4.1. Frequency distribution of whistled vowels

The vowels correspond to the most stable parts of a whistled sentence; they also contain most of its energy. Their mean frequency is much more easy and precise to measure than spoken formants. The statistical data resulting from the analysis of an original corpus of Greek and Spanish natural sentences on one hand, and Turkish lists of words[3] on another hand, show that for a given distance of communication and an individual whistler, each vowel is whistled in a specific interval of frequency values. A vowel space is characterized by a band of whistled frequencies corresponding to its articulation with the mouth. The limitations of this articulation define the frame in which the relative frequencies can vary. Indeed, the pronunciation of a whistled vowel is in direct relation with the specificities of the vocal tract maneuvers occurring in the usual speech - including their variability -, to the extent that these could be achieved while maintaining an alveolar/apical whistle source. The whistled systems of vowels follow the same general organization in all the non-tonal languages. The highest pitch is always attributed to /i/. Its neighbouring vowels in terms of locus of articulation and pitch are for example [ʏ] or [ɨ]. /o/ is invariably among the lowest frequencies. It often shares its interval of frequencies with another vowel such as /a/ in Greek and Turkish or /u/ in Spanish. /e/ and /a/ are always intermediary vowels, /e/ being higher in frequency than /a/. Their respective intervals overlap more or less with the neighbouring vowels, depending of their realization in the concerned language. For example, when there are a lot of intermediary vowels like in Turkish, their frequencies will overlap more, up to the point that they seem not to be easily distinguished without complementary information given by the lexical context or eventual rules of vocalic harmony. Finally, the vowel /u/ has a particular behaviour when whistled: it is often associated to an intermediary vowel in Turkish, Greek but in Spanish it is the lowest one. One reason of this variation is that the whistled /u/ looses the stable rounded character of the spoken equivalent because the lips have a lessen degree of liberty during whistling.

Finally, each language has its own statistical frequency distribution of whistled vowels. As these language-specific frequency scales are the result of purely phonetic adaptations of usual speech, the constraints of articulation due to whistling exaggerate the tendencies of vocalic reductions already at play in the spontaneous spoken form. They also naturally highlight some key aspects of the phonetic-phonologic balance of each language. The analysis of the functional frequencies of the vowels shows that some phonetic reductions characterize the whistled signal when compared to the spoken signal.

4.1.1. Spanish Silbo

Silbo vocalic system is based on the Spanish spoken dialect of La Gomera island for which /o/ and /a/ are sometimes very near and /u/ is very rare (7 %) and often pronounced /o/ (Classe 1957). The spoken vowels (i, e, a, o, u) are therefore whistled in five bands among which some overlap strongly. Four intervals can be statistically distinguished with overlapping in their extrema: (/i/, /e/, /a/ and /o, u/) in a decreasing order of mean frequencies (figure 7). Moreover, some very good whistlers distinguish clearly /u/ from /o/ when necessary by lowering the /u/ and using the extremes of the frequency intervals. These results confirm the analysis of Classe (Classe 1957, Busnel & Classe 1976) but

[3] The recordings of Turkish used here were done during the expedition organized by Busnel in 1967, the data used for the analysis concern a list of 138 words (Moles 1970). Since, Bernard Gautheron preserved them from degradation.

opposes firmly to the theory of Trujillo (1978) stating that only two whistled vowels (acute and low) exist in Spanish Silbo. This theory is confirmed wrong even if perceptive performances of vowel identification are considered (see 5.2). Yet, it has been erroneously taken as a reference in Carreiras & al (2005) for analysing the first neuroscience perception experiment on a whistled language and also for a teaching manual supposed to be used by the professors of Silbo taking part of the process of revitalization through school teaching in La Gomera (Trujillo, 2005). Most of the native whistlers contest the Trujillo theory, some of which are also professors in school don't apply it for a didactic teaching preferring the traditional form of teaching by imitation (pers. com. Rodriguez 2006).

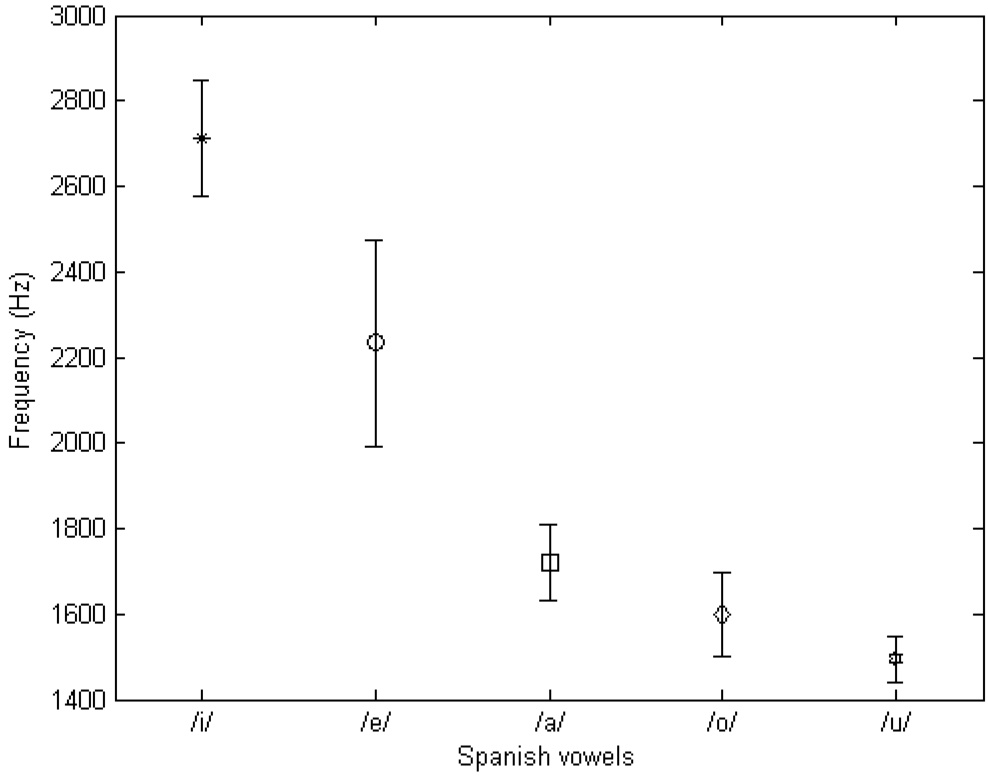

Figure 7: Frequency distribution of Spanish whistled vowels

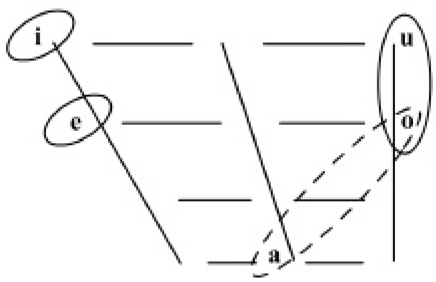

Figure 8: Vocalic triangle of Spanish with statistical grouping underlined

4.1.2. Greek

The five phonological Greek vowels (i, ɛ, ɑ, ɔ, u) are whistled in five intervals of frequencies that overlap in unequal proportions (figure 9). The whistled [i] never overlaps with the frequency values of the other vowels, which overlap more frequently. In a decreasing order of mean frequency, /u/ and /e/ are whistled at intermediary frequencies, and /a/ and /o/ at lower frequencies. The standard deviations of /u/ and /e/ show that they overlap up to the point that they are statistically difficult to distinguish. Such a situation is an adaptation to the loss of the rounded aspect of /u/ by fixation of the lips during whistling. Similarly, the frequency intervals /a/ and /o/ also overlap highly. Indeed, /a/ is the back vowel [ɑ], which is phonetically close to [ɔ] when it looses its rounded character with the fixation of the lips during whistling. Finally, the whistled vowels define statistically three main distinct bands of frequencies: (i), (u,ɛ) and (ɑ,ɔ). These reductions are only phonetic and they do not mean that there are only three whistled vowels. All the whistlers recorded have the same pattern of frequency distribution of whistled vowels, which is rooted in the way Greek vowels are articulated. When the context is not sufficient to distinguish the statistically overlapping vowels, the whistlers use the extremes of the intervals. Yet, most of the time, the whistlers use the lexical context to distinguish the vowels /u/ and /ɛ/ on one hand, and /ɑ/ and /ɔ/ on another hand.

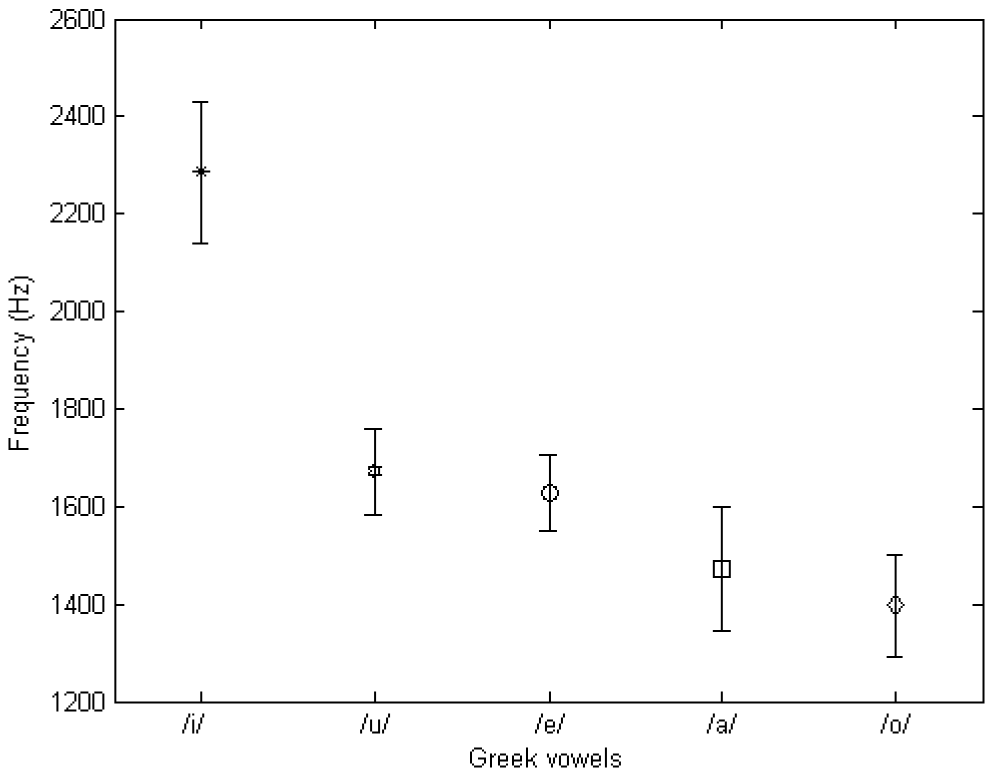

Figure 9: Frequency distribution of Greek whistled vowels

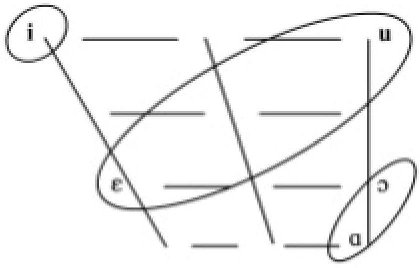

Figure 10: Vocalic triangle of Greek with statistic groupings underlined

4.1.3. Turkish

The 8 Turkish vowels are whistled in a decreasing order of mean frequencies in 8 intervals (ɪ, ʏ, ɨ, ɛ, œ, ʊ, a, o) that overlap quite a lot The pattern of distribution in the frequency space is the same for all the whistlers. The vowel i [ɪ] bears the highest frequency, it is relatively well distinguished from its nearest neighbouring vowel in frequency: ü [ʏ]. Some intervals overlap much more than others: first of all, the vowels i [ɨ] and ü [ʏ] which have bands of frequencies nearly confused even if ü [ʏ] is meanly slightly higher than i [ɨ]. Secondly, the intervals of frequencies of the vowels e [ɛ], ö [œ], u [ʊ] overlap largely. Thirdly, the respective intervals of the whistled frequencies of a [a] and o [o], with o [o] at a lowest frequencies of all.

As a consequence, such a complex vocalic system of eight whistled frequency intervals highlights 4 statistical groupings [(ɪ), (ɨ, ʏ), (ɛ , œ, ʊ), (a, o)] showing that some phonetic reductions effectively exist (figure 11).

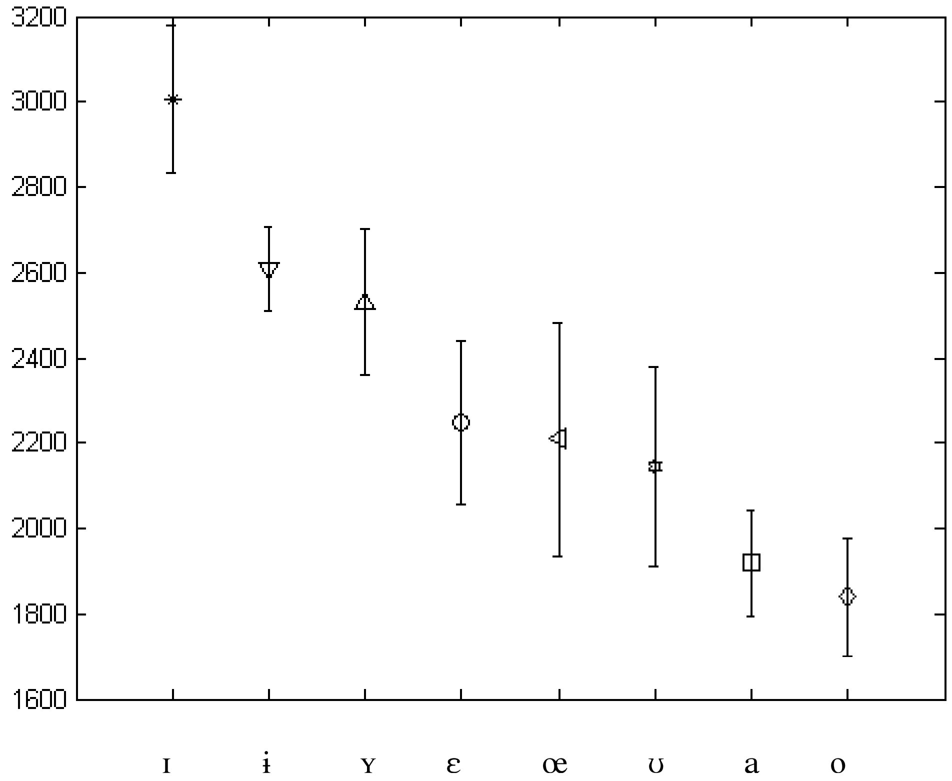

Figure 11: Frequency distribution of Turkish whistled vowels

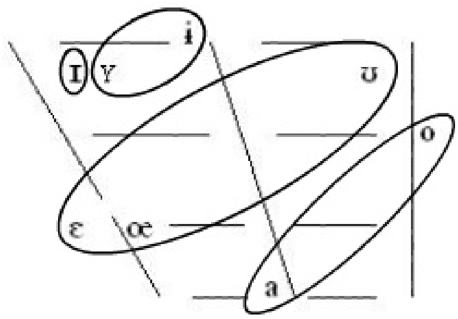

Figure 12: Vocalic triangle of Turkish with statistic groupings underlined

These groupings are compatible with the vocalic harmony rules that order some aspects of the syllable chain in a glutinous Turkish word. They even provide a simplified space of possibilities enabling the speakers to identify the vowels with a reduced number of variables. This means that the whistled system is unravelled by the rules of vocalic harmony. The latter in Turkish is a process through which a part of the vowel quality oppositions is neutralized by an effect of assimilation between one vowel of a syllable and the vowel of the following syllable. It applies from left to right and therefore only the non-initial vowels are concerned. The rules are the following:

(a) If the first vowel has an anterior pronunciation (i[ɪ], e[ɛ], ü[ʏ], ö[œ]) or a posterior one (i [ɨ], u [ʊ], a[a], o[o]), the subsequent vowels will respectively be anterior or posterior. Which classifies the words in two categories.

(b) If one closed (or High) vowel is unrounded or rounded, the following vowel will respectively be also unrounded or rounded. On another hand, a diffuse (or Low) vowel in non-initial position will always be unrounded (étirée). The direct consequence is that the vowels ö and o will always be in intial syllable. We can sum up the possibilities opened by the vocalic harmony as the following:

a et i ---------- can be followed by ---------- a et i

o et u ---------- can be followed by ---------- a et u

e et i ---------- can be followed by ---------- e et i

ö et ü ---------- can be followed by ---------- e et ü

For every non-initial syllable the system is reduced to 6 vowels. It can be noticed that the only oppositions that subsist thank to this process are those between High and non-High vowels.

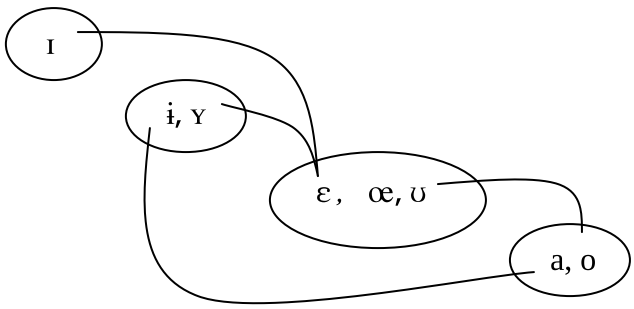

Figure 13: combination of vocalic frequency intervals and harmony rules

The statistical data suggest that in Turkish the 4 links created by harmony rules combine with the 4 whistled statistical bands of frequencies. Each frequency band is in relation to each of the other three frequency bands through only one harmony rule (figure 13). Because of human abilities such as phonetic and auditory memory in vowel discrimination (Cowan & Morse 1985), the nature of two consecutive vowels not whistled in the same band of frequencies will always be identified. A few cases of confusions may subsist but only for two syllables words with the same consonants: (i) two consecutive [ʏ] might be confused with two consecutive [ɨ]; (ii) two consecutive [ʊ] might be confused with two consecutive [ɛ]; (iii) [œ] followed by [ɛ] might be confused with [ɛ] followed by [ɛ]; and (iv) [a] followed by [a] might be confused with [o] followed by [a] or [o] followed by [o]. However the ambiguities that are not solved by the harmony system are sometimes overcome by the use of extremes of the bands of frequencies. For example for the common words “kalay” (/kalaj/) and “kolay” (/kolaj/): in the latter /o/ and /a/ are effectively distinguished because /a/ bears a higher pitch despite the fact that its two vowels are usually whistled in the same way. It is relevant to ask the question whether this process also helps in spoken form as it would mean that we perceive frequency scales through some timbral characteristics of the vowels. A part of the answer will be given in the final discussion of this paper.

4.1.4. Stress in Greek, Turkish and Silbo

For Greek, Turkish and Silbo, stress is usually preserved in whistled speech especially if it eases the intelligibility of the sentences. Most of the time it is expressed by a combined effect of amplitude and frequency increase. Stress doesn't change the level-distribution of the intervals but acts as a secondary feature influencing the frequency. A frequency stressed vowel is often in the highest part of its typical interval of frequency. But it's not always the case as the frequency variation of a stressed vowel depends of the whistled frequency of the preceding vowel.

4.1.4.1. Stress in Silbo

The rules of the tonic accent of Spanish are mostly respected in Silbo. The stress is performed in two different ways in function of the context: either it is marked by a frequency and amplitude increase of the whistled vowel, or by a lengthened quantity of the vowel when the usual rules of stress are disturbed, for example for proparoxyton words (Classe 1956).

4.1.4.2. Stress in whistled Greek

The whistlers produce stress in 80 % of the measured cases through an increase of the amplitude and an elevation of the frequency of the whistled vowel. This has the effect of situating the frequency of the stressed vowel in the upper part of its interval. Indeed, for the spoken form *«It is generally admitted that in a neutral intonative context the stressed vowels are longer, higher and more intense than the unstressed ones* (Dimou and Dommergues 2004: 177). "*First, in this language some minimal pairs exists that are differentiated only by the location of the stress. Second the morphemes allow a stress occurrence inside a definite zone: independently of the number of words* »

4.1.4.3. Stress in whistled Turkish

Spoken Turkish uses an intonative stress that takes place on the particles preceding the expression of interrogation, negation and on imperative expressing prohibition. Among the sentences of the examined corpus, several present the required conditions for an analysis. For example in the interrogative sentence [kalemin var mɨ] meaning "Do you have a pen?" or (pen-POSSESSIVE2sg there is INTERROGATION), the /a/ of /var/ is stressed in spoken voice, at least in intensity. In the 6 examined whistled pronunciations of this sentence, only one is not stressed at the frequency level. For the others, the /a/ has a frequency value in the highest part of the interval of values of Turkish whistled /a/. However this stress is also developed through a slight increase of the amplitude. Other examples presenting the three different configurations of stress in Turkish are available in Meyer (2005).

4.1.5. Two other non tonal languages: Siberian Yupik and Chepang

Siberian Yupik and Chepang are representative of the variability of the impact of both vocalic timbre and stress on the whistled signal of a non-tonal language. For these radically different languages, the underlying whistled frequency-scale carried by the nucleus of the vowels contributes to the whistled pitch but it has not anymore systematically the dominant influence like in Turkish, Greek or Silbo. For both languages, three groups of vowels have been identified in function of the influence of the timbre on whistled pitch: /i/ which «pulls » the frequencies of the syllable toward high values so that /i/ always remains the highest pitch without being disturbed by the prosodic context. /e, a, u/ which are more neutral and therefore influenced by prosody and sometimes consonantal contexts and /o/ which pulls the frequencies to low values but is more influenced by the context than /i/.

For Siberian Yupik, a first corpus of whistled speech constituted in summer 2006 in bilabial whistling has shown that /a, e, u/ (/e/ being the schwa) are very variable and highly overlap in between each other while /i/ is statistically distinguished (Figure 14). The rhythmic complexity of Siberian Yupik (Krauss 1985) and the tonal tendency of Chepang affect the whistles in such a way that they underline that whistles adapt to various forms of language structures. For the incipient tonal language Chepang, Caughley (1976) has observed that pitch is influenced in spoken intonation and whistled forms by two criteria of the articulation of the vowel nucleus affecting its weight: height (High, Medium or Low) and backness (Non-Back vs. Back). Meyer (2005) has verified that the frequency bands of /a/, /u/ and /e/ whistles effectively vary more than for /i/ (In bilabial whistling /a/ varies from 1241 to 1572 Hz, /e/ from 1271 Hz to 1715 Hz and /u/ from 1142 to 1563 Hz whereas /i/ stay around 1800 Hz).

Siberian Yupik and Chepang have a typological behaviour characterized by a research of balance between Fo and timbre at the level of the vowel. With broader corpuses of whistled speech in each language, it might be possible to go deeper in the analysis, but very few speakers still master this whistled knowledge.

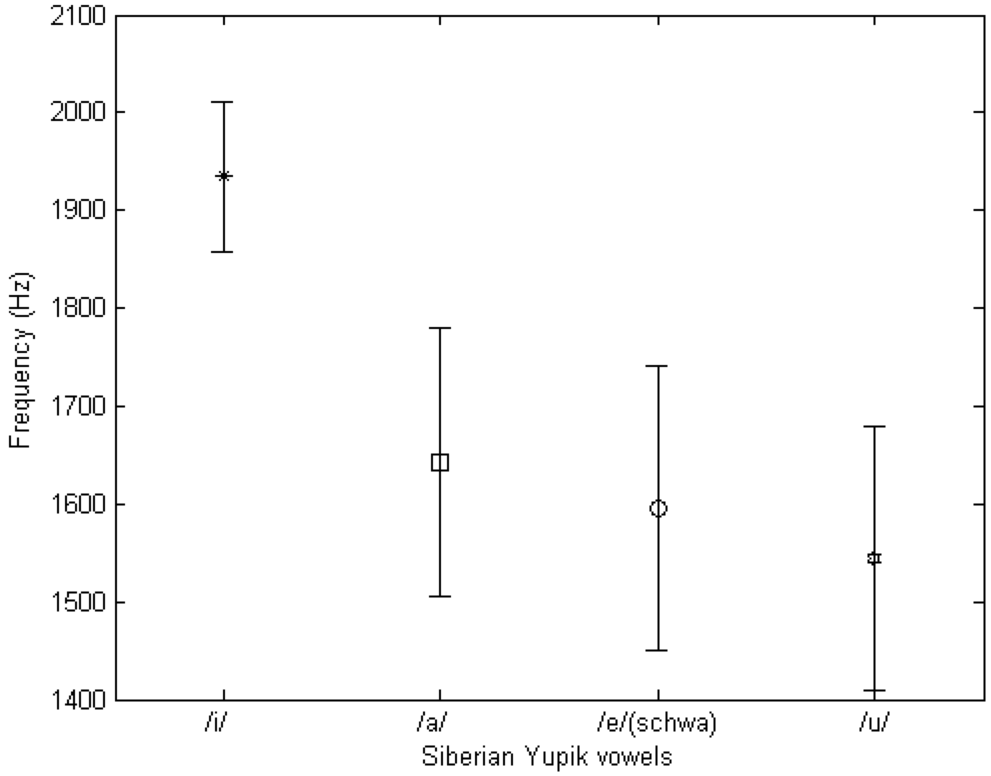

Figure 14: Frequency distribution of Siberian Yupik whistled vowels

4.1.6. Other common characteristics relying on vowels

Each vowel is characterized by a relative value that can vary with the technique and the power of whistling. The farther the whistlers have to communicate the higher are the relative frequencies of the whole set of vowels, /i/ staying beyond 4 kHz and the lowest vowel above 1kHz. In a single sentence the limit of one octave is systematically respected between the lowest and the highest frequency, this phenomenon also observed in tonal whistled languages might be due to risks of octave ambiguities in a same perceptive flux because of the spiral-like perception of pitch of the human ear (Shepard 1965, Risset 2000). Another interesting aspect concerns vowel durations: in the languages that don't have phonologic distinctions of vowel quantities, the durations of the vowels are adapted to ease the intelligibility of the sentence. It's not rare to hear that a vowel is maintained during one second, especially in very long distance communication. In the languages having long and short vowels (like Siberian Yupik), this effect is only done on long vowels. On another hand, when the final and the initial vowels of two consecutive words are identical they are nearly systematically whistled in a single vowel. In fact, exactly like in spoken speech, the word by word segmentation is not always respected even if two words present two different vowels as consecutive sounds: for example in the Spanish sentence «Tiene que ir. », /ei/ from «que ir » is whistled like a diphthong similarly as in « Tiene ». Diphthongs are treated as pairs of vowels; a modulation going from the frequency of the first vowel to the frequency of the second vowel makes the transition.

4.2. Consonants: frequency and amplitude modulations

Whistled consonants are expressed through modulations in frequency and amplitude of the simple frequencies of neighbouring vowels. When the amplitude modulation shuts up the whistle, they are also characterized by silent gaps. In all the languages that are clearly non-tonal (Spanish, Greek, Turkish and Siberian Yupik in this paper), the whistled modulations are determined by the articulatory locus of the consonants. The resulting system reaches a degree of complexity that is limited by the articulation constraints dues to whistling. This naturally highlights frequency classes of consonants mostly confined to congeners, i.e. sounds formed at close articulatory loci (the figure 15 represents the most representative whistled articulations in function of the locus of explosion or resonance of the superior part of the mouth (teeth, palate, upper pharynx)).

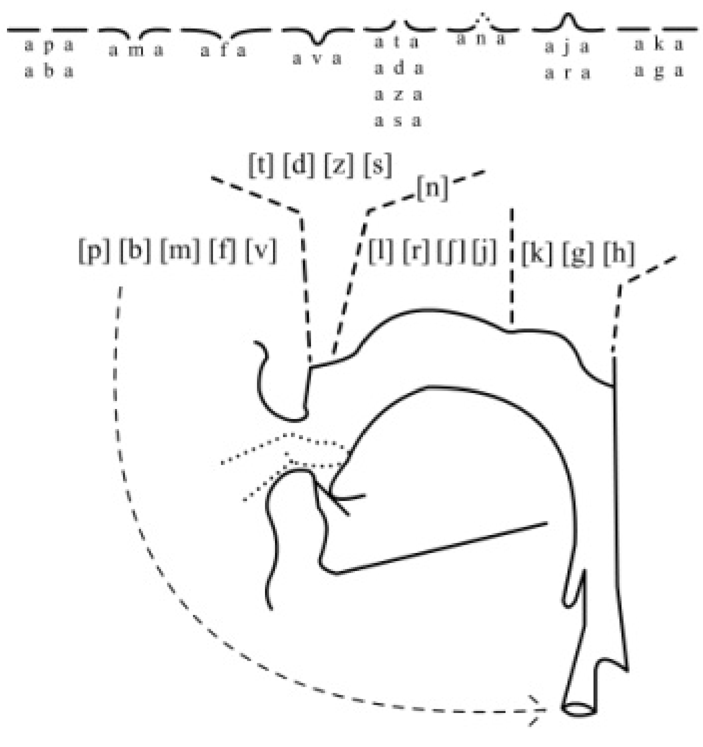

Figure 15: Loci of explosion or main resonance area of the whistled consonants and corresponding phonetic shapes (the glottal occlusion plays a role for [p], [b],[m], [v],[f], [k],[g],[h])

The figure 15 not only explains the origin of the consonant frequency classes but also why some of the consonants oscillate between two categories (as for [n]). On another hand, the fixed lips modify slightly the pronunciation of [f] and [v], of the nasal [m], or of the occlusives [p] and [b] which all often use in compensation a glottal occlusion to produce a stop ([p] and [b] are expressed mainly with this process and are therefore very similar to [k] and [g] in whistles).

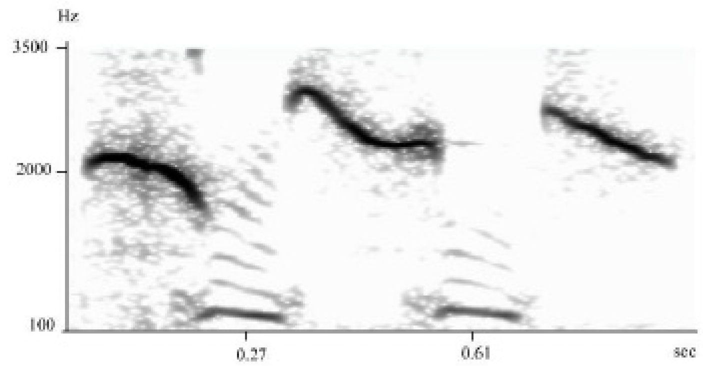

Figure 16: « Kom -yun-köp » in Turkish. [m] like [k] are the occasion of a glottal constriction (controlling the power of the blow) that can be observed on recordings made from close distance (en champ proche).

The fact that each language has a specific frequency distribution of whistled vowels is mirrored in consonant frequency modulations. The representative example of /t/ combined with the vowels of Turkish, Greek and Spanish is displayed on figure 17.

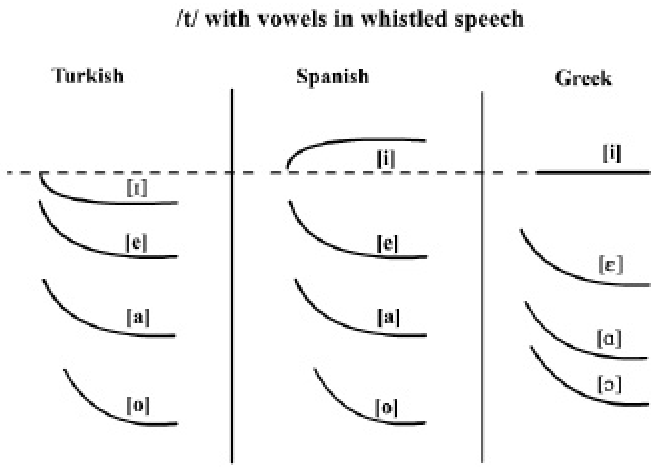

Figure 17: Mean relative frequency distribution of the whistled vowels in three languages, the modulations are due to the articulation of /t/ and the pointing to its locus.

Amplitude modulations of the envelope of whistled speech also provide information about the consonants that are whistled, especially to discriminate them inside a frequency class. For example an amplitude modulation producing a clear silence characterizes occlusives, with a very rapid modulation for unvoiced consonant and slower ones for voiced ones. Fricatives and nasals are characterized by slow progressive amplitude modulations sometimes finished by a short silence (depending of the distance of listening). Finally continuants are expressed through continue modulations that appear with a silence only at long distance (/j/ , /l/ ,/r/ trills, taps or flaps for example). (example of /l/ on figure 18)

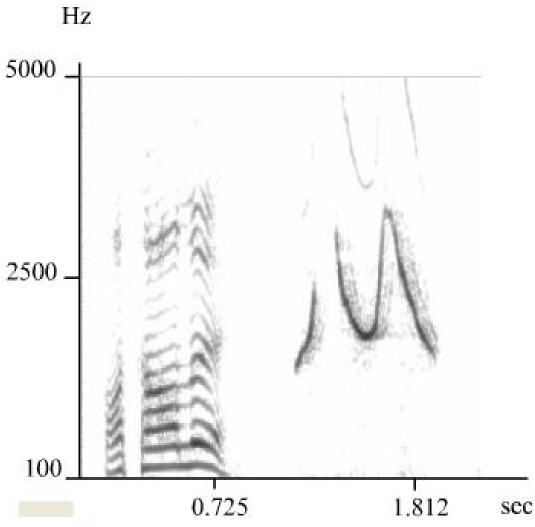

Figure 18: /atalɨ/ in Turkish

Consonant clusters in whistled speech cumulate the frequency, amplitude and duration information of their constituent consonants but truncate them at their encounter. When there is a silent gap in one of the consonants, the truncation is at this level (figure 19). In dusters constituted only by continuants, the cumulated effect of the concerned consonants is developed in a single longer modulation, like for [ll] or [rl]. However, a concave and a convex frequency from two distinct continuants only appear together separated by a silent gap. Some small differences exist between the languages, for example Spanish eludes the /r/ after a plosive but only truncates it in Greek or Turkish. In Siberian Yupik, the tendency is to truncate more the consonants than in the other three languages, even if not in clusters.

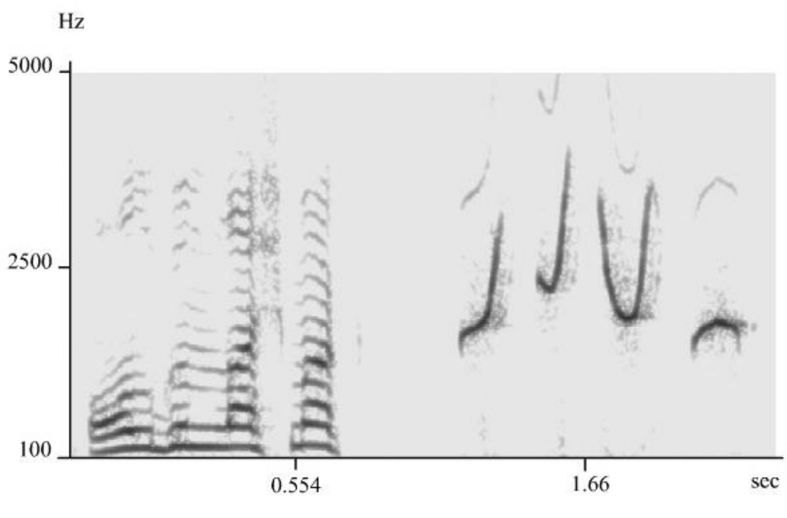

Figure 19: /olgunlaʃmak/ spoken and whistled /ol/--/gun/--/laʃ/--/mak/ where each silence is apart of a consonant cluster

4.3. Discussion and conclusion

Generally speaking, the results presented here show that the whistlers of non-tonal languages rely on articulation considerations but render both segmental and prosodic features because of the complementary control of the air flux. Vocalic and consonantal groupings are mainly due to articulatory proximities shared with the spoken speech, except in the case of those due lip constraints imposed by whistling affecting /u/ for the vowels and imposing a new strategy of pharyngeal control of air pressure for some consonants like /b/ and /p/. Therefore, most of the time, the groupings underline phonetic reductions that are common to the ones observed in spontaneous natural speech (Lindblom 1963, Gay 1978, Lindblom 1990) and they do not allow theorising on hypothetic phonological simplifications. The vowel systems are expressed in floored frequency scales while the consonantal shapes can be grouped in classes explained by similar loci of articulation or similar coupling and decoupling movements.

Some acoustic correlations between usual and whistled speech can be observed, they are mainly due to common combinations of tongue height and anterior-posterior position. As the second formant of the voice is the result of the cavity formed between the tongue and the palate, it is often in correlation with whistling, but many aspects suggest that it is not the whole story. For example, even is we exclude the /u/, whistled groupings suggest that broader considerations of the vowel frequency spectrum must be considered. It is the same for consonants: the similarity between formant two and whistled modulations are clear for occlusives or [j] but not for the liquids [l, r] despite the common mouth movements. One reason is that the whistled modulations are also greatly influenced by the tension of the low vocal tract that changes in function of the consonants and vowels that are pronounced.

The simple frequency band of whistles is easier to analyse than traditional complex voice spectra, the formants being much more diffuse in comparison. This natural phenomenon underlines key features of the phonology of each language while suggesting which acoustic cues carry them. For example, salient parts of the formant distribution are embodied in whistles as pure tones for vowels and combinations of frequency and amplitude modulations for consonants. Moreover the loci of articulation are very precisely pointed. As the whistled elements are known to be sufficient for reaching high intelligibility performances at the level of the sentence, all the information given by whistles on segmental features is relevant, especially for non tonal languages. Another example is the little studied aspect of the role of the combination of frequency and amplitude modulations (also called phase), which is more and more underlined as an important aspect carried by consonants to ease the intelligibility of speech, with an increasing role in current noisy conditions (Zeng and al 2005). In fact, whistled phonetic cues first show that vowels are always modulated underlining the coherency of the spectral elements of voice that they transpose. Then, with a large scale of observation like the sentence level, which obliges the phonetician to consider speech as dynamic[4], the combined rapid modulations of amplitude and frequency give a representative overview of the relative location of the events in the sentence by signalling *"lines edges and other narrow events"* (Oppenheim & al 1979: 534). Thanks to the whistled

[4] and not as quasi-stationary composed of stationary segmental features

signal the consonants providing these perceptively relevant cues in noise or in a cocktail party are the following: occlusives and affricates (rapid opening and closing of the cavities of the mouth), nasals (coupling and decoupling of nasal cavities in usual speech and opening and closing of the mouth), and laterals like /l/ or /r/ (produced by a coupling and decoupling of two parallel oral cavities).

5. Perception experiment on whistled vowels

In order to deepen the understanding of the perception of the whistled vowels and their natural link to the vowels of the usual spoken speech, two variants of a same experiment were developed. The idea was to test subjects who knew nothing about whistled languages in order to measure if they performed the same categorisations of whistled vowels as the ones made by the trained whistlers. For this, they only knew they had to recognize the four vowels /i, e, a, o/ in a simple and intuitive task. They never received any feed back of their performances before the end of the test, or any explanation of the real distribution of the whistled vowels. The first variant experiment presented the vowels isolated; the second one presented the vowels in the context of their sentence.

5.1. Method

5.1.1. Participants

The tested subjects were 40 students of 19 to 29 years old whose mother tongue was French. 20 persons have performed each variant of the experiment. The persons of the variant 1 (alone vowels) were not the same as the one of the variant 2 (sentences). The students' normal hearing performances have been tested by audiogram. The subjects had never listened to whistled speech before and never heard an explanation of the distribution of the vowels in non tonal whistled speech.

5.1.2. Stimuli

The selected vowels were /i/, /e/, /a/, /o/ from the Spanish whistled language of La Gomera island. Indeed, these vowels exist also in French with similar pronunciations. Another reason for this choice was that these four whistled vowels (or close pronunciations) have the same frequency distribution also in Greek and Turkish. Given the structure of French, one can reasonably think that whistled vowels of French would bear the same scale. The sound material was made of 84 vowels, all extracted from the recording of 20 long sentences whistled relatively slowly in a single session by the same whistler in the same conditions (targeted distance and whistling technique, distance from recorder, background noise). These 84 vowels (21 /i/, 21 /e/, 21 /a/ and 21 /o/) have been chosen by taking into account statistical criteria based on the above analysis of Silbo (4.1.1). First, the final vowels of a sentence have been excluded because they are often marked by an energy decrease. Moreover, the selected vowels are 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 don't overlap, while staying relatively wide.

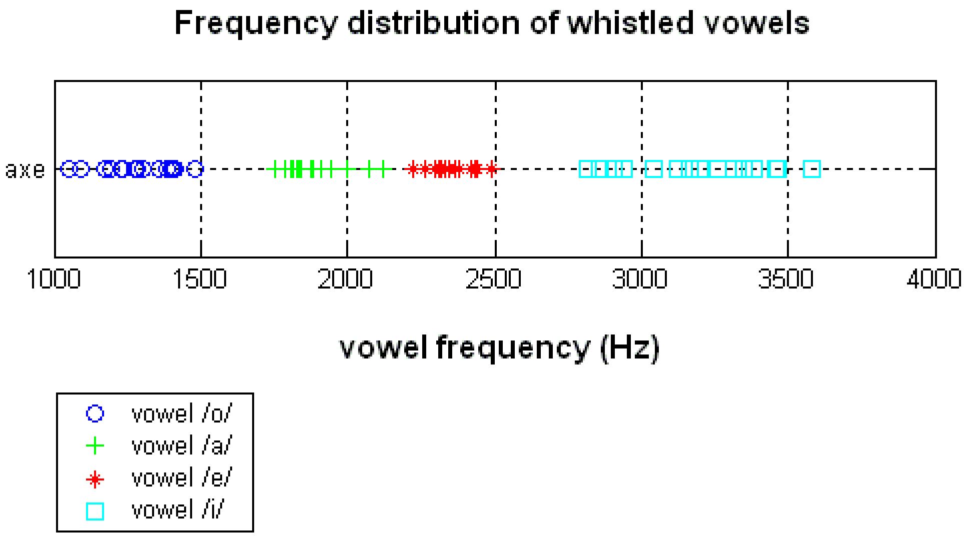

Figure 20: Played vowels of the experiment

While the sounds of the variant 1 concern only the vowel nucleus without the consonant modulations, the stimuli of the corpus of variant 2 keep the part of whistled sentence preceding the vowel during 2 to 3 seconds. This was made to test 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 was made 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.

5.1.3. Design and procedure

For each variant, the task of the experiment was the following: after listening to each whistled vowel the subject selected the vowel type that he estimated the nearest to the one heard by clicking on one of the four buttons in French orthography « a », « é », « i », « o ». The task was therefore a forced-choice among four solutions (4-AFC). The interface, programmed in Flash-Actionscript controlled first the presentation of the 20 sounds of the training phase in an ordered list containing all the possible successive combinations of vowels and then the presentation of the successive 64 sounds in a non-recurrent random algorithm during the test. The subjects where tested in a quiet room with high quality Sennheiser helmets.

5.2. Results

The collected data put together the played sounds and the answer given. They were analysed for each individual and then put in common. A specific *Matlab* program was developed to summarize the answers in confusion matrixes and to present them graphically by reintegrating some information concerning for example the frequency distribution of played vowels (figure 20).

5.2.1. Results for isolated vowels (variant 1)

The mean level of success corresponding to right answers is of 55%. Considering the protocol and the task these results are largely above the random or chance of 25 %. But the mean rates of right answers vary largely in function of the vowels. For the other percentages, most of the confusions can be qualified of logical in the sense as a vowel is generally confused with its frequency neighbouring vowels in 83% of the cases of confusion (in bold letters on table 1).

Table 1 : Confusion matrix of 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 precise 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 of the experiments (figure 21). On this figure, appear also the estimated curves of the answers of the subject, averaged by polynomial interpolations of the second order.

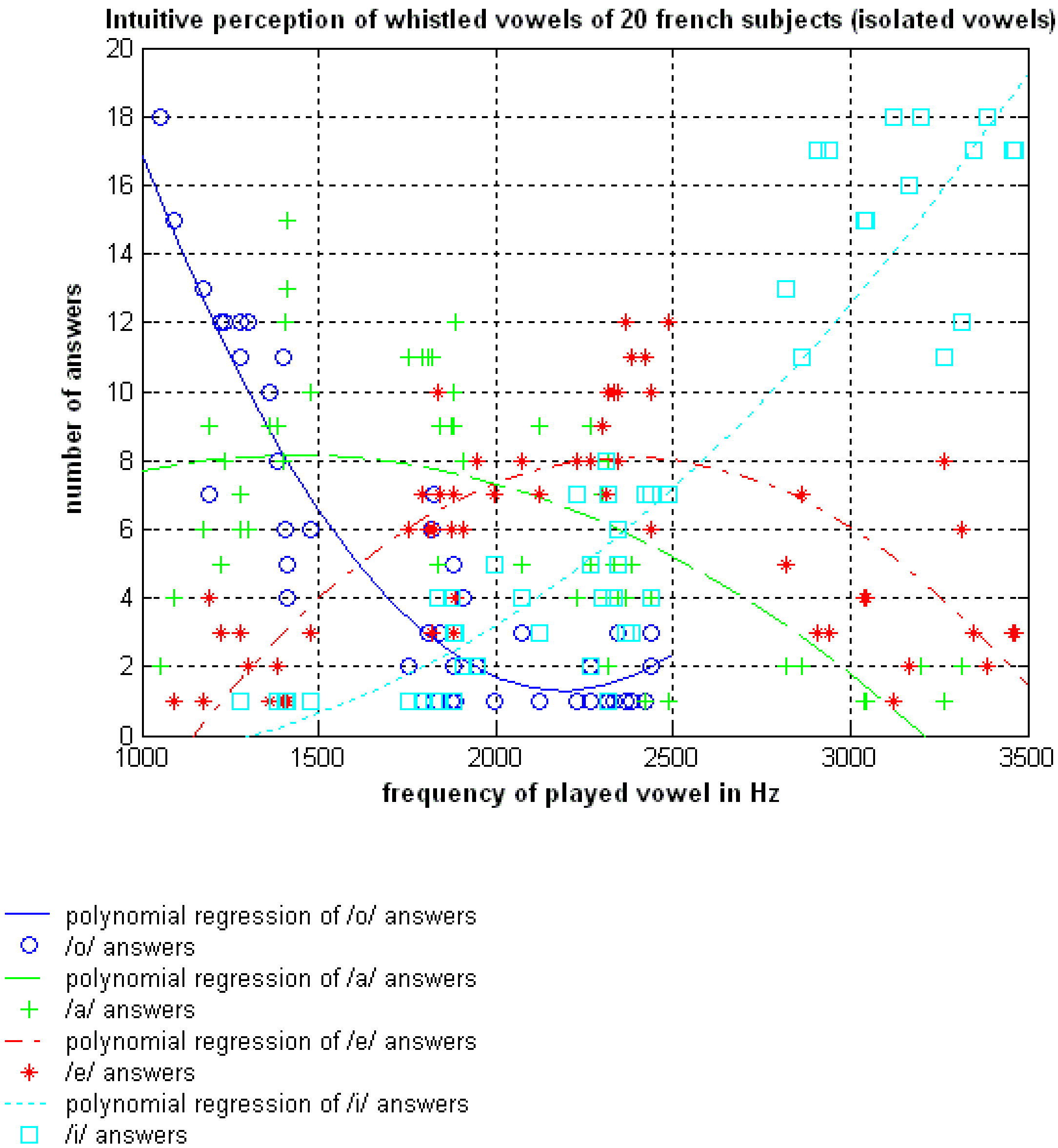

Figure 21: Distribution of the answers in function of frequencies of the whistled vowels

5.2.1.1. Individual variability and other confusions

The rates of success and of confusion distinguish the subjects. Indeed two subjects have very high performances with 73,5 % of right answers. Then a group of 6 persons has more than 40 right answers for 64 sounds (62,5%). 4 other persons follow them with more than 58% of right answers. This means that half of the subjects have good performances to the test. The 10 other subjects all have performances above 37% of right answers. The four persons with the lower results are between 37 and 40% land the 6 other ones have better results to the test because they obtain performances between 45 and 54%.

Generally speaking, the less efficient still have a confusion matrix with logical confusions. Their relatively low rate is often due to a systematic confusion between two whistled vowels neighbour at the level of the frequencies. The variability of the subjects depends of the concerned vowel: (i) for /i/ most of the subjects have very good success results because 16 of them have a score over 12 for16 and 2 even have 100% right answers. The less efficient has 9 right answers (56%). (ii) For /o/, 6 persons identify more than 10 vowels among 16 (62,5%). All the others often think the /o/ is an /a/. (iii) The /a/ is the less well-identified letter because the subject often think it is a /e/ or sometimes a /o/. (iv) The /e/ is 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 subjects can categorize them very well as different vowels, and this uniquely thanks to the pitch they perceive. Finally, the more frequent confusions are the following: the /o/ is often thought as a /a/, the /a/ and the /e / are reciprocally often taken one for another, partly because of the proximity of their intervals of whistling.

5.2.1.2. Differences between musicians and non-musicians

Among the subjects of this variant, 6 were musicians. The results of this group are significantly distinguished form the non musicians ($F(1,18)=6,71$, $p<0,018$). One can therefore say that the musicians have more success to the task of the test than the non-musicians.

5.2.1.3. Conclusion

All the analysis detailed above support the fact that the French subjects categorise the whistled vowels « a », « é », « i », « o » like the whistlers of la Gomera island. The tendencies of the curves of good answers show that the subjects in general have good performances to the task.

These results occur despite one experimental bias that can be observed and that relies on the fact of presenting isolated vowels without any sound context except the one of the preceding listened vowel. Indeed, some subjects reflect some confusion on one vowel on the following answer. For example, if while listening a whistled /e/ they had answered /a/ and that the following vowel is a /a/ they sometimes have the tendency to answer /o/. As a consequence one can observe a cascading effect of logical confusions that stops when there is a significant frequency jump. This phenomenon is difficult to control with only 4 types of vowels that is why the variant 2 of the experiment was then developed. It has the interesting effect to confirm that the non-whistlers subjects perceptively floor their vowel prototypes in a distribution that depends of the frequency. In these conditions, it is not surprising to note that the musicians have better performances because they are more used to associate an isolated pitch to a sound reference culturally marked.

5.2.2. Results for the sentences ended by the tested vowel (variant 2)

This experiment aimed at verifying the effect of the context on the perception of vowels. Particularly the hypothesis was emitted that by an approach nearest to the real conditions of listening of whistlers -

who don not perceive the vowels isolated but integrated in a sound flux – one could observe a suppression of the effect of cascading effect of logical confusions.

The same general tendencies as for variant 1 could be measured with slightly better performances to the identification task: 60,2%. The whistled vowels /o/ and /i/ are even better identified than for the variant 1 (respectively 73,13% and 87,81%) whereas the vowels /a/ and /e/ are slightly less well identified (see table 2 for percentages and figure 22 estimated curves of answers).

Table 2 : Confusion matrix of the answers of 20 subjects for whistled vowels in context (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

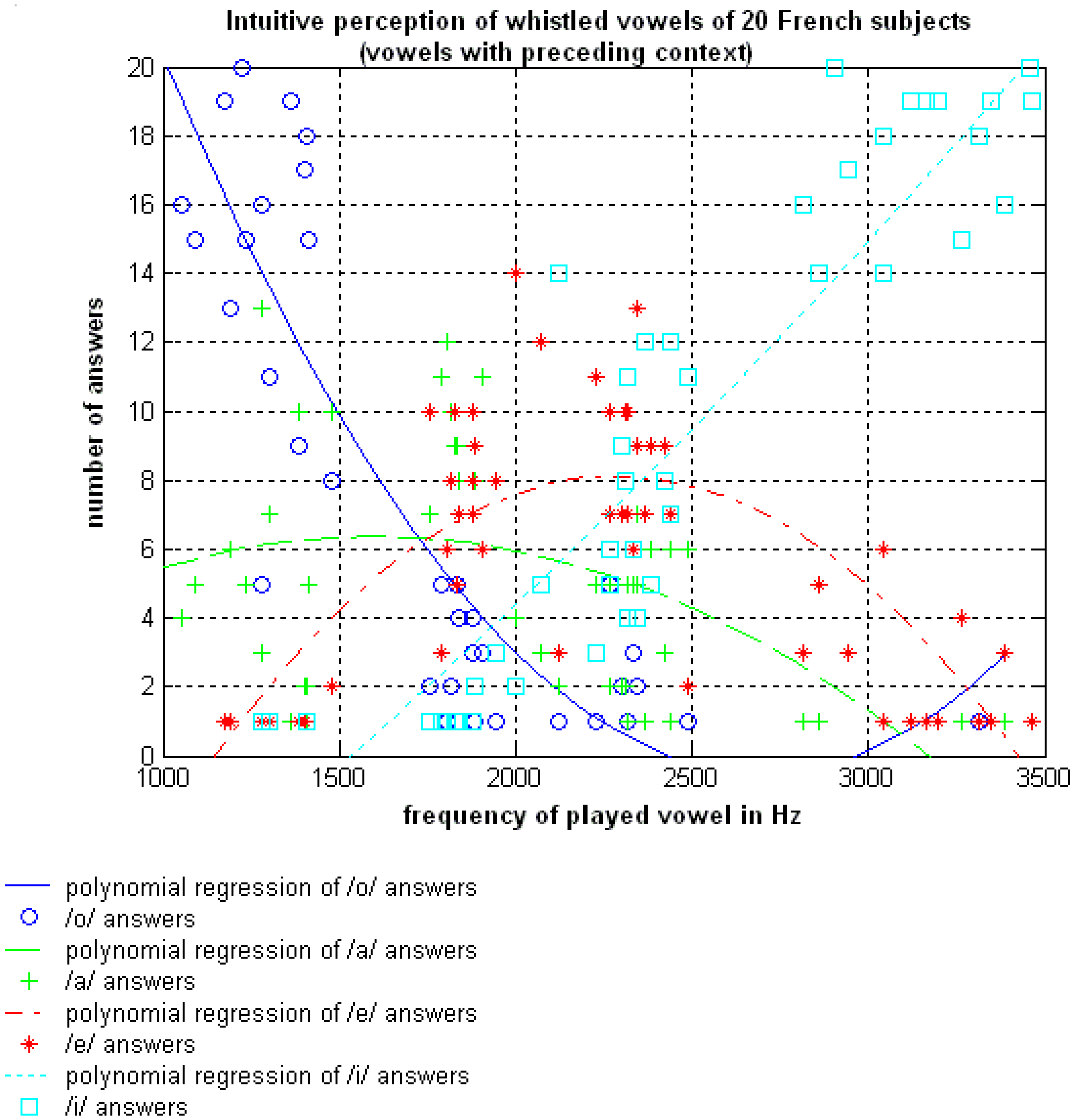

Figure 22: Distribution of the answers in function of frequencies of the whistled vowels

5.2.2.1. Confusions and inter-individual variation

8 persons have a success score above 40 right answers for 64 (62,5%). 3 other even have more than 47 among 64 (73%). The better 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 has an overall score of 75%. At the opposite the less efficient subject in has a score of 46% of identification, which is more than for variant 1. For the confusions, the better scores for /o/ show that it is less thought as a /a/, whereas /a/ is still a lot confused with /e/. Finally, and this is new, /e/is often thought as a /i/ with strong differences between subjects. It's often at the level of the identification of /a/ and /e/ that the differences have been made between subjects with high scores and subjects with lower scores. Finally,

one confusion of a /i/ with a /o/might be explained either by the absence of attention at this moment or by the effect of octave ambiguity.

5.2.2.2. No difference between musicians and non-musicians

Again, in this variant there were 6 musicians among the subjects (despite the fact that the subjects of the two variants are distinct). An analysis of the variance similar to the one performed for variant 1 has shown that this time the results of the musicians were not significantly different from those of non-musicians (F(1,18)=6,71, NS). The context effect has eased the choice of the non-musicians without affecting the performances of the musicians.

5.2.2.3. No learning effect of training

Because of the elimination of the confusions specific to variant 1 in the test because of successive presentation of isolated vowels, it is relevant in variant 2 to compare the performances of the test to the performances of the training phase in order to see if a learning effect can be noticed (table 3). One can note that the tendencies described in the test were already at play in the training. The fact that these result were obtained in a first contact to whistled vowels with only 20 vowels support the fact that the subjects rely on categorizations that are already at play in their linguistic use.

Table 3: Confusion matrix of 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

5.3. Discussion

The results obtained for the two versions of the experiment of identification show that French subjects without any idea of whistled speech or of the frequency distribution in scales of the Spanish whistled vowels, succeed in categorizing intuitively them even if they listen such sounds for the first time in their life. The distribution of their answers is similar to the cognitive representation of the whistlers. The fact that this ability is already stable during the training phase shows that this perceptual representation in frequency scales already exists in the tested subjects with /i/ identified as an acute vowel and /o/ as a low one and /e/ and /a/ in between with /e/ a little more high in pitch than /a/. This suggests that it is also useful to identify the French vowels /i, e, a ,o/. Moreover these results confirm that the whistlers rely on a perceptive reality at play in usual speech to transpose the vowels in whistled frequencies.

6. General discussion

The present study has examined the strategy of whistled speech and its reflection in the phonetics of several types of whistled languages. These linguistic practices have been found by the author to be much more spread on the planet than the literature mentions it. They are defined here under the terminology *style of speech* and it has been shown that they are the logical continuity of shouted voice. While having an acoustic behaviour of a real telecommunication system, whistled forms of languages are also identified in this study as natural phonetic descriptions of salient key features of a language. The direct consequence is that whistled speech classifies the world languages in frequency types, the main distinction being between tonal and non-tonal languages. One of these language types, represented here by Greek, Turkish and Spanish, is particularly interesting to further elucidate the functional use of the formant distribution in vowel qualities and consonants. The analysis of newly collected data of these three languages shows that their vowels are organized in frequency scales. The consonants are whistled in a combined frequency and amplitude modulations of the neighbouring vowels. They can be grouped in classes in function of their articulatory loci, which is common to usual spoken speech. Moreover, this paper shows through a psycholinguistic experiment that the frequency distribution of whistled vowels is also perceptively relevant for non-whistlers. Indeed, French subjects knowing nothing about whistled languages categorize Spanish whistled vowels /i, e, a, o/, which also exist in French language, in the same way as the cognitive representation of Spanish whistlers, and this without any learning. This suggests that the subjects already used a frequency scale to identify the spoken vowels.

But which part of the spoken energy distribution in the frequency spectrum is used both by whistlers and speakers to identify the vowels? Several researchers thanks to various tasks of identification, discrimination or matching have tested the mechanism of vowel perception. The experiments with closer results to the present study were based on the notion of « *effective upper formant* » (Carlson & al. 1970, Bladon & Fant 1978) also called F2'. F2' is the derivation of formant 2 (F2) at variable degrees in order to take into account the upper frequency values. This formant is therefore considered as the perceptual integration of all the upper formants (above formant F1).

In Carlson et al (1970) the experiment consisted in the examination of the perceptual drawbacks of the manipulation of the spacing between two spectral prominences in synthetic vowels. The results showed that for F1 and F2 relatively close (from 3 or 3,5 Bark) the subjects placed F2' at the level of F2 showing therefore a perceptual integration on the peak with more spectral density (see vowels /a/, /o/, /u/ on figure 23). Carlson & al (1970) also noticed that for vowels for which the formants F2 and F3 where the closest, the subjects placed F2' in between these two peaks of energy (typically see vowels /e/ or /ae/ on figure 23). Finally the /i/ was the only vowel of their experience for which the subjects placed F2' between F3 and F4. Bladon & Fant (1978) reproduced the same experiment from a corpus of cardinal vowels with similar results. Later Chistovitch and her team have confirmed the phenomenon of *perceptual integration* (Chistovitch & Lublinskaya 1979 ; Chistovitch & al 1979 ;

Chistovitch, 1985) for a 3,5 bark proximity between two of the formants as well as Escudier & al (1985).

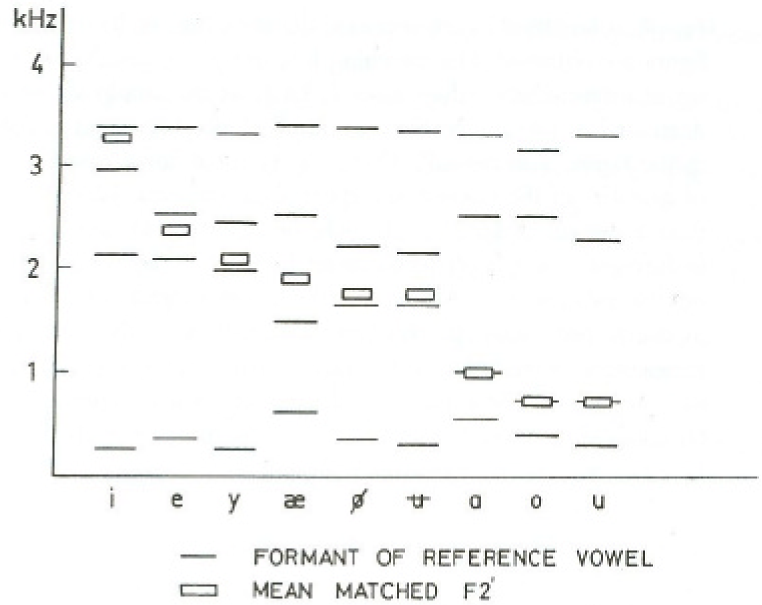

Figure 23. Results of the experiment of Carlson & al (1970) (in Stevens, 1998: 240)

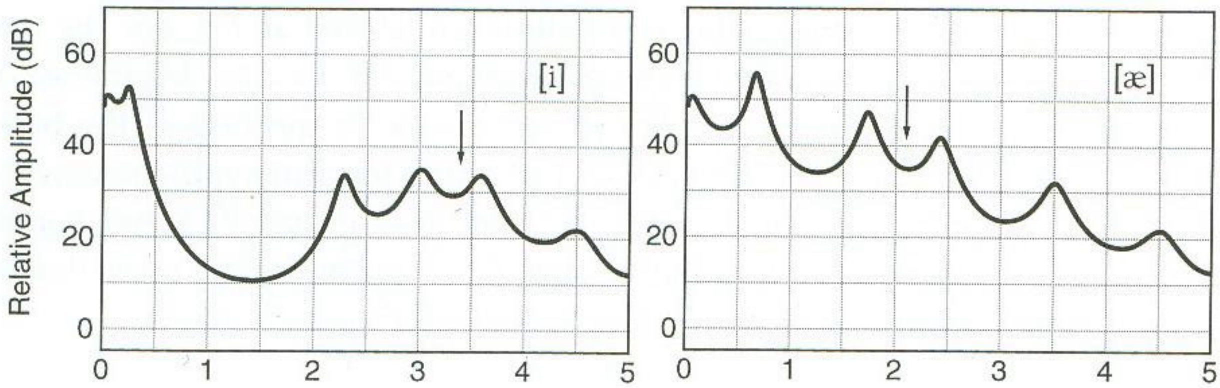

Figure 24: Localisation of the F2'(arrow) on the spectral analysis of two types of vowels (Stevens, 1998: 289)

The distribution of the *effective formant* F2' seems particularly adapted to a comparison with the strategy of transposition of vowels into whistles in non-tonal languages. Indeed, one can find again the clear distinction between /i/ and the other vowels that was found for whistled Turkish, Greek and Spanish, but also the grouping of posterior and central vowels in two different categories. Moreover the F2' provides a more perceptively reliable explanation than the theory of transposition of the only Formant 2 proposed by Rialland (2003). By underlining the role of the perceptual integration of two formants of speech in one F2' these researches highlight the importance of highly compact areas in the perception of the vowel quality. According to Stevens they point the role of the critical spacing and have strong consequences for the classification of vowels: *« some aspects of the auditory system* *undergoes a qualitative change when the spacing between two spectral prominences becomes less than a critical value of 3,5 bark* » (Stevens 1998: 241). He made an illustration by situating the F2' on a spectral analysis of some vowels (figure 24). One can see that the common band of frequencies of the two formants forms a compact area. Inside this grouping of frequencies the spectral valleys are not profound and not wide whereas they are more profound and wider around. Whistled languages support that these conditions are key acoustic cues that ease the perceptive prominence for speech perception. Boë and Abry (1986) called such a convergence of two consecutive formants *focalisation* and showed that this mechanism was the basis of the stability of a decreasing spectral variability and an increasing acoustic salience (Abry & al 1989). This led them to propose that focalisation could provide a benefit for speech perception. Schwartz & Escudier (1989) gave a first evidence for this with a discrimination experiment proving that patterns of synthetic vowels with the greatest formant convergence were more stable in short term memory and produced a lower level of false alarm.

For whistler and for the subjects of the experiment of this paper, the easy correspondence made between nearly pure frequencies of whistles and the four types of vowels /i, e, a, o/ supports that frequency convergence of the vowel prototype they had in memory helped them. The /i/ is marked by a convergence of F3 and F4, it's timbre can be qualified of bright or « acute », the /e/ is intermediary, marked by a convergence of F2 and F3, while /a/ and /o/ have timbres sometimes qualified as « low ». These terms, borrowed to musical acoustics correspond well to the reality of whistles. Actually, formant convergence is one of the aspects highlighted by whistled speech that ease the salience of a perceptive flux in an auditory image according to the auditory scene analysis (Bregman 1990).

At last, whistled languages provide a relevant way to listen to both language diversity and language cognitive processes because they give a complementary insight to both phonology and phonetics. It has been shown here for the first time that they represent a strong model to investigate the perception of languages. All these assets are tempered by the fact that whistled speech rapidly looses vitality nowadays in all the concerned cultures because it is linked to traditional rural ways of life. The alternative point of view on the language faculty that they carry has long been neglected, however it has still say many things to say both at the cultural and scientific level. That is why an international network of research on this subject has been developed under the control of the local traditional leaders.

Acknowledgement

I would like to thank the whistlers and the cultural leaders who took some time to work with me on the field, the engineer Laure Dentel for her volunteer recording during my fieldwork and her advices in programming, Prof. René-Guy Busnel for his strong scientific support during the last 4 years, Bernard Gautheron for his advice, the organisers of the FIPAU 2006 Forum for their help to invite two Siberian Yupik whistlers in France, Ross Caughley for lending me some material in Chepang, the staff of the DDL Laboratory for providing advices and technical material, the team of the Laboratory of Applications in Bioacoustics (LAB) of the Polytechnic University of Catalunya (UPC). This research

was a third financed by a BDI Ph.D. grant of the CNRS, a third by a Post-Doc Fyssen Foundation grant and a third by personal financial resources.

Bibliography

Abry, C., Boë, L. J. & Schwartz, J. L. (1989). Plateaus, catastrophes and the structuring of vowel systems. *Journal of Phonetics,* 17, 47-54.

Bladon R. A. W. & Fant G. (1978). A two-formant model and the cardinal vowels. *STL-QPSR,* 1-1, 1-12.

Boë, L. J. & Abry, C. (1986). Nomogrammes et systèmes vocaliques. *Actes des 15èmes journées d'études sur la parole*. Société Française d'Acoustique, 303-306.

Bregman A. S. (1990). *Auditory scene analysis. The perceptual organization of sound.* Cambridge, Massachussetts: MIT Press.

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.

Busnel R-G, Moles A. & Vallencien B. (1962). Sur l'aspect phonétique d'une langue sifflée dans les Pyrénées françaises. *Proceedings of the International Congress of Phonetical Science, Helsinki*, 533-546. The Hage : Mouton,

Busnel R-G., Alcuri G., Gautheron B. & Rialland A.(1989). Sur quelques aspects physiques de la langue à ton sifflée du peuple H'mong. *Cahiers de l'Asie du Sud-Est*, 26, 39-52.

Carlson R., Granström B. & Fant, G. (1970). Some studies concerning perception of isolated vowels, *STL-QPSR* 2-3/1970, 19-35.

Carreiras M., Lopez J., Rivero F. & Corina D. (2005). Linguistic perception: neural processing of a whistled language. *Nature*, 433, 31-32

Caughley R. (1976). Chepang histled talk. In Sebeok T. and Umiker-Sebeok D.J. (Eds). *Speech surrogates: drum and whistle systems*, 966-992. Paris: Mouton.

Chistovitch, L. A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, 77, 789-805.

Chistovitch, L. A. & Lublinskaja, V. V. (1979). 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.

Chistovitch L. A. Sheikin, R. L. & Lublinskaja, V. V. (1979). Centres of gravity and spectral peaks as the determinants of vowel quality. In *frontiers of speech communication research* (ed. Lindblom and S. Öhman), 143-157. New York: Academic Press.

Classe A. (1956). Phonetics of the Silbo Gomero. *Archivum linguisticum,* 9, 44-61.

Classe A. (1957). The whistled language of La Gomera. *Scientific American*, 196, 111-124.

Cowan G. (1948). Mazateco whistle speech. *Language 24*, 280-286.

Cowan G. (1976). Whistled Tepehua. In Sebeok T. and Umiker-Sebeok D.J. (Eds). *Speech surrogates: drum and whistle systems*, 1400-1409. Paris: Mouton.

Cowan N. & Philip A. Morse. 1986. The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*. 79(2), 00-507.

Dimou A. L. & Dommergues J. Y. (2004). L'harmonie entre parole chantée et parole lue: comparaison des durées sylla biques dans un chant traditionnel grec. In. *Proceedings of Journées d'Etudes de la Parole*, 177-180.

Escudier, P., Schwartz, J.L. & Boulogne, M., 1985. Perception of stationary vowels : internal representation of the formants in the auditory system and two formant models. *Franco-Swedish seminar,* Société Française d'Acoustique, Grenoble. 143-174.

Gay T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*. 63, 223-230.

Green D. M. (1985) Temporal factors in psychoacoustics, In A. Michelsen (ed) *Time Resolution in Auditory Systems*, 122-140. Berlin: Springer-Verlag.

Handley T. D. & Steer J. (1949). *Journal of speech and Hearing disorders*. 14, 363-368.

von Helmholtz H. L. F. (1862). *On the Sensation of Tone*. Green & co. (4th ed). London: Longmans.

Jacobson, S. A. (1985). Siberian Yupik and Central Yupik prosody. In M. Krauss (Ed.), Alaska Native Language Center. *Yupik Eskimo prosodic systems: descriptive and comparative studies.25-46*

Lindblom B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America,* 35, 1773-1781.

Lindblom B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle and A. Marchal (eds.), *Speech production and speech modelling*, 403-439. Dordrecht: Kluwer.

Lombard E. (1911). Le signe de l 'élévation de la voix. *Annales des maladies de l 'oreille, du larynx, du nez et du pharynx*, 37, 101-119.

Meyer J. (2005). *Description typologique et intelligibilité des langues sifflées: approche linguistique et bioacoustique*. Ph.D thesis. Université Lyon 2.

Meyer J. & Gautheron B. (2005). Whistled speech and whistled languages. In *Encyclopedia of Language and Linguistics 2nd Edition*. Elsevier.

Moles A. (1970). Etude sociolinguistique de la langue sifflée de Kuskoy. *Revue de Phonétique Appliquée*, 14/15, 78-118.

Oppenheim, A. V., Lim, J. S., Kopec, G. & Pohlig, S. C. (1979). Phase in speech and pictures, *Proceeding of IEEE, ICASSP* 1979, 632-637.

Padgham M. (2004). Reverberation and frequency attenuation in forests - implications for acoustic communication in animals. *Journal of the Acoustical Society of America*, 115, 1, 402-410.

Plomp R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America*, 41,1526-1533.

Rialland A. (2003). A New Perspective on Silbo Gomero, In Solé M.J., Recasens D. & Romero J., (eds), *Proceedings of the 15th ICPhS conference*, 2131-2134. Barcelona.

Risset J. C. (1968). Sur certains aspects fonctionnels de l'audition, *Annales des Télécommunications*, 23, 91-120.

Risset J. C. (2000). Perception of musical sound: simulacra and illusions. In Tsumotu Nakada (ed.), *Integrated Human Brain Science: Theory, Method, Application* (Music), 279-289. Elsevier.

Schwartz, J. L. & Escudier, P. (1989). A strong evidence for the existence of a large scale integrated spectral representation in vowel perception. *Speech communication*. 8, 235-259.

Sebeok T. & Umiker-Sebeok D.J. (1976). *Speech Surrogates: Drum and Whistle Systems*. La Haye and Paris: Mouton.

Shepard R. N. (1968) Approximation to uniform gradients of generalization by monotone transformation of scale. In D. I. Moskosky (ed.), *Stimulus generalization*, 343-390. Stanford: Stanford University press.

Stevens K. N. (1998). *Acoustic Phonetics*. Cambridge: MIT Press.

Stevens S. S. & Davis H.(1938). *Hearing: Its Psychology and Physiology*. New York: Wiley.

Trujillo R. (1978). *El Silbo Gomero: analisis linguistico*. Ed. I. Canaria -I. Andres Bello, Tenerife.

Trujillo, R., Morera M., Guarro A., Padrón U., Ortiz I., Botanz R., Darias E. (2005). *El Silbo Gomero : materiales didácticos*. Islas Canarias, Consejería de educación, cultura y deportes del Gobierno de Canarias - Dirección general de ordenación e innovación educativa.

Wiley R. H. & Richards D. G. (1978). Physical constraints on acoustic Communication in the Athmosphere : Implications for the Evolution of Animal Vocalizations. *Behavioral Ecology and Sociobiology*, 3, 69-94.

Xirometis N. & Spyridis H. C. (1994). An acoustical approach to the vowels of the village Antias in the Greek Island of Evia, *Acustica*, 5, 425-516.

Zeng F. G, Nie K, Stickney G. S., Kong Y. Vongphoe M., Bhargave A., Wei C. & Cao K. (2005). Speech recognition with amplitude and frequency modulation. *Proceedings of the National Academy of Sciences*, 102, 7, 2293-2298.