The multiple dimensions of language sound systems

Contribution of African and Amerindian languages

Didier Demolin

MOTS CLES. – Systèmes sonores, amarique, hadza, clics, karitiana, nasa Yuwe.
RESUME. – Cet article souligne la contribution des langues africaines et sud-américaines à la connaissance des mécanismes de production et de perception de la parole. Il passe en revue quelques-unes des dimensions qui sont impliquées dans la réponse aux questions que posent la compréhension de la structure et du fonctionnement de la parole humaine. Ces questions sont illustrées par l'examen de la production et de la perception des clicks en Hadza et de la nature des systèmes vocaliques de deux langues sud-américaines, le Karitiana et e Nasa Yuwe.

KEYWORDS. - Sound systems, Amharic, Hadza, Clicks, Karitiana, Nasa Yuwe.
ABSTRACT. - This paper highlights the contribution of African and South American languages to the knowledge of the mechanisms of speech production and perception. It also reviews some of the dimensions that are involved in answering the questions raised by research on the structure and working mechanisms of human speech. These questions are illustrated by examining the production and perception of clicks in Hadza and the nature of vowel systems of two South American languages, Karitiana and Nasa Yuwe.

TERFWOORDEN. - Geluidssystemen, Amharic, Hadza, Clicks, Karitiana, Nasa Yuwe.
SAMENVATTING. - Dit artikel belicht de bijdrage van Afrikaanse en Zuid-Amerikaanse talen aan de kennis van de mechanismen van spraakproductie en perceptie. Het evalueert enkele dimensies die een rol spelen bij het beantwoorden van vragen over de structuur en de werking van menselijke spraak. Dit wordt gedaan aan de hand van onderzoek naar de productie en perceptie van kliks in Hadza en de aard van klinkersystemen van twee Zuid-Amerikaanse talen, Karitiana en Nasa Yuwe.

1. Introduction

The study of languages' sound systems is associated with the phonetic and phonological modules of human language. Phonetics seeks to describe and compare the sounds produced and perceived in human languages. Phonology shows how sounds work in particular language systems. In phonology, the distinctive sounds of each language are categorized as a finite number of abstract units called phonemes.

There is considerable diversity among the phonological systems of the world's languages. Languages like Rotokas spoken in Papua New Guinean (West Bougainville family) and Hawaiian (Austronesian family) have about a dozen phonemes while the languages of the Khoesan family of southern Africa may have as many as 141. How can this diversity be explained? What may be its underpinning constraints? What are the main principles, biological and physical, upon which we should base an explanation of the production and perception of speech sounds?

To answer these questions, it is necessary to observe and study as many phonetic and phonological systems as possible, based on biological and physical principles and on experimental methods. There are about 6,000 languages in the world, but too few of

them have so far received a detailed description of their phonetic and phonological systems based on the criteria discussed above. It is likely that some answers to these questions will come from the description and understanding of as yet little studied systems.

The objectives of this article are, on the one hand, to emphasize the contribution of African and South American languages to our knowledge of the mechanisms of speech production and perception, and on the other hand to review some of the dimensions that are relevant for answering questions about the structure and function of human speech. The central question being, of course, to understand *how* and on which fundaments human speech is structured.

2. Speech and its dimensions

How can we define the phenomenon of speech, beyond reference to one specific language? Speech can be defined as one of the communication modalities (and even the main modality) of human language. Speech is essentially a signal, an acoustic wave varying over time with modulations of amplitude and frequency. These modulations are due to the articulatory movements of the organs of the vocal tract. Motor controls are necessary to achieve these movements whose interactions with aerodynamic parameters (pressure and air flow in the vocal tract) produce the acoustic signal. What are the dimensions involved in the working mechanisms of human speech?

The first dimension that we can immediately think of is symbolic. A phonetic or orthographic transcription represents the symbolic encoding of speech sounds. In this regard, it must be remembered that in the International Phonetic Alphabet (IPA), symbols represent sounds observed in the phonological systems of languages, (they are phonemes) therefore the IPA is not the code of speech. This observation therefore raises a fundamental question: is there a speech code and, if so, what are its basic units? What could this code be based on? Distinctive features? Articulatory gestures? Are there invariants? What are the primitives involved?

This perspective leads us to consider an important issue, about which linguists are still far from reaching consensus, namely phonological universals. Hombert & Ohala (1978) propose the following hypothesis to characterize them: '… *We will define phonological universals as follows: by observing similar patterns of phonological phenomena in many languages chronologically, geographically and genetically distant, we will examine, of necessity, the physical properties of the systems of production and perception of human speech. This is to explain the origin and directionality of these sound patterns'*. This hypothesis makes it possible to integrate the dimensions of speech production and perception which are necessary to determine a universal set of phonological features. The debate about their innateness, or not, is outside the scope of the present paper and will therefore not be discussed. The question of the code, and in a way of universals, immediately raises another one: what are the limits of phonetic systems? Are these systems open or closed? The IPA framework shows that there are combinations of places and modes of articulation that are not yet observed or not established in the languages of the world, such as the bilabial approximant of Karitiana that are shown in Figure 10 for example. What conditions their presence in the languages where they are met? Biomechanical developments or simple products of the

diachronic evolution of systems? Considering sound systems as open or closed is not trivial, because the answer ties up with the evaluation of complexity in the sound systems of languages. We will see later that open systems can evolve towards a complexification of their structures and that this makes it possible to formulate hypotheses about the evolution of complex structures that are found in languages.

The anatomy of the speech apparatus, although its overall shape is similar in all humans, is an important dimension in understanding variation and certain constraints exerted on the acoustic output of sound productions. The difference in the length of the vocal tract and vocal folds (often called vocal cords) between women and men is a mark of sexual dimorphism. It exerts an influence on the fundamental frequency which is higher in women (on average around 220 Hz) than in men (on average around 120 Hz), as well as on the timbre of the voice which is higher in women than men, by the presence of higher frequency components in the spectrum of women's voices. This is a consequence of the difference in average length of the vocal tract between women and men. The importance of anatomical features has also been raised by Traill (1985), Sands et al. (1996) and Nakagawa (2010), who showed that speakers of !xóõ, G|ui, Hadza and Dahalo do not have a marked alveolar ridge behind the teeth. This has led Moisik and Dediu (2020) to study the possible effect of vocal tract morphology on the ability to learn to produce clicks. Their work investigates whether the failure or substitution of click production in a learning task is biased by the shape of the vocal tract and particularly by the anterior part of the palate. So far, no concluding results have been obtained but this kind of study opens paths to test empirical observations and reinforces the necessity to obtain strong experimental data on these phenomena.

The physiology of speech is a dimension that involves the respiratory and laryngeal systems and the functioning of the vocal tract. These complex phenomena affect the position, shape and volume changes of the vocal tract and result from the coordinated movements of the articulators: the tongue, the mandible, the soft palate and the lips. Controlling the activity of the respiratory system plays a crucial role in the production of speech, which is typically produced with an egressive pulmonary airflow. Several speech-related phenomena are directly linked to the management of breathing during speech: management of the breath group, control of subglottic pressure (Ps) and increase of Ps to produce emphatic stress (Benguerel 1970, Bouhuys 1977, Fant 2000). We will see that articulatory mechanisms and their synchronization play an important role in explaining the functioning and production of certain classes of sounds such as non-pulmonary consonants. These require the precise synchronization of multiple articulators to generate the acoustic targets. This points to the dynamics, and therefore the temporal dimension, of the articulatory movements involved in the production of sounds. These self-organized phenomena are at the origin of what can be described as state of the vocal tract variables that generate the acoustic targets. Figure 1 illustrates this point with aerodynamic data inferring the movement over time of some articulators[1].

Figure 1 makes it possible to infer different articulatory events from the aerodynamic parameters of air flow and intraoral pressure on sound production. Points 1 to 5 on the audio waveform indicate successively: 1 the maximum of the turbulence noise of the first alveolar ejective fricative /s'/; 2 an interruption due to the glottal occlusion which marks the end of the ejective fricative, that is produced with the glottis

closed; 3 a peak of oral air flow at the start of the second fricative, due to the opening of the glottis followed by its closure marked by the interruption of oral air flow; 4 the maximum of turbulent noise of the second alveolar ejective fricative /s'/. The gradual increase in the turbulence marks the vertical movement of the larynx characteristic of the ejective consonant; 5 reflects the same process as described for 2. Points 6 and 7 on the Po plot show an increase in intraoral pressure to a peak followed by a gradual descent. From this curve, it is possible to infer the duration of the up and down movement of the larynx characteristic of the production of ejective consonants[2].

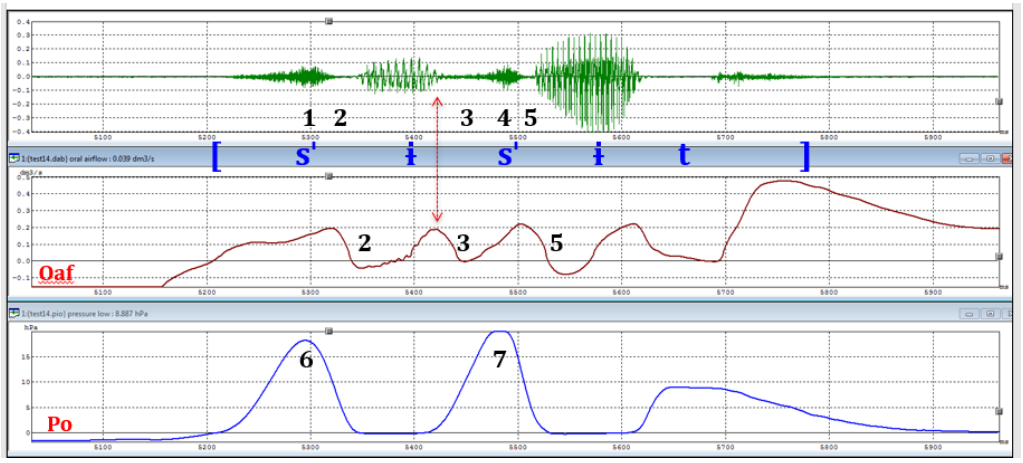


Figure 1. Audio waveform, oral air flow (Oaf) in dm^3/s and intraoral pressure (Po) in hPa (Hecto Pascal) of the word /s'ɨs'ɨt/ "regret" in Amharic.

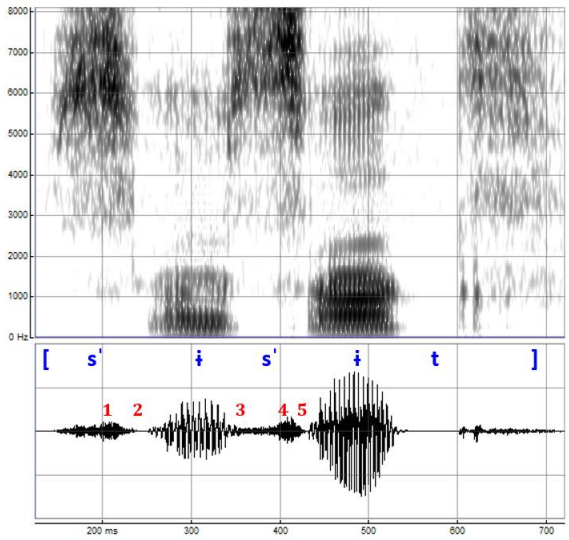


Figure 2. Wideband spectrogram and audio waveform of the word /s'ɨs'ɨt/ 'regret' in Amharic. The figure corresponds to the aerodynamic data presented in Figure 1.

The spectrogram in Figure 2 shows the acoustic product of the interaction of articulatory and aerodynamic parameters involved in speech production of the same word represented in Figure 1[*2]. The spectrogram gives three parameters of the sounds

produced: the frequency on the ordinate, the time on the abscissa and the intensity of the sounds reflected by the degree of darkness of its parameters. The fricative turbulent noise (a white noise) is represented by the irregular structures of varying intensity located above 4000 Hz. The vertical bars at the end of /t/, marked by a silence on the audio signal, mark the release of the vocal tract closure produced during the alveolar occlusion. The interruption of airflow for this voiceless consonant, where vocal folds do not vibrate, can be seen on the oral airflow plot in Figure 1. The interaction of articulatory and aerodynamic parameters over time clearly shows dynamic aspects of speech production.

The perceptual dimension will be discussed several times in what will follow, but the interpretation of the data from the spectrogram in Figure 2 helps to show its role in the working mechanisms of speech. The turbulence noise that characterizes fricatives is most intense (the plot is darker) above 5000 Hz. This noise starts and ends abruptly and lasts for about 100 ms. How is it generated and how is it perceived? Signorello et al. (2019) showed that the beginning and end of fricative consonants are determined by thresholds in the ratio between subglottic pressure (Ps), intraoral pressure (Po) and atmospheric pressure (Pa). The level of Po has to reach a certain level for frication noise to be generated at the constriction place. The Δ between Po and Pa must be above a threshold for frication noise to occur. This is easily reached for voiceless fricatives as the glottis is open. Voiced fricatives simultaneously necessitate a sufficient Δ between Ps and Po and between Po and Pa. This is not easily reached because the adducted glottis impedes Po to rise very high. This explains why voiced fricatives are among the least frequent fricatives in the world's languages. Perceptually the noise at Figure 1 is easily interpreted. The lowest zone of the frication noise (its center of gravity) makes it possible to differentiate between the fricatives, for example between /s/ and /ʃ/ in the French words *sans* 'without' [sã] and *chant* [ʃã] 'song'. The frication noise is higher for /s/ (there are more high frequencies in the spectrum) compared to that for /ʃ/.

These aspects of sound production in Amharic clearly show the complexity and interplay of several dimensions in explaining the way sounds are produced, perceived and represented.

3. Clicks and their complexity

The study of the non-pulmonary consonants found in the Khoesan languages and in two isolates in Tanzania (Hadza and Sandawe) reveals the interaction of several dimensions in the production and perception of speech. It also allows us to ask fundamental questions about the diachrony and even the phylogeny of speech sounds.

Hadza, spoken by just under 1000 people in the Lake Eyassi region of Tanzania, is a language that has clicks in its phonological inventory. There are 4 of them, bilabial [ʘ], dental [|], lateral [||] and alveolar [!]. These clicks can have various accompaniments: glottal, aspirated, nasal, glottal nasal for the dental click [|]; simple, aspirated, nasal glottal, nasal for the lateral click [||] and simple, nasal, aspirated, and nasal glottal for the alveolar click [!] (Demolin at al. 2021).

Clicks are defined as non-pulmonary consonants because the sound initiation mechanism does not come from the pulmonary air flow, but from a lingual airstream mechanism which creates an ingressive air flow (see Miller et al. (2009) for the

definition of the lingual airstream mechanism). The description of the alveolar click [!] illustrates this point (Figure 3)[3]. The production of the alveolar click [!] (a noise which is similar to that of a fairly brief snap) is done as follows: a first phase corresponds to a double occlusion in the vocal tract: one in the alveolar region (1) and another in the velar or sometimes uvular region (2). The second phase, which lasts about 120/150 ms, corresponds to an expansion of the cavity between points 1 and 2. The consequence of the increase in volume between points 1 and 2 (those of the double occlusion) is that the pressure in this volume decreases (Ps- in image 2) compared to atmospheric pressure (Pa). The release of the alveolar closure (which gives its name to the click) with the maintenance of the second occlusion will cause an ingressive air flow since the oral pressure (Po) is lower than atmospheric pressure. The acoustic effect results in the abrupt, grave noise characterizing the alveolar click.

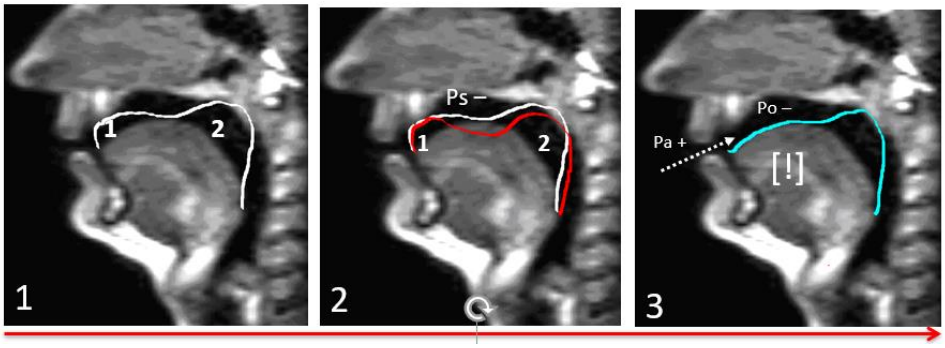


Figure 3. Phases of the production of the alveolar click [!]: 1 double occlusion (1 alveolar and 2 velar/uvular); 2 increase in the volume of the cavity between the two occlusions (1 & 2) with a decrease in pressure (Ps-) included in this volume and 3 relaxation of the alveolar occlusion and ingressive air flow due to the difference between atmospheric pressure (Pa) and oral pressure (Po).

Figure 4 illustrates four realizations of the alveolar click [!] in Hadza (simple, aspirated, nasal and nasal glottal) in the words: /la!o/ "stumble"; /!ʰaku/ "jump"; /ŋ!ʔoje/ 'beeswax' and /kuŋ!ãna/ 'kudu', Demolin et al. (2021)[4].

Figure 4 also shows in (a) that the noise of the alveolar click release is abrupt and more intense in the low frequencies; (b) the abrupt noise is colored by a resonance around 1 kHz and is followed by a slight aspiration; (c) the abrupt noise is colored by resonances around 1 and 2 kHz, followed by a period of silence due to the glottal closure; (d) the abrupt noise is preceded by a nasal and is followed by a slight aspiration. This click is therefore voiced. These examples of the alveolar click production show the complexity of the articulation of clicks and their accompaniments.

An important point in describing clicks is to know what are the features or gestures characterizing them better. Traill (1997) clearly showed that an articulatory description is not enough and that it is necessary to use acoustic features to characterize them. The alveolar click that we have just described is then described by the features [abrupt and grave].

The click production mechanism is quite complex but is very efficient in perception as the salience of the release noise makes it very perceptible. One final point worth mentioning is click coding. Most of the consonants in languages interact with the vowels around them and this produces what is called coarticulation phenomena. Clicks do not seem to be affected by these phenomena. Traill (1997) has shown that the

features which set the clicks apart and distinguish them from the burst noises from pulmonary consonants are contained in the release noise (or transient attack). Their noise burst can be extremely brief since it can be of the order of 2 ms at the start of the alveolar click. This is quite extraordinary as Traill points out, since it represents the lower limits of human capacities in the temporal processing of the auditory system (Gerber 1974: 177, Pickles 1977: 91). What Traill's work shows quite clearly is that clicks are salient and exploit fundamental sensitivities of the hearing system. The spectral information contained in the click release noise (the transient attack) is conveyed to the auditory system with such a temporal salience and/or intensity that its perception is assured. This last point emphasizes that the contribution of clicks to the knowledge of speech phenomena is not limited to complex articulations but that their perceptual dimensions also contribute to our understanding of fundamental aspects of human hearing capacities.

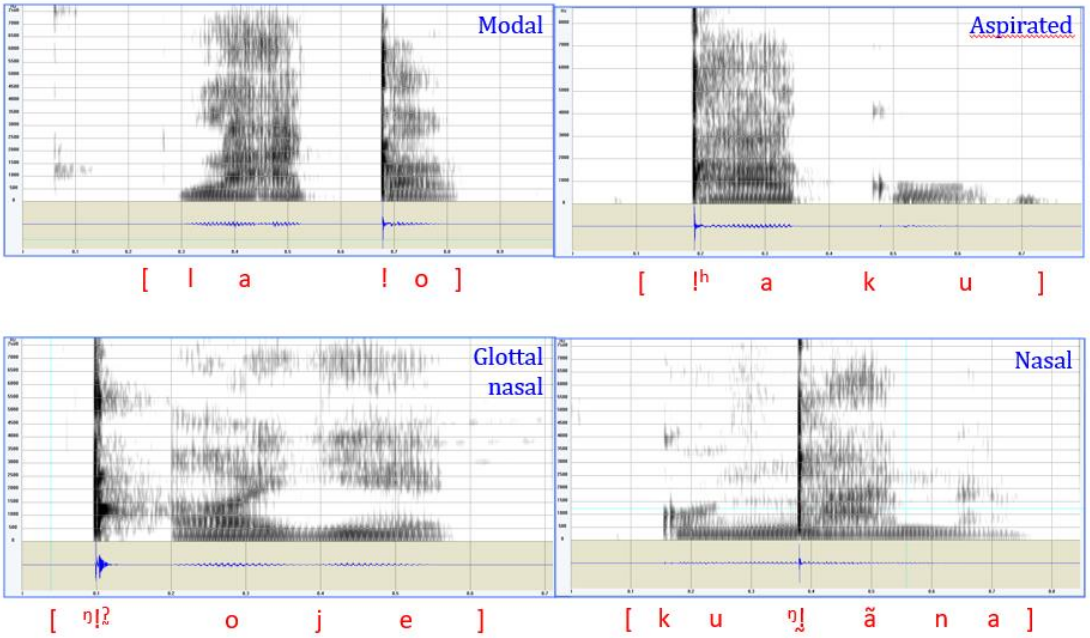


Figure 4. Wideband spectrograms and audio waveform of the words [la!o] 'stumble' ; [!ʰaku] "jump"; [ŋǃʔoje] 'beeswax' and [kuŋ!ãna] 'kudu' in Hadza.

4. The historical dimension of clicks

Is rarity a factor in determining the age of clicks? Are they ancient traces of the first sound systems in human language? Knight et al. (2003) have suggested that clicks could be a sort of residue of the first sounds or sound systems produced by man and of which the speech of Khoesan peoples would be the last witnesses. Sands & Güldemann (2009) points out, contrary to current opinions and to the assertions of Knight et al. (2003), that the emergence of clicks as phonemes in Africa could represent a late episode in the diversification of sound systems in human languages. Sands and Güldemann also suggest that clicks may have developed several times in human history, for various reasons. This would mean that innovation, contact-induced borrowing and the transmission of clicks are more frequent than is usually thought and that the emergence of clicks in Africa would represent a late episode in the diversification of

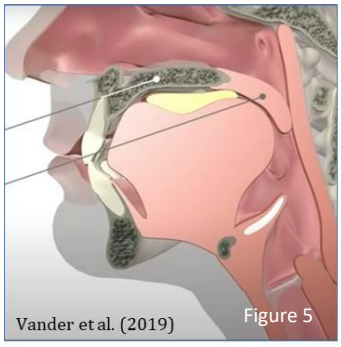
human speech. If we adopt this point of view, the emergence of clicks would allow different types of explanations: retentions with great time depth; independent innovation; or variation in the timing of articulatory movements and contact. It is also important to stress that the emergence of phonemic clicks in present-day Khoesan language varieties is the end result of historical processes rather than a starting point (Kusimba 2003). The emergence of clicks in Khoesan languages also occurs in an area where there is a great diversity of sounds and systems and therefore where the possibility of the emergence of a greater complexity of sounds and sound systems increases. The current sound systems including clicks and the reconstruction of the old form (the proto-system) would reflect the different language families as basins of attraction centered on particular types of sounds, if we considers language and their sound systems as dynamical systems.

Along the same lines as Sands and Güldemann (2009), Mayer et al. (2017), Demolin (2021), and Demolin and Kingston (2021) hypothesized that the movements observed during swallowing are recruited into the production of consonants produced with non-pulmonary airflow mechanisms. These consonants are the ejectives and implosives produced by a glottal initiation mechanism and especially the clicks and labio-dorsals produced with a lingual initiation mechanism. The basis for this hypothesis comes from the fortuitous observation of movements of the tongue and larynx in the swallowing of saliva between statements made in the context of real-time MRI recordings of French speakers pronouncing short sentences. The examination of the phases of swallowing is shown in Figures 5 to 8, which compare this mechanism with that which is involved in the production of clicks. These data show that the similarities between the movements observed in swallowing and the production of clicks are quite obvious. This hypothesis, which has yet to be supported by experimental data, is part of an embodied speech theory which specifies that speakers have learned to control certain highly specialized parts of the body to produce sounds.

The presence of the Hadza and Sandawe, in Tanzania, mostly hunter-gatherers until recently, shows the presence of languages with clicks sounds in East Africa. The disputed affiliation of Hadza to the Khoesan family leads to the hypothesis that other linguistic families, now extinct, have come into contact with the Khoesan languages. The Hadza, sometimes considered as an isolate, would be a final witness. The production mechanisms of clicks and ejectives in Khoesan and Cushitic suggest very marked links in the production of these sounds and therefore possibilities of borrowing or diffusion following linguistic contacts. The production of these sounds shows that they have mechanisms that appear to resist contact, even as they change and adapt to new constraints. If the hypothesis of the reuse of the swallowing mechanism is correct, it makes it possible to account for the presence of clicks in Khoesan and related languages not as old remnants, but as proofs of the complexification of the languages' sound systems over time. It will also be possible to assess the role of non-pulmonary consonants as evidence in the study of the short-term and long-term history of languages in this region of Africa.

In the same vein, the labio-dorsal consonants [kp, gb, gɓ, kʁ̥, qp, qɓ, ɢɓ] which are found in some African languages, are produced with an ingressive airstream mechanism similar to that of clicks (Ladefoged 1968). This amounts to saying that the production of these sounds could also be based on the reuse and specialization of movements made

during the mechanisms of swallowing. The acoustic characteristics of labio-dorsals make them sound less salient than clicks, especially when the lip occlusion is released, but the similarity in ingressive airstream mechanism shows that this type of sound is therefore much less rare than what we generally think, since it is shared by dozens, if not hundreds, of African languages.



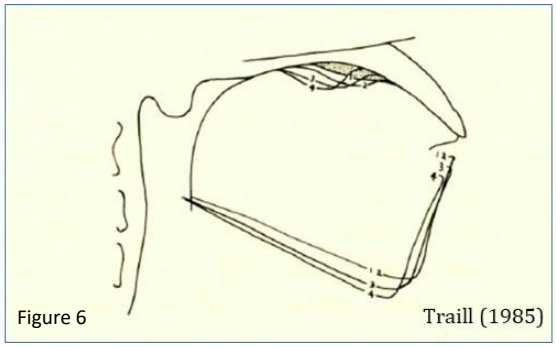


Figure 5. Double occlusion observed at the start of the swallowing phase.

Figure 6. Tongue and jaw contours extracted from an X-ray film of the jaw movement and the part of the tongue between the two occlusions during the production of a palatal click [ǂ]. This sequence shows the volume increase in the oral cavity between the two occlusions.

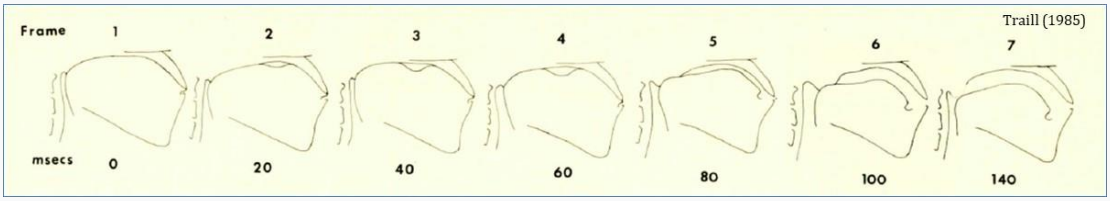


Figure 7. Sequence of tongue movements recorded every 20 ms during the production of a palatal click [ǂ].

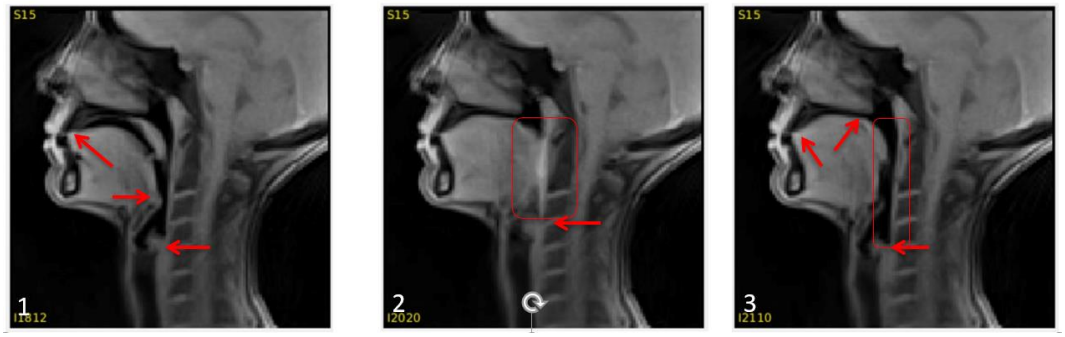


Figure 8. Sequence of swallowing movements obtained by real-time MRI, in the mid-sagittal plane. 1 shows the preparatory phase, the arrows indicate the position of the tongue apex, the tongue root and the relative position of the larynx observed in relation to the position of the arytenoids. 2 the backward propulsion phase where one can observe, in the box, the contact of the posterior part of the tongue and the upper pharyngeal constrictor and also the elevation of the larynx indicated by the arrow. 3 The return to the position which precedes the propulsive phase where one can observe the contact of the apex on the front part of the hard palate and the back of the tongue in the palatal region indicated by the arrows. The box shows the pharyngeal cavity and the horizontal arrow shows the position of the arytenoids and larynx (Demolin & Kingston 2021)[5].

The hypothesis of reusing the phases of swallowing to produce clicks falls within the theoretical framework of embodied phonetics which studies the neurophysiology and biomechanics of speech, Gick et al. (2014), Mayer et al. (2017). This research program shows that the vocal tract offers only a small inventory of reliable and biomechanically robust actions that are exploited over many repetitions for linguistic expression, providing a basis for universals. This is consistent with a phonetic theory built on an inventory of functionally defined primitives, each of which fulfills a specific function in speech (Gick & Stavness 2013). An appropriate model for the sounds of speech in an embodied theory of phonetics is not that we learn to control an inventory of sounds, but rather that we build and control certain highly specialized body parts, each of which is constructed by the use and optimization, to fulfill a specific phonetic function (Gick 2019). The hypothesis of the reuse of swallowing mechanisms, which has yet to be tested experimentally, suggests a complexification in the production of speech sounds.

5. Dispersion of vowel systems and Tupi languages

The theories of dispersion of Liljencrants and Lindblom (1972), Lindblom (1986, 1990), the quantal theory of Stevens (1972, 1989) and of the dispersion focalization of Schwartz et al. (1997) have been the anchor of many descriptive and theoretical works on vowel systems. An important contribution of this work has been to predict the shape of vowel systems by incorporating spaces of perceptual representation. In general, vowel systems account for the principle of dispersion which obeys a principle of maximum or sufficient contrast. If a language has only three vowels, it will be the vowels /i, a, u/ which are perceptually the furthest from each other. Likewise, if a language has five vowels, it will be mainly /i, e, a, o, u/ (Crothers 1978, Maddieson 1994, Vallée, 1994, Schwartz et al. 1997). Crothers (1978) and the optimal systems derived from Lindblom (1986) show that systems with 5 vowels /i, e, a, ɨ, o/ are less frequent but possible. These systems are found in Amerindian languages and in particular in the languages of the Tupian family (Storto & Demolin 2012). Karitiana, a language of the Arikem subfamily inside Tupian (Storto 1999), shows a system of 5 vocalic qualities /i, e, a, ɨ, o/ (Figure 9) which can be oral short or long and nasal short or long[6].

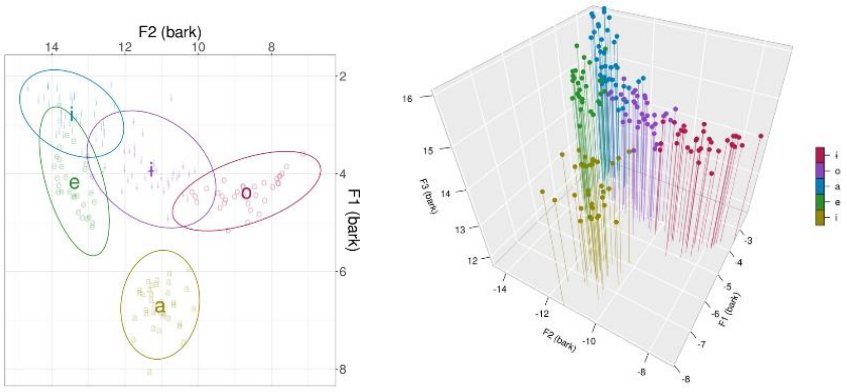


Figure 9. Distribution of Karitiana short oral vowels measured in Bark, in the F1/F2 space, for 5 speakers and in an F1/F2/F3 space.

The absence of a posterior high vowel /u/ in this vowel system can be explained either by a diachronic process or by an intrinsic feature. This phenomenon is found in other languages of the Tupian family. How to explain the absence of the high posterior vowel?

One possible explanation relates to the absence of the rounding feature combined with a palatal vowel prototype. This proposition arises from the fact that the posterior middle vowel /o/ of Karitiana lacks the rounding feature which is characteristic of the posterior mid-closed or middle vowels. Note that this is not an unrounded vowel, it is indeed an /o/ timbre but without rounding. Figure 10 shows the shape of the lips during the realization of two vowels /o/ in the word /koβ̞ot̚/ "sweet". It can be seen that neither on the profile nor on the frontal view of the lips, it is possible to detect any rounding or protrusion of the lips[7]. On the contrary, one can even detect a certain stretching of the shape of the lips. Stavness et al. (2013) have shown that it is possible to simulate the production of the lips' shape and the production of this vowel from a biomechanical model.

Figure 10. Shape of the lips seen from the front and in profile during the realization of two vowels /o/ in the word koβ̞ot̚ 'sweet. Images are taken in the middle of the vowel. The bilabial approximant /β̞/ is also taken in the middle of this articulation.

It is also possible to simulate the production of this vowel from an articulatory/acoustic model which makes it possible to generate the timbre of the observed vowel (Figure 11) on the basis of articulatory parameters: position of the jaw, shape and position of the back of the tongue, opening and protrusion of the lips and vertical position of the glottis. We can then simulate the shape of the vocal tract and the position of the articulators with the mean values of the formants measured on data produced by speakers. This is then compared with that of a rounded vowel of a language like Brazilian Portuguese whose system includes the vowel /o/. This vowel of Brazilian Portuguese has almost the same formant values as the /o/ of Karitiana. This makes it possible to note, as shown in Figure 11, that the /o/ of Karitiana is produced without protrusion of the lips and with the glottis markedly more lowered when comparing its position with that of a Brazilian Portuguese /o/ which is produced with the glottis higher and with a much more advanced lip position. The two vowels show an almost identical F-Pattern in this simulation (F1 378 Hz, F2 1126 Hz and 1920 Hz for the /o/ of Karitiana and F1 319 Hz, F2 1127 Hz and 2025 Hz for the /o/ in Portuguese).

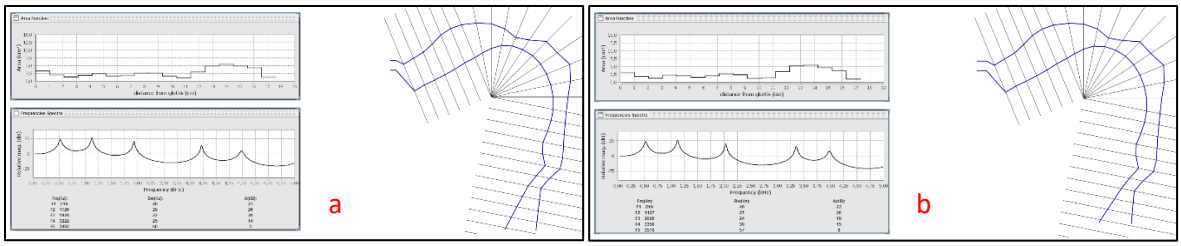


Figure 11. (a) Simulation of the posterior vowel /o/ articulation of Karitiana from the mean values of formants measured on vowels in context. The left part of the mid-sagittal slice shows the area function at the top left and the transfer function at the bottom left. (b) Simulation of the articulation of the posterior vowel /o/ of Brazilian Portuguese from the mean values of formants measured on vowels in context. The left part of the mid-sagittal section shows the area function and the transfer function. The shape of the tongue was not changed in the simulation, only the protrusion and the position of the larynx were changed[8].

Ménard (2002) showed from vowel prototypes, obtained from a MRI database, that the closed posterior vowel /u/ actually has two prototypes for the position of the back of the tongue, a palatal and a velar. The result of the combination of a palatal prototype without lip rounding gives a timbre quite close to the Karitiana /ɨ/ (Figure 12). In Karitiana this vowel is more open and more central than the reference version of the IPA.

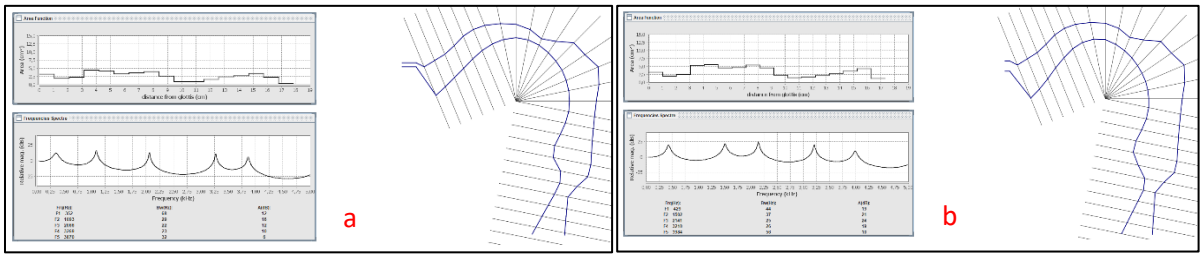


Figure 12. (a) Simulation of the articulation of the posterior vowel /u/ of Brazilian Portuguese from the mean values of the formants measured on vowels in context. The part to the left of the mid-sagittal section shows the area function and the transfer function (The F-Pattern is F1 360 Hz; F2 1093 Hz and F3 2088 Hz). (b) Simulation of the articulation of the central vowel /ɨ/ of Karitiana from the mean values of formants measured on vowels in context. The left part of the mid-sagittal slice shows the area function at the top left and the transfer function at the bottom left. The F-Pattern of Karitiana /ɨ/ is F1 429 Hz; F2 1502 Hz and F3 2143 Hz.

There is another complementary answer based on the size of the space of possible articulatory configurations to produce the [u] vowel compared to other vowels such as [i] as suggested by Potard et al. (2008). Indeed, it appears that there are very few articulatory solutions to produce the vowel [u] compared to [i]. Consequently, the production of this vowel requiring a great articulatory precision articulatory precision, other vowel systems without [u] can emerge to get rid of this difficulty.

The Karitiana data show that predictions made from deductive models are found among the world's languages and particularly in less frequent systems such as those which do not include a high posterior vowel /u/. It should be noted that these systems are also found in other Native American languages such as Navajo, spoken in North America (McDonough 2003). Explaining the shape of these systems involves considering

the articulatory and acoustic constraints of speech production and perception. Lindblom (1986) defines the notion of "possible vowel" considering the following criteria: articulatory, acoustic and perceptual. A system like that of Karitiana makes it possible to show that the predictions of the model of maximum contrast must be refined and there is a need to introduce a concept of sufficient contrast to consider the presence of interior vowels. The framework in which the Karitiana system is described refers to the notion of perceptual contrast. Universal phonetic space is then defined as the relationship between articulatory space and perceptual space.

6. Vowels and phonation types in Nasa Yuwe (Paez)

Vowel systems which have 4 qualities are predominantly /i e a u/, according to Crothers (1978), Lindblom (1986), Maddieson (1994), and Vallée (1994). Nasa Yuwe a language spoken in the Cauca department in Colombia has these vowel timbers, but has a phonological system with 32 vowels, as described by Rojas (1998) and Diaz (2019). This is made possible by adding features of nasality, length, aspiration and glottalization which can also be combined. The vowels can be grouped into two sets. Oral vowels: short /i, e, a, u/; long /i:, e:, a:, u:/; aspirated /i̤, e̤, a̤, ṳ/ and glottals /iʔ, eʔ, aʔ, uʔ/. Nasal vowels: short /ĩ, ẽ, ã, ũ/; long /ĩ:, ẽ:, ã:, ũ:/; aspirated /ĩ̤, ẽ̤, ã̤, ṳ̃/ and glottals /ĩʔ, ẽʔ, ãʔ, ũʔ/.

This vowel system is a superb example of complexification in phonetic and phonological systems implemented by the addition of different phonatory types. The addition of these phonatory types requires motor commands and control in addition to those needed to produce the “common” vowels. Examination of glottal vowels helps to demonstrate this point (Demolin et al. (2016). It should be noted that to date there is very little description of glottal vowels in the languages of the world. The only known other case comes from !xóõ spoken in Botswana (Traill 1985). This shows that the mechanism is not an areal phenomenon, but rather reflects a possibility in the production of speech sounds. A recent model of speech production, the LAM (Laryngeal articulator model) considers that the larynx functions as an articulator in the full sense of the term and that the larynx should no longer be considered as being only the place of the vocal fold vibration (Esling (2005), Esling et al. (2019)). In the LAM, the larynx is made up of several valves distributed vertically. The first three which concern the description of the glottal vowels of Nasa Yuwe are: the vocal folds (valve 1); the ventricular bands (valve 2) and the aryepiglottic folds (valve 3), as in Figure 13.

Figure 13 shows video laryngography that allows visualizing the position of the valves and the configuration of the epilaryngeal tube, which is the small tube that runs from the glottis (the space between the vocal folds) and the basis of the epiglottis. The three images are separated by a 20ms interval.

The images in Figure 14 show a strong insertion of the ventricular bands which largely cover the vocal folds. The epilaryngeal tube is not closed. These images were recorded with a Nasa Yuwe speaker who produced a glottal vowel /iˀ/ in isolation. Data taken in the context of more spontaneous speech show that the glottal character of the vowel is marked at the end of the vowel, by a constriction of the epilaryngeal tube at the level of one of the three valves of the larynx. In any case, it is not a common laryngealization produced by a vibration of the aryepiglottic folds (Esling et al. 2019: 64), but rather a constriction of one part of the epilaryngeal tube. What is visible in the

images in figure 14 shows the configuration of valves 1 and 2 before the tightening of the epilaryngeal tube which appears at the end of the vowel.

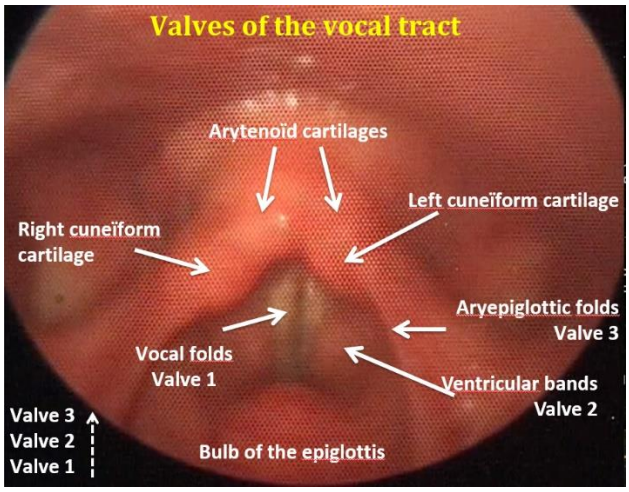


Figure 13. Valves of the vocal tract and position of the main anatomical parameters involved in the production of the glottal vowels of Nasa Yuwe[1]

Figure 14. Position of valves 1, 2 and 3 during production of a glottal vowel in Nasa Yuwe. The three images are spaced by an interval of 20 ms. The figure on the right shows the position of the valves 1 (glottis-vocal folds); 2 ventricular bands and 3 aryepiglottic folds[11].

The interaction of articulatory, acoustic and aerodynamic dimensions in the production of speech sounds is illustrated in Figure 15, where the red arrow points at the final constriction, which appears at the end of the glottal vowel /iˀ/ in a Nasa Yuwe word[12]. The amplitude of the EGG signal (which measures the electrical impedance at the vocal folds) shows a marked decrease in amplitude towards the end of the vowel before a very brief increase. This decrease in amplitude is also visible on the acoustic signal. On the spectrogram, we can observe that it becomes irregular towards the end of the vowel. The detection of the fundamental frequency (F0) is interrupted at this time. This shows that the glottal aspect of the vowel (the constriction at the level of the epilaryngeal tube) manifests itself at the end of the vowel and is controlled to produce the acoustic signature which characterizes glottal vowels in Nasa Yuwe. The oral and nasal flow tracings allow the inference of the speech phases, the opening and closing of the velum and vocal tract as well as the opening and closing phases of the glottis on the EGG signal and the vibration of the vocal folds on the fundamental frequency trace.

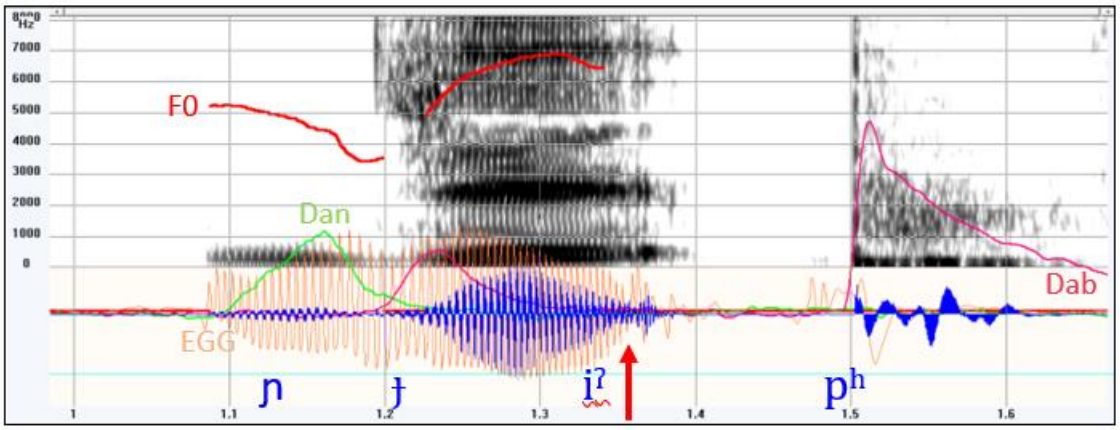


Figure 15. Wideband spectrogram, fundamental frequency (red), nasal airflow (Dan) plot in green, oral airflow (Dab) in pink, EGG signal in orange and audio waveform in blue during production of the word /ɲɟiʔpʰ/ 'face' in Nasa Yuwe. Only the scale of the spectrogram is given.

7. Conclusion

Understanding the synchronization of various parameters involved in speech provides a better understanding of how the sounds of languages are produced and which acoustic features stimulate the auditory system. Speech is a complex system that manifests itself in multiple dimensions that are far from being all known. The contribution of experimental data from African and Amerindian languages and their integration in models of speech production, biological and physical principles shows phenomena that still need a better understanding.

Data from African languages permit discussing some crucial issues in phonetics and phonology. Amharic offers a good example to study the concept of articulatory control (Kingston & Dhiel 1994). This language shows that not only articulatory but also auditory aspects of speech have to be taken into consideration to understand the issue. Indeed, Amharic shows that changes in aerodynamic parameters, as changes in Po, allow to infer the speed of larynx elevation in ejectives. Hadza data indicate that the characterization of clicks necessitates the acoustic and perceptual dimensions to make an adequate description of their features. Clicks also illustrate details in the fundamental sensitivities of the hearing system in languages (Pickles 1977).

Amerindian languages showed some other dimensions of speech. The production of the Karitiana vowel [o] is a good example of the possibility to produce the same acoustic target with different articulatory settings. This language also illustrates that the integration of production and perception dimensions allow the validation of universal tendencies in less frequent vowel systems, such as those lacking the vowel [u] (Lindblom 1986, Potard et al. 2008). Nasa Yuwe provides evidence that reinforce the validity of Esling et al. (2019) LAM model. The language also offers a superb example in the complexification of vowel systems where 32 phonological vowels are made possible by the addition of several phonation types to the 4 basic vowels qualities.

Finally, another aspect in the complexification of linguistic structures is given by the possible reuse of the physiological dimensions of swallowing to produce clicks. If true, this possibility leads to a new evaluation of clicks since in this case they could

reflect a complexification of speech sounds and they would then not be the remnants of ancient speech features following the hypothesis made by Sands & Güldeman (2009). Labio-dorsal consonants, on the other end, illustrate an ingressive airstream mechanism that is therefore much more common than usually assumed in African languages.

African and Amerindian languages examined in this paper show that we need more details and more empirical data if we want to understand *how* human speech works and we want to integrate data from as many languages as possible in a coherent theoretical framework.

Notes

This paper is the extension of a presentation made at a meeting of the Section Sciences humaines of the KAOW-ARSOM held on 2nd of February 2021. This contribution sums up and discusses more quantitative findings presented in several publications that are mentioned in the text.

[1] Figure 1 presents results of experiments made by using aerodynamic (oral airflow and intraoral (Po) pressures) and acoustic parameters on Amharic ejectives. These data allow making inferences on the dynamics of the articulatory movements that produce ejectives consonants. Aerodynamic recordings were made using the *Physiologia* workstation (Teston and Galindo 1990). Oral airflow measurements were taken with a small flexible silicon mask placed against the mouth. Intraoral pressure (Po) was recorded with a small flexible plastic tube (ID 2 mm) inserted through the nasal cavity into the oro-pharynx. Acoustic recordings were made via a High Fidelity microphone set on the hardware equipment connecting the transducers to the computer. The audio signal was digitized at 16,000 Hz and the physiological data at 2,000 Hz. Aerodynamic data were processed with the *Phonedit* software.

[2] Acoustic data were processed with the *Signal Explorer* software.

[3] These static MRI images were obtained by a method described in Demolin & Metens (2009).

[4] Words including all Hadza clicks were recorded with an *AKG head microphone*. Data were processed with the *Signal Explorer* and *Winpitchpro* software.

[5] MRI These real time MRI images were obtained by a method described in Isaeva et al. (2021).

[6] Words including all Karitiana vowels were recorded with an *AKG head microphone*. Data were processed with the *winpitchpro* software to measure the vowel formants and *Visible vowel* software to make the presentation of the results.

[7] Video data using simultaneous front and profile images were recorded at a speed of 25 fps.

[8] The simulations were made using the *JVTCal* articulatory/acoustic model provided by Shinji Maeda and Jacqueline Vaissière.

[9] The simulations were made using the *JVTCal* articulatory/acoustic model provided by Shinji Maeda and Jacqueline Vaissière.

[10] In order to establish the glottal, ventricular and/or epilaryngeal settings fiberscopic data were recorded with 1 Nasa Yuwe male speaker. Larygoscopy was made with a Xion video-stroboscopic & flexible microphone. This view shows the laryngeal settings. Images were processed with a *Matlab script* made available by Angélique Amelot.

[11] In order to establish the glottal, ventricular and/or epilaryngeal settings fiberscopic data were recorded with 1 Nasa Yuwe male speaker. Larygoscopy was made with a Xion video-stroboscopic & flexible microphone. Images were processed with a *Matlab script* made available by Angélique Amelot.

[12] Data were recorded with the *EVA2* workstation that can simultaneously record sound, oral airflow (Oaf), nasal airflow (Naf), intra oral pressure (Po) and electroglottography (EGG). The experimental protocole is similar to what is presented for the first note. Data were processed with the *Winpitchpro* software.

Acknowledgements

I would like to thank John Kingston, Tulio Rojas, Esteban Diaz, Philippe Martin, Lise Crevier Buchman, Jacqueline Vaissière, Shinji Maeda, Luciana Storto, Andrew Harvey, Richard Griscom, Alain Ghio, Angélique Amelot, Yves Laprie, Pierre André Vuissoz, Clara Ponchard, Moges Yigezu and Bonny Sands, for discussions that helped to improve the manuscript.

Parts of this research have been financed by the Labex EFL, Axe 1 *Phonetic and phonological complexity* and by the ANR project *Full3DTalkinghead*.

References

Benguerel, P.A. 1970. Some Physiological Aspects of Stress in French. *Natural language studies*, **4**. University of Michigan.

Bouhuys, A. 1977. *The Physiology of Breathing*. New-York. Grune and Statton.

Crothers, J. 1978. Typology and Universals of Vowels systems. – In J.H. Greenberg (ed.) *Universals of Human language*. Stanford, Stanford University Press, pp 93-152.

Demolin, D. 2021. *Speech embodiment and non-pulmonic consonants*. – International Seminar of Speech Production. New Haven. Issp2020.yale.edu/S07

Demolin, D., Griscom, R., Andrew Harvey, A. & Ghio, A. 2021. Acoustic features of Hadza clicks. – *The Journal of the Acoustical Society of America.* **150**, A68.

Demolin, D. & Metens, T. 2009. L'imagerie par résonance magnétique en temps réel pour l'étude de la parole. In A. Marchal & C. Cavé (eds.). – *Techniques d'imagerie médicale pour l'étude de la parole*. Paris, Hermès, pp 257-271.

Demolin, D., Amelot, A. Crevier-Buchman, L., Diaz, E. & Rojas Curieux, T. 2016. The vowel system of Nasa Yuwe. – *The Journal of the Acoustical Society of America*. **140**, (4): 3106.

Demolin, D. & Kingston, J. 2021. *The evolution and history of non-pulmonic consonants*. – Workshop honoring John Ohala. Laboratoire de Phonétique et Phonologie. Université Sorbonne nouvelle. www.ilpga.univ-paris3.fr

Diaz, E. 2019. *El habla nasa (páez) de Munchique: nuevos acercamientos a su sociolingüistica, fonología y morfosintaxis*. – Thèse de doctorat Université Lumière Lyon 2.

Esling, J. 2005. There are no back vowels. The Laryngeal Articulator Model. – *Canadian Journal of Linguistics* **50**: 13-44.

Esling, J., Moisik, S.R., Benner, A. & Crevier-Buchman, L. 2019. *Voice Quality. The Laryngeal Articulator Model*. – Cambridge. Cambridge University Press.

Fant, G., Kruckenberg, A. & Liljencrants, J. 2000. Acoustic-phonetic Analysis of Prominence in Swedish. In: Botinis, A. (eds) *Intonation*. Text, Speech and Language Technology, **15**. Springer, Dordrecht.

Gerber, S.E. 1974. *Introductory Hearing Science*. London. W.B. Saunders company.

Gick, B. 2019. *Emergence in Embodied Speech: Sound Change, Ontogeny and Phylogeny.* – Labex EFL invited seminar. Institut de Linguistique et Phonétique Générales et Appliquées. Sorbonne 3, Paris, France.

Gick, B. & Stavness, I. 2013. Modularizing speech. – *Frontiers in Psychology*. **4**: 1-2.

Gick, B., Anderson, P., Chen, H., Chiu, C., Kwon, H.B., Stavness, I., Tsou, L. & Fels, S. 2014. Speech function of the oropharyngeal isthmus: a modelling study, – *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. **2** (4): 217-222.

Hombert, J.-M. & Ohala, J. 1978. A model of tone systems, – In D. Napoli (ed.) *Elements of Tone, Stress and Intonation*. Georgetown Univeristy Press.

Isaieva, K., Laprie, Y., Leclère, J., Douros, I., Felblinger, J. & Vuissoz P-A. 2021. Multimodal dataset of real-time 2D and static 3D MRI of healthy French speakers, Scientific data, **8** (1). www.nature.com/scientificdata

Kingston, J. & and Diehl, R.L. 1994. Phonetic Knowledge. *Language.* **70** (3): 419-454.

Knight, A. Underhill, P.A., Mortensen, H.M., Zhivotovsky, L.A., Lin, A.A., Henn, B.M., Louis, D. Ruhlen, M. and Moutain, J.L. 2003. African Y chromosome and mDNA divergence provides insight into the history of click languages, –*Current Biology* **13**: 464-473.

Kusimba, S.B. 2003. *African foragers: environment, technology, interactions. Landham*, – AltaMira Press.

Ladefoged, P. 1968. *A phonetic study of West African languages*, – Cambridge. Cambridge University Press.

Liljencrants, J. & Lindblom, B. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. – *Language* **48**: 831-852.

Lindblom, B. 1986. Universals in vowel system. – In J. Ohala and J. Jaeger (eds.) *Experimental Phonology*. Orlando, Academic Press, pp 13-44.

Lindblom, B. 1990. On the notion of possible speech sound. – *Journal of Phonetics*, **18**: 135-152.

Maddieson, I. 1994. *Patterns of Sound*. – Cambridge, Cambridge University Press.

Mayer, C., Roewer-Despres, F., Stavness, I. and Gick, B. 2017. Do innate stereotypies serve as a basis for swallowing and learned speech movements? – *Behavioral and Brain Sciences* **40**.

McDonough, Joyce. 2003. *The Navajo sound system*. – Dordrecht, Kluwer Academic Publishers.

Ménard, L. 2002. *Production et perception des voyelles au cours de la croissance du conduit vocal : variabilité, invariance et normalisation*, – Thèse de doctorat Université Stendhal Grenoble.

Miller, A., Brugman, J., Sands, B., Namaseb, L., Exeter, M., & Collins, C. 2009. Differences in airstream and posterior place of articulation among N|uu clicks*. Journal of the International Phonetic Association*, 39 (2), 129-161.

Moisik, S. & Dediu, D. 2020. The ArtiVarK click study: documenting click production and substitution strategies by leaners in a large phonetic training and vocal tract imaging study. – In B. Sands (ed.), *Click consonants.* Leiden, Brill, pp 384-417.

Nakagawa, H. 2010. *Aspects of the phonetic and phonological structure of the G|ui language*. – PhD thesis. University of the Witwatersrand.

Pickles, J.O. 1977. *An Introduction to the Physiology of Hearing*. London. Academic Press.

Potard B., Laprie, Y., and Ouni, S. 2008. Incorporation of phonetic constraints in acoustic-to articulatory inversion. *The Journal of the Acoustical Society of America*. **123** (4). 2310–2323.

Rojas, T. 1998. *La lengua Paez*. – Ministerio de Cultura. Bogota.

Sands, B. & Güldemann, T. 2009. What clicks can and can't tell us about language origins. – In R. Botha and C. Knight (eds.) *The Cradle of language*. Oxford. Oxford University Press, pp 204-218.

Sands, B., Maddieson, I. and Ladefoged, P. 1996. The phonetics structures of Hadza. – *Studies in African Linguistics* **25** (2): 171-204.

Schwartz, J.-L., Boë, L.-J., Vallée, N. and Abry, C. 1997. The dispersion-focalization theory of vowel systems. – *Journal of Phonetics*, **25**: 255-286.

Signorello, R. Hassid, S. & Demolin, D. 2018. Toward an aerodynamic model of fricative consonants. – *The Journal of the Acoustical Society of America Express Letters*, **143** (5), EL 386-392.

Stavness, I., Nazari, M. A., Perrier, P., Demolin, D. and Payan, Y. 2013. Effects of Oribicularis Oris and Jaw Position on Lip Shape: A Biomechanical Modeling Study of the Effects of the Orbicularis Oris Muscle and Jaw Posture on Lip Shape. – *Journal of Speech Language and Hearing Research*, **56** (3): 878-890.

Stevens, K. 1972. The quantal nature of speech: evidence from articulatory-acoustic data. – In E.E. Davis Jr. and P.B. Denes (eds.) *Human communication a unified view*. New_York. McGraw Hill, pp 51-66.

Stevens, K. 1989. On the quantal nature of speech. – *Journal of Phonetics,* **17**: 3-45.

Storto, L. 1999. *Aspects of Karitiana Grammar*. – PhD thesis, MIT.

Storto, L. and Demolin, D. 2002. The phonetics and phonology of unreleased stops in Karitiana. – *Proceedings of the Twenty-Eighth Annual Meeting of the Berkeley Linguistics Society*, pp 487-497.

Storto, L. and Demolin, D. 2012. The phonetics and phonology of South American indigenous languages. – In: L. Campbell & V. Grondona (eds.) *The Indigenous*

languages of South America: A comprehensive guide. Berlin. De Gruyter Mouton. pp 331-390.

Traill, T. 1985. *Phonetic and phonological studies in !xóõ Bushman*. – Hamburg. Helmut Buske Verlag.

Traill, T. 1997. Linguistic phonetic features for clicks: articulatory, acoustic and perceptual evidence. – In R.K. Herbert (ed.) *African Linguistics at the Crossroads: Papers from Kwaluseni*. Köln. Rüdiger Köppe, pp 99-117.

Vallée, N. 1994. *Systèmes vocaliques de la typologie aux prédictions*. – Thèse de doctorat Université Stendhal Grenoble.

Vander, A. Widmaier, E.P., Raff, H., Strang, K.T. and Pradel, J.-L. 2019. – *Physiologie humaine.* Paris, Maloine.