Lingue e Linguaggi
Lingue Linguaggi 59 (2023), 127-150
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v59p127
http://siba-ese.unisalento.it, © 2023 Università del Salento


# HYPOTHESIS OF SOUNDS SPREADING FROM WHALES TO ANCESTRAL HOMININS

AMEDEO DE DOMINICIS[1],
WITH AN APPENDIX BY ALBERTO PETRI[2]
[1]UNIVERSITY OF TUSCIA, [2]ISC (INSTITUTE FOR COMPLEX SYSTEMS), CNR

**Abstract** - This paper explores the acoustic characteristics of the human speech signal (used for communication purposes), proposing that they can derive from an adaptive evolution of the cetaceans' echolocation signals. Nevertheless, the modern human speech signal is far more complex than that of animal echolocation. Indeed, this evolution began before *Homo sapiens*, probably at the time of the *H. erectus*. The comparison between the whale and *Homo sapiens* can allow us to describe the acoustic features of human speech as the result of a co-evolution of the system of acoustic localization of objects in the common space inherited from mammals and specialized only by humans in order to allow them more sophisticated uses of their sensory apparatus. Although it is impossible to adduce material proofs, there is inferential evidence arising from comparing archaeological, paleontological, biological, acoustic, and linguistic data.

**Keywords**: phonetics; biolinguistics; formants; ultrasounds; biosonar.

## 1. Introduction

By the term 'speech' here we mean the acoustic-phonetic characteristics of the human voice if these characteristics are used for a linguistic function, i.e., if they are a vehicle of oral "language". On a spectrogram, human speech appears as an acoustic signal comprising a sequence of formants and formant transitions: not a succession of simple harmonics, not merely noises (grunts, hoo sounds, barks, screams, panted grunts, etc.), but a complex arrangement of formants, transitions, and noise bursts. How did humans come to use formants and formant transitions to implement the acoustic substrate that we find today in languages worldwide?

Ultrasonic echolocation includes both the emission and reception of signals. Here we will refer to emission, although it is clear that the two components coevolve phylogenetically.

According to the current literature in neurosciences, phylogenetically the main function of hearing and of the auditory system in animal species is localizing the source of the sounds. Ultrasonic echolocation is then the most refined system to achieve this goal. However, it is not universal. Only part of the animal world has this ability. The *Homo* species does not.[1] Nevertheless, there are remarkable acoustic similarities between animal echolocation signals and human vocal signals. This paper discusses these similarities and tries to describe a possible interface between them.

The paper is organized as follows.

In section 2 we introduce the point of view of neuroscience on the original function of the hearing system in animal species.

[1] Although blind humans are known to be able to use sounds to echolocate, they do not use ultrasounds.

In section 3 the paper analyzes the auditory mechanism of echolocation, in particular for cetaceans. In addition, we focalize on the acoustic features of steady and modulated harmonics through which whales locate dangers and preys.

In section 4 the paper describes the acoustics of human speech and compares it to the animal ultrasonic echolocation.

In section 5 we discuss some evidence in order to corroborate a possible interface between animal and human hearing, which arise from the comparison of archaeological, paleontological, biological, acoustic, and linguistic data. This interface is then described in section 6.

In section 7 we present two main resources to explain a possible interface between animal and human hearing: Vocal Mimicry, and Vocal Production Learning.

In section 8 we depict the further evolution of human speech features, enabled by the unique anatomy of man (compared with other primates); these features form the background to the enormous variety of human languages and speeches.

## 2. Main function of hearing in animal phylogenesis: location of targets in space

Examining the main function of hearing in animal species, according to the well-established literature, the first goal of the auditory system is localizing the source of the sounds, as evidenced by the most comprehensive reference work in neurosciences (Kandel *et al.* 2013, pp. 682-684).[2]

In animal species the hearing system has the primary function of generating information about the presence of a prey, or of a predator, or of a mate for reproduction (i.e., the main functions of the entire living world). As a consequence, the connection between the auditory and the animal neural apparatus was born and evolved to primarily fulfill these survival functions. Consequently, it is particularly sensitive to those acoustic characteristics that are necessary to locating the source of sound in space and to defining its identity (shape, size, etc.). This sensory and neural predisposition phylogenetically precedes its use for linguistic purposes.

As for the human linguistic function and purposes, there is a co-evolutionary relationship (Deacon 1997) between hearing and speaking (as well as between articulated speech and the respiratory system). In addition to a vocal tract that is anatomically capable of producing a large variety of formant patterns, human speech requires sophisticated nervous control. The most obvious aspect of this feature is the possibility that speech requires enhanced motor control over the vocal articulators (tongue, lips, velum, jaw, etc.). Thus, a co-evolutionary framework can afford the task of accounting for a continuity

[2] In particular, Kandel *et al.* (2013) wrote: "In most animals hearing is crucial for localizing and identifying sounds; for some species, hearing additionally guides the learning of vocal behavior" (Kandel *et al.* 2013: 682). And they added: "The auditory system differs from most other sensory systems in that the location of stimuli in space is not conveyed by the spatial arrangement of the afferent pathways. Instead, the localization and identification of sounds is constructed from patterns of frequencies mapped at the two ears as well as from their relative intensity and timing. The auditory system is also notable for its temporal sensitivity; time differences as small as 10 μs can be detected. Auditory pathways resemble other sensory systems, however, in that different features of acoustic information are processed in discrete circuits that eventually converge to form complex representations of sound" (Kandel *et al.* 2013, pp. 682-683).

between animal echolocation and human speech signals, connecting both hearing and speaking.

We believe that determining how vocal signals evolved is likely to become clearer with concerted efforts in testing comparative data from many non-human animal species. There is some genetic evidence for an evolutionary continuity between ultrasonic echolocation animal calls and human vocal signals (Davies *et al.* 2012)[3]: the authors examine three 'hearing genes', namely the proteins *Prestin*, *Tmc1* and *Pjvk*. Genetic analysis shows that "All three genes encode proteins that are expressed in the cochlea and implicated in mammalian hearing, and all have mutant forms that have been linked to non-syndromic hearing loss in humans and/or mice" (Davies *et al.* 2012, p. 487).[4]

## 3. Acoustics features of ultrasonic echolocation

Echolocation has been studied in terrestrial and aquatic mammals, for instance in bats and in cetaceans. As for bats, although the pioneering study by Suga *et al.* (1983) is clearly dated, it deals with experiments on bats that would not be possible today because they would not be allowed under current ethical standards. Many animals emit sounds above 17-20 kHz, which represent the limiting frequency perceivable by human hearing. The phenomenon was discovered in 1938, when Donald Griffin was able to pick up ultrasound emissions from bats of the *Vespertilionidae* family (Griffin 1959). Griffin's discovery stimulated much research in this field; actually, echolocation signals have been detected in many other bat families. Generally, the ultrasonic signals useful for echolocation are emitted from the bat's nostrils and the echo is received through the animal's auricles. Bats also emit short audible *clicks* from their mouths. These communication systems are more widespread in the animal world than is believed (Sales, Pye 1974). In the bat brain, differentiated hearing territories have nerve cells that respond selectively to one, two, or all of these stimuli, and the acoustic system therefore acts as a feature detector (Suga *et al.* 1983).

An echolocation system similar to that of bats has also been demonstrated for cetaceans (whales, dolphins, porpoises). Cetaceans' ultrasonic echolocation dates back to the Oligocene era, i.e. from 33 to 23 million years ago, long before the separation between humankind and African apes (about 5 million years ago) and before hominins – *Homo habilis* and *Homo erectus* – appear on Earth (about 2 million to 250,000 years ago, in the Pleistocene era). Likewise, *Homo neanderthalensis* appears about 400,000 to 130,000 years ago. Also, according to the fundamental review by Fordyce and de Muizon (Fordyce, de Muizon 2001), echolocation and filter-feeding in cetaceans occurred both in

[3] In particular, Davies *et al.* (2012) wrote: "The 'hearing gene' *Prestin* was recently shown to have undergone unprecedented levels of sequence convergence between lineages of echolocating mammals (Li *et al.* 2008, 2010; Liu *et al.* 2010)" (Davies *et al.* 2012, p. 480). And they added: "The molecular basis of mammalian hearing involves over 50 candidate genes identified via studies of mutagenesis and non-syndromic hearing loss (Accetturo *et al.* 2010; Dror, Avraham 2010). The mammalian hearing apparatus has evolved into a wide range of auditory systems, the most specialised of which arguably occur in echolocating bats and cetaceans (Vater, Kossl 2004)" (Davies *et al.* 2012, p. 480).

[4] "Given their ability to echolocate, it is perhaps not surprising that bats have long served as important models for understanding the neurophysiology of auditory processing (see, for example, Kossl *et al.* 2003). The data of sequence convergence between taxa with ultrasonic hearing in three separate hair cell genes suggest that echolocating mammals might be equally useful for unravelling the molecular basis of hearing" (Davies *et al.* 2012, p. 487).

Odontocetes and Mysticetes around the Eocene/Oligocene boundary, i.e. over 28 million years ago. This dating is demonstrated by anatomical characteristics found in fossil data.[5]
The source of ultrasounds is the upper part of the cetacean nasal passages; thus, cetacean melon acts as a lens or transmission pathway to project sound into the environment. Although receiving echo signals occurs via their special earbones,[6] [7] many other anatomical mechanisms play a role in the perception of echo signals, such as the lower jaw, air cavities adjacent to ears, special 'acoustic lipids' and air sinuses. Briefly, the echo is transmitted through a fat pad located in the lower jaw to a tympanic bulla.
These anatomical characteristics are testified by fossil remains[8] and were probably induced by a response to changes in early Oligocene food resources.[9] Thus, cetaceans' ability to manage ultrasonic echolocation is far older than the appearance of humans.

[5] "All odontocetes, fossil and recent, have a distinctive and unique pattern of skull bones, in which the maxilla (the main tooth-bearing upper jaw bone) extends back over the frontal bone usually beyond the orbit (Miller 1923). The two maxillae and nearby bones form a voluminous face which carries muscles associated with the soft tissues of the nose (Cranford *et al.* 1996; Mead 1975). In living species, this complex of structures is implicated strongly in echolocation. Because this behaviour is associated with a distinct skull form also seen in fossils, echolocation is also inferred for all fossil odontocetes. Fordyce (1980) suggested that the evolution of echolocation in Oligocene times was a key factor in the origin of odontocetes" (Fordyce, de Muizon 2001, p. 195).

[6] "Many field and laboratory observations identify sound as critically important in cetacean communication, navigation, and prey detection underwater. Hearing is linked with the issue of sound generation, below, which includes extremes of high frequencies in odontocetes and low frequencies in mysticetes (Wood, Evans 1980; Heyning 1989; Heyning, Mead 1990; Ketten 1991, 1992; Hemila *et al.* 1999; Nummela 1999a, 1999b). The auditory complex is modified from that of land mammals, which evolved to function in air (Luo, Gingerich 1999). Earbones are amongst the most distinctive elements of the Cetacea. Of these, the large dense tympanic bulla is a rather simple bone which contains an internaI cavity filled with an expanded eustachian tube. The more complex periotic (petrosal), which includes the organs of hearing (cochlea) and balance, has prominent anterior and posterior (mastoid) processes and a complex range of muscle attachments and nerve and vascular foramina. These distinct earbones can be traced back into the early (pakicetid/protocetid) beginnings of Cetacea (Pompeckj 1922; Kellogg 1936; Luo, Gingerich 1999)" (Fordyce, de Muizon 2001, p. 217).

[7] "Ongoing research addresses the evolution of echolocation. Amongst living cetaceans, only odontocetes are known to echolocate (Wood, Evans 1980; Ketten 1991, 1992; Cranford *et al.* 1996). This pattern implies that the most recent common ancestor also echolocated, a notion supported by some observations on fossils. Echolocation requires a sound-making system (Norris 1968) as weIl as a highly evolved ear. Modern studies of odontocetes point to the upper part of the nasal passages, between the blowhole and the skull, as the likely source of high-frequency sounds (Mead 1975; Heyning 1989; Heyning, Mead 1990), though some authors consider the larynx to play a role in sound generation (Purves, Pilleri 1983; Reidenberg, Laitman 1988). Odontocetes probably make high frequency echolocation sounds by moving recycled air in a network of asymmetrical sacs – paranasal sinuses – and valves of the nasal passages. The large fatty melon, seen in aII living odontocetes and inferred for ancestral odontocetes, may act as a lens or transmission pathway to project sound into the environment (Norris 1968; Cranford 1999; Cranford *et al.* 1996)" (Fordyce, de Muizon 2001, p. 218).

[8] "The fossil record of odontocetes shows that unique features implicated in echolocation (listed above) evolved early and only once in odontocete history, by the Late Oligocene. Early squalodontids, for example, have a symmetrical but depressed facial region which held an enlarged complex of facial muscles and, presumably, nasal diverticula and melon. Periotics of Late Oligocene squalodontids show adaptations for receiving high frequency sound (Fleischer 1976; Luo, Eastman 1995). Asymmetrical skulls evolved by the Late Oligocene (e.g., in *Waipatia*; Fordyce 1994). The later Miocene radiation of the delphinoid lineage marked the evolution of much more elaborate pterygoid sinuses and presumably better-isolated earbones than seen in older fossil groups" (Fordyce, de Muizon 2001, pp. 219-220).

[9] "Fordyce (1980, 1992) suggested that, as for mysticete, there was a link between feeding behaviour and the origins of the group: echolocation evolved to help hunt single prey in about the Early Oligocene, in response to changing food resources (especially below the photic zone), changing oceans, and continental rearrangement" (Fordyce, de Muizon 2001, p. 221).

The ultrasonic calls of cetaceans consist of a series of harmonics with a stable or constant frequency over time (CF), and a series of harmonics with a modulated frequency (FM): rising and/or falling. Figure 1 shows the spectrogram of these calls.

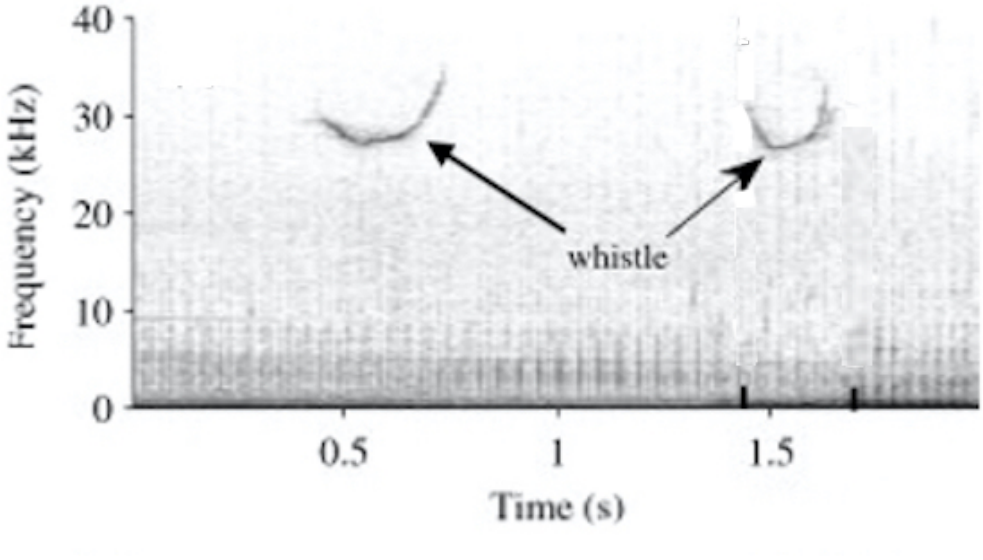


Figure 1
Example spectrograms of ultrasonic whistles from Icelandic killer whales (*Orcinus orca*). The figure is adapted from Samarra *et al.* (2010).

The FM (falling) component is not only a prerogative of cetaceans, but also of some terrestrial mammals that use echolocation, such as the bats studied by Suga (Suga *et al.* 1983). The FM (falling) sounds are used to determine the distance to the target and to characterize it. The animal measures the interval between the emitted sound (pulse or call) and the returning echo, which corresponds to a particular distance, based on the relatively constant speed of sound.

The CF components of the animal's calls are used to determine the relative speed of the target with respect to the animal, and the acoustic image of the target. When an echolocating cetacean is swimming toward an object, the sounds reflected from the target (echo) are Doppler-shifted to a higher frequency at the animal's ear, for the animal is moving toward the returning sound waves from the target, causing a frequency increase of these waves at its ear. Similarly, a receding target yields reflections of lowered frequency at the animal's ear.

More specifically, the whale's problems are three-fold: determining the source of the sound, understanding if that source is moving or not, and, if it moves, estimating its direction and speed. The animal produces a sound with CF and calculates the sonar echo in order to measure the time delay of the echo and its variation in frequency (compared to the original pulse). The delay provides the distance between the animal and the source, while if the echo has a frequency higher than the signal emitted, this means that the distance between the animal and the target is decreasing (Doppler effect). If, on the other hand, the echo has a frequency lower than the call signal, then the distance between the whale and the target is increasing. At this point the animal only knows if the target is stationary or if it moves; but in this second case – that is, if the target moves – the animal

does not yet know how fast and in which direction. The FM (falling) component of the acoustic signal is functional for this purpose. It too produces a sonar echo. In the case of a stationary target, the sonar echo is an FM signal that falls in the frequency scale, and this provides the whale with a second, redundant confirmation that the target is stationary, a redundancy useful for avoiding false alarms. To find out if the target is moving and in which direction, the animal sends further consecutive signals and receives a series of consecutive echoes: if the delay of the CF varies in the series (increases or decreases), then the measure of these delays will correspond to as many points in space and, joining them, the direction of movement of the target is reconstructed, while the speed of the target will be calculated on the basis of the differences in the delay times of the different consecutive echoes. The FM component, with frequency drop, is used to explore the environment at greater distances, as with the same energy used, the lower frequencies propagate further away. Therefore, the use of the decreasing FM component allows the whale to reconstruct the shape of the target, to obtain more information about the target, and to identify it.

However, unlike bats and terrestrial mammals, marine mammals use not only falling FM but rising ones as well (see Payne, McVay 1971; Tyson *et al.* 2007). In order to explain this specificity of rising FM harmonics in marine mammals, two hypotheses can be formulated.

1. Cetaceans produce sounds while swimming, that is, while moving in a medium denser than air, and sounds move faster. Consequently, the Doppler effect is enhanced (Rosen, Gothard 2009) and differs in front of the sound source and behind it. In front of the sound source, the acoustic waves decrease by compression, and this produces an increase in frequency. Behind the animal, the process is the opposite, with an elongation of sound waves and a decrease in frequency. A conclusion that cetaceans must modify their relative swimming speed and sound frequency to minimize the Doppler effect can be hypothesized. This alternation of falling and rising FM triggers a mismatch in predators' ultrasound receivers, and allows cetaceans to avoid detection by them, especially by killer whales (*Orcinus orca*). Thus, according to this hypothesis, rising call frequencies is a strategy to avoid predation, because this gives the predator false information about its prey.

2. Marine mammals use rising FM calls to achieve highly directional sonar beams while hunting in open water. All else being equal, more directional (i.e., long, narrow) beams focus more energy along the acoustic axis, increasing sonar range while minimizing potentially distracting off-axis echoes. According to this hypothesis, rising calls are hunting tools.

## 4. Acoustics features of human vocal signal (*Homo sapiens*)

Acoustically, human speech consists of a series of periodic and aperiodic signals. The spectrogram of periodic signals is characterized by 'packets' or clusters of harmonics particularly amplified (by the vocal tract) in energy and called formants; formants are relatively stable over articulation time and identify the place of articulation of vowels and sonorant consonants.

In particular, spectral analyses of human speech sounds exhibit three basic acoustic patterns or components: CF components (formants), FM components (transitions), and noise bursts. Vowels and sonorant consonants are identified mainly by the first and second formants (F1 and F2), although the third formant (F3) has some influence on recognition. Plosives, some of the fricatives, and affricates are identified by the burst or by the noise,

and by a combination of the formant transitions belonging to the adjacent vowels.

In a spectrogram of aperiodic signals – if adjacent to a periodic signal – its place of articulation is identified by formant transitions.

In short, the human vocal signal shows a specific acoustic pattern consisting in a succession of *Steady Formants* and *Formant Transitions*. Similarly to the CF and FM harmonics of cetaceans' calls, Steady Formants are constant in frequency, and Formant Transitions are modulated in frequency: they can be falling or rising. In addition to a difference between human vocal signals and the ultrasonic signals of cetaceans, here we also find a remarkable similarity, which consists in the fact that both signals are composed of a recurring sequence of stable and modulated spectral components. The only difference is that whales use harmonics, while humans use formants. The causes of this difference, which have been well described in the pioneering work of Gunnar Fant (Fant 1960, 1966), depend on the anatomical and articulatory characteristics of the human resonator tube (also called filter) or human vocal tract. Unlike the human vocal tract, the ultrasonic resonator of cetaceans is the so-called 'melon' – that is a mass of adipose tissue located in the forehead of Odontocetes – comparable to the sound box of a musical instrument: it has no internal moving parts and therefore is simply intended to vibrate at the same harmonic frequencies generated by the source (i.e., a dense concave bone and an air sac located at the top of the head, near the blowhole).

A further reason for this differentiation is that animal ultrasounds are only useful as a measuring tool; hence, the greater their evolutionary utility, the greater the precision and accuracy of the measurement. Therefore, the ultrasonic calls must be pure sounds, with only one harmonic. Conversely, this metrological purpose is absent in human speech.

These two features (formants and transitions) in human speech are used to characterize vowels/sonorants and non-sonorant consonants respectively. The steady parts of the formants discriminate the timbre of the vowels and of the sonorants (periodic signals); the transitions of the aforementioned formants (in particular that of the second formant, also called T2) discriminate the formantic locus, that is, the acoustic index that identifies the place of articulation of the non-sonorant consonants (aperiodic signals) that are adjacent to the vowels. The reference acoustic model for these phenomena is the so-called locus theory by Delattre and colleagues (Liberman *et al.* 1952; Durand 1954; Delattre *et al.* 1955; Delattre 1970; Blumstein 1980). Figure 2 illustrates this point.

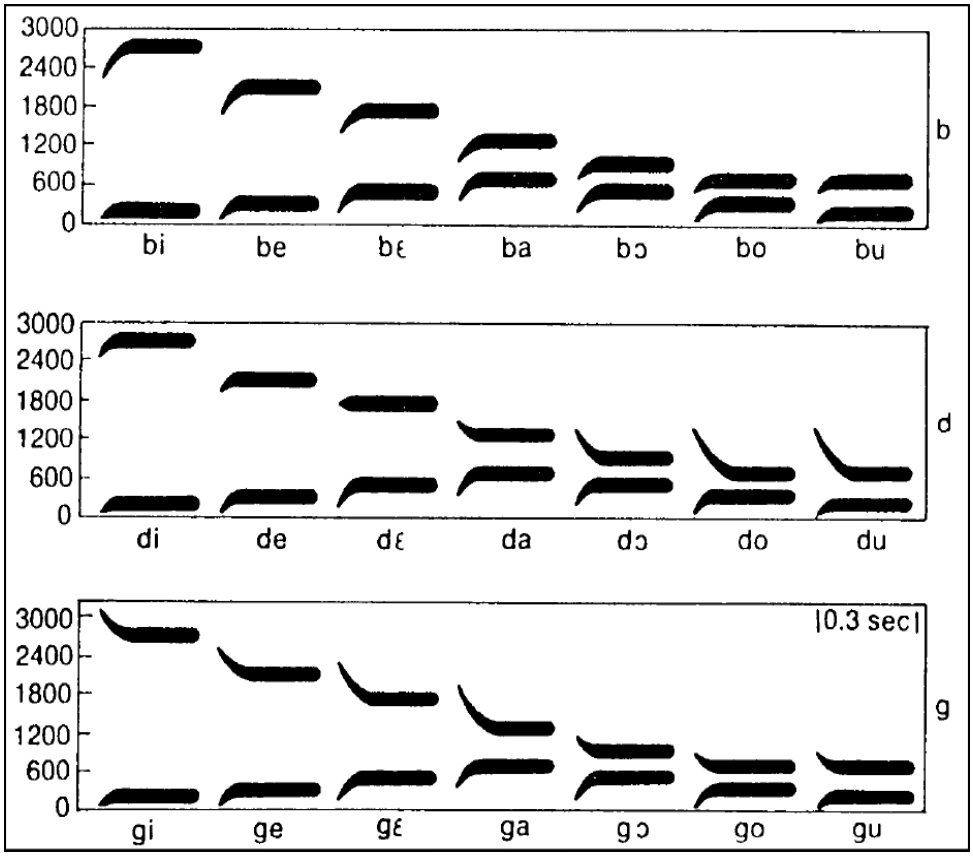


Figure 2
Schematic spectrograms of formants and formant transitions in CV syllables with voiced bilabial, dental and velar C (adapted from Cooper *et al.* 1967, p. 274). V = vowel; C = non-sonorant consonant.

This acoustic representation concerns *Homo sapiens*. In his vocal signal we find the acoustic features mentioned above. However, from paleontological and archaeological evidence, some studies had already found support for an older origin of a speech signal broken down into speech sounds, dating back to the stages of nonmodern hominins, before *Homo sapiens* (for instance, MacNeilage 2008; Milo *et al.* 1993).

Given the aforementioned spectral similarities between animal echolocation signals and human vocal signals, and given that animal echolocation ability is older than human's appearance on earth (see section 3), we plan to further explore the hypothesis of their possible causal relationship.

The question we try to answer is whether or not there is a possible continuity between animal ultrasound echolocation and human vocal signals. If this continuity is possible, and if it squares with archaeological, paleontological, biological, acoustic, and linguistic data, then it can be assumed that human speech stems from interspecific imitation and not from intraspecific evolution. Obviously, in order to obtain a valid proof, it is first necessary to verify that there is a possible "natural" context in which the ultrasonic frequencies of cetaceans have become "audible" to the human ear, and that at the time of this possible "encounter" between human beings and cetaceans, human beings had not already independently developed a vocal signal with formants and transitions (see section 6). Moreover, this demands and implies that at the time of the aforementioned "encounter" the cetacean echolocation system was fully developed, and therefore prior to the encounter (see section 3), and that – always at the time of the "encounter" – human beings were able both to imitate the sounds of non-conspecifics (Vocal Mimicry) and to

learn to reproduce them from human conspecifics (Vocal Production Learning) (see section 7).

## 5. Making ultrasounds "audible"

A "natural" mechanism of ultrasound down conversion to the human frequency range does not exist. In a "natural" context, a frequency conversion is possible only if the sound spreads through a nonlinear medium. Therefore, a possible context for this type is an ultrasound coming from the water and propagated in the air, after passing through a possible "natural" nonlinear transducer.

Thus, as for the natural context allowing humans to "hear" the ultrasonic calls by cetaceans, the answer may lie in the net of air bubbles produced by whales and cetaceans while hunting. Air bubbles can act as a possible mechanism of ultrasound down conversion since the bubbles are efficient nonlinear resonators. As regards the physical-acoustic details of this mechanism, we refer the reader to the Appendix.

Furthermore, we have pre-*sapiens* evidence that *Homo erectus* roamed the seas (see section 6) and hunted whales. It was a so-called 'opportunistic' (or passive) hunt, datable to the Oligocene, i.e. long before the appearance of *Homo sapiens*. In this context, *Homo erectus* could easily have come into auditory contact with whale signals filtered and down converted by the air bubbles.

As for human whale hunting, Savelle and Kishigami (2013, pp. 2-5) date the human hunting of whales, according to archaeological and petroglyph evidence, to around 10,000-6,000 B.P., that is long after *Homo neanderthalensis*, *H. erectus* and *H. habilis* had disappeared. Likewise, Seersholm *et al.* (2016) report fossil and DNA-based evidence of bowhead whale hunting by the Saqqaq Paleo-Inuit culture (Greenland) 4,000 years ago. The origins of active whaling have been tied to the development of toggling harpoons that appear about 4,000 BCE among North Pacific and Bering Sea peoples for hunting small sea mammals like seals in ice-infested waters (Yamaura 1980).

As opposed to active whale hunting, scavenging of stranded cetacean carcasses was common in pre-historic times and has been described across multiple sites in Europe (Clark 1947), North America (Monks *et al.* 2001) and Africa (Smith, Kinahan 1984). This relates to what can be termed 'opportunistic' whaling use.

Nevertheless, this does not necessarily imply that humans did not hunt whales before the *Homo sapiens* era. Still today in the Paamiut area in Southern Greenland, humpback whales are traditionally hunted using simple lances and toggling harpoons (Seersholm *et al.* 2016): "By approaching the docile animals noiselessly, the hunters could kill the whales by spearing them behind the flipper. Similarly, single kayak-equipped eighteenth-century Unangan (Aleut) hunters of the Bering region used barbed non-toggling harpoons coated with aconite poison to immobilize the whales by spearing them near the flipper (Fitzhugh, Chaussonnet 1994). After a few days the whale could no longer remain upright and would drown and be towed to shore. While it is unlikely that aconite poison was part of the Saqqaq hunting strategy, a similar effect might have been achieved from harpoons infested with rotten meat or blubber, as even small flesh wounds can cause inflammation and, within days, immobilization of the flipper or death of such large whales" (Seersholm *et al.* 2016, p. 7).

In any case, setting aside the possible human interaction with whales due to their hunting, some research hypothesizes the existence of a collaborative relationship between killer whales and *Neanderthals*, which therefore dates back to an era before *sapiens*.

According to marine biologist Manuel Esteve (2020), orcas and *Neanderthals* consciously collaborated together to hunt tuna in the Strait of Gibraltar. This research is based on the analysis of archaeological remains and on whale behavior: this collaborative relationship between orcas and *Neanderthals* was intentional and is confirmed by the fact that, in order to flee the killer whales, tuna would jump onto the beach where the hominins would capture them and throw unwanted fish parts back into the sea. In fact, according to the records of fossils found in the caves of *Neanderthals* in Gibraltar, it was found that these people lived on tuna, despite lacking the tools or boats to fish them. *Neanderthals* took advantage of this phenomenon of fleeing tuna to catch them, and, far from this behavior being fortuitous for both the hominins and the killer whale, the collaboration was intentional, because the whales benefitted from the fish remains that were cast back into the sea by the hominins.

Thus, *Homo neanderthalensis*, *Homo erectus* and *Homo habilis* – long before the era of *Homo sapiens* – could have 'opportunistically' hunted whales by approaching and poisoning them in the sea, and then scavenging the stranded cetacean carcasses on the seashore. During this slow approach to cetaceans in the sea, humans could listen to the echolocation calls of cetaceans possibly down converted by air bubbles.

## 6. Vocal development of hominins before *sapiens*

According to the so-called source-filter theory developed by Gunnar Fant (Fant 1960, 1966), in order to produce acoustic formants and formant-transitions, the vocal signal requires an advanced filter or resonator (or vocal tract) able to change its shape and dimensions, and this outcome depends primarily on changes in the shape and position of the articulators (tongue, lips, jaw, etc.).

*Homo erectus* (or *ergaster*) appeared in Africa between 1.7 and 1.8 million years ago, while *Homo neanderthalensis* appeared later, in Europe, approximately 127,000 years ago – although anthropologists do not always agree on how to classify early hominin species (Boyd, Silk 2000). *Homo erectus* and *Homo habilis* could not yet produce complex vocal signals, that is, signals characterized by acoustic formants and formant-transitions. But he was able to sail in the sea where cetaceans live.

According to archaeological and fossil evidence (Bednarik 1999; Bednarik 2003), Daniel Everett (2017) argues that *Homo erectus* was a hunter and a mariner (and a fisher): "archaeologist Robert Bednarik and others have provided extensive and convincing evidence that *Homo erectus* built watercraft and crossed the sea at various times in the lower Palaeolithic era, around 800,000 years ago (and three-quarters of a million years before *Homo sapiens* made sea crossing)" (Everett 2017, p. 59).[10]

[10] According to Daniel Everett (2017), *Homo erectus*, which appeared in Africa some 1.8 million years ago, already knew how to build boats for navigation, and even had a basic language for communicating at sea. Indeed, fossils of *Homo erectus* have been found in southern Europe, but also in China and Indonesia. For instance, archeological sites have been discovered in Socotra, Flores, Crete, and other islands (Everett 2017, p. 60). And some scholars hypothesize that *Homo floresiensis*, the short-sized hominin who lived on the Indonesian island of Flores up to 50,000 years ago, was descended from *Homo erectus* (van den Bergh *et al.* 2016).
Everett hypothesizes that the hominin's seaborne movements were intentional and coordinated, with at least 20 individuals per shipment. To understand each other and survive, these prehistoric sailors had to have some shared linguistic code.
Everett (2017, pp. 53-54) also observes that other species of *Homo* co-existed or existed in close

Everett also concludes that *Homo erectus* had some kind of speech, but that this was accompanied by gestures as aids to communication. Everett (2017, pp. 117-118) writes that *Homo erectus* would have been unable to make the same range of sounds as we can, not least because he lacked the version of a gene necessary for human control of the muscles used in speech (known as FOXP2) found in modern humans and in *Homo sapiens*, but – as we will explain below – because of the anatomy of his vocal tract, which did not allow him an accurate realization of formants and transitions. These limits also affect the anatomy of *Homo habilis*.

In particular, *homo erectus* was not able to carry his vocal signal over long distances, because of his "inability to form the same range of vowels that *sapiens* can produce" (Everett 2017, p. 116). Moreover, his "speech perhaps sounded more garbled relative to that of *sapiens*, making it harder to hear the differences between words", partly because he lacked a modern hyoid bone (Everett 2017, p. 117). Lastly, "*Erectus* faces were more distinguished by prognathism than modern humans', which would have impeded speech as we know it" (Everett 2017, p. 117).

The main differences between the *H. erectus* and *H. sapiens* vocal apparatus were in the hyoid bone and pre-*Homo* vestiges, such as air sacs in the center of the larynx. "The hyoid bone sits above the larynx and anchors it via tissue and muscle connections. By contracting and relaxing the muscles connecting the larynx to the hyoid bone, modern humans are able to raise and lower the larynx, altering the $F_o$ (fundamental frequency) and other aspects of speech. In the hyoid bones of *erectus,* on the other hand, though not in any fossil *Homo* more recent than *erectus*, there are no places of attachment to anchor the hyoid" (Everett 2017, p. 186).

Air sacs are relevant to human vocalization because their presence would render many sounds emitted less clear than they are in *sapiens*. The evidence that *erectus* had air sacs is based on fossils of *erectus* hyoid bones. Capasso *et al.* (2008) describe a hyoid bone body, without horns, attributed to *Homo erectus* from Castel di Guido (Rome, Italy), dated to about 400,000 years BP. The hyoid bone body shows the bar-shaped morphology characteristic of *Homo*, in contrast to the bulla-shaped body morphology of African apes and *Australopithecus.* Its measurements differ from those of the only known complete specimens from other extinct human species and early hominin (Kebara Neanderthal and *Australopithecus afarensis*) and from the mean values observed in modern humans, suggesting that the morphological basis for human speech didn't arise in *Homo erectus*.[11]

succession soon after and before *erectus* (*Homo habilis*, *ergaster*, *heidelbergensis*, *rudolfensis*, etc.), but "the story of human language evolution changes in no significant way, whether *erectus* and *ergaster* were the same or different species".

On the other hand, even denying the intentional and coordinated nature of these maritime movements of *Homo erectus*, and therefore attributing them e.g. to a tsunami, nevertheless those of *Homo neanderthalensis*, recently discovered in the Mediterranean sea, are intentional and targeted, since they concern several sites. In fact, Tristan Carter, Thomas Strasser and Curtis Runnels in 2008-2009 discovered stone tools along the shores of some Greek islands, dating back to 130,000 years ago. The finding indicates that *Neanderthals* too had the technological and cognitive means to navigate (Carter *et al.* 2019). So if not *Homo erectus*, at least the *Neanderthal* was able to navigate the sea. In any case, navigation by sea precedes *Sapiens*.

[11] "This shape seems to confirm that the earlier phases of human evolution, not associated with the capacity for speech, were characterized by a bulla-shaped hyoid body. On the basis of the few fossil hyoid bones available for examination, it seems reasonable to admit that the bar-shaped hyoid body is a characteristic of the genus *Homo*. In addition, the small anatomical differences between the *Homo erectus* of Castel di Guido and *Homo sapiens* hyoid bones consist primarily of a few impressions from the attachment of the major supra-hyoid muscles, whose activity modulates the high end of the vocal tract together with the sub-

These issues have been discussed in the literature on language origins. For instance, Deacon (1997), Tobias (1998), and Wynn (1998) have analyzed the language capabilities of both *Homo habilis* and *Homo erectus* with respect to brain structure and respiratory control. Lieberman (1992, 1993), referring to formants and transitions, has argued that any speech production capabilities in *Homo neanderthalensis* would have been severely limited by the physiology specific to that species.

Lieberman (1992, 1993, 2013; Lieberman, Crelin 1971) provides several pieces of evidence regarding the structure of the Neanderthal supralaryngeal tract, which would have prevented *Homo neanderthalensis* from producing the range of speech sounds that modern humans are capable of producing. He asserts that Neanderthals were incapable of producing the vowels [i], [a], and [u]: the Neanderthal tongue "is largely contained within the oral cavity" (Lieberman 1992, p. 410). This positioning of the tongue would have prevented Neanderthals from "accomplishing the abrupt changes in airway shape that are necessary for producing the vowels [i], [u], and [a] and from sealing off the nasal cavity from the rest of the supralaryngeal airway" (Lieberman 1992, p. 410). As a consequence, vowel production would have been limited, and only nasalized speech sounds would have been possible for Neanderthals. Thus, Lieberman argues that Neanderthals were unable to produce unnasalized speech sounds, that is, the sound that "enhances the perceptual recovery of the formant frequency patterns that make human speech a rapid means of communication" (Lieberman 1992, p. 409).

Previous studies have claimed that Neanderthals could produce human-like speech, based on the presence of a hyoid bone. Lieberman (1992, 1993) disagrees with these studies and notes that the supralaryngeal airway of modern humans is not defined by the simple presence of the hyoid bone. In fact, the hyoid bone of modern humans is similar in size to the hyoid bone of pigs. "The metrical similarity between pig and human hyoids, furthermore, indicates that hyoid bone morphology is not related to the position of the hyoid and the shape of the supralaryngeal vocal tract" (Lieberman 1993, p. 174). Thus, the presence of a hyoid bone in Neanderthal specimens does not suggest that Neanderthals had a supralaryngeal tract similar to that of modern humans.

Another evidence comes from the structure of the basicranium or base of the skull. Its anatomical shape and size affect the dimensions of the supralaryngeal tract. The size of the basicranial hump differs in all primates (Boyd, Silk 2000). A noticeably high basicranial hump in adult humans makes room for the modern human vocal tract, and a long vocal tract is crucial in the production of a wide range of formants and articulatory places (formant transitions). Modern humans are also distinguished from previous hominin species based on the presence of a flexed basicranium. Thus, Lieberman (1993, p. 174) specifically states that "classic Neanderthal fossils retain the primitive condition – an unflexed basicranium and a supralaryngeal vocal tract ill-suited to speech production".[12]

hyoid muscles; this muscular deficiency may reflect a minor ability of *Homo erectus* to communicate in an articulate language" (Capasso *et al.* 2008, p. 1011).

[12] More recent studies (Dediu, Levinson 2013, 2018) have revised the hypothesis according to which Neanderthal was unable to produce complex vocal signals. Moreover, recent evidence shows that a baboon can make its larynx descend (Fitch, Reby 2001; Nishimura 2006). Thus, it is presumptive that a Neanderthal could do so too. This means that the simplification of the larynx might have favored formant-based communication. However, for our purposes, this revision is of no consequences. Neanderthal appears later, after Erectus. Instead, we hypothesize that earlier hominins – notably *Homo erectus* – lacked the ability to produce complex vocal sounds.

In short, the ancestors of *Homo sapiens* could encounter cetaceans at sea and did not yet possess the articulatory skills necessary to produce vocal sounds with acoustic formants and transitions.

## 7. VPL and VM

In this section we argue that VPL and VM were fully developed at the time of the possible "encounter" between hunting whales and humans. As noted in section 3, the cetaceans' ability to manage ultrasonic echolocation is far older than the appearance of humans. Vocal Production Learning (VPL) and Vocal Mimicry (VM) in humans could also date back to very early evolutionary stages, referring to *Homo erectus*, or *H. habilis*, or *H. neanderthalensis*.

Vocal production learning (VPL) is the ability to modify the structure of vocalizations as a result of hearing those of conspecifics or of other species. The most comprehensive compilation of studies on this topic was published in 2014 (Janik 2014; Knörnschild 2014; Reichmuth, Casey 2014; Stoeger, Manger 2014).

Phylogenetically, VPL relates to old stages of animal evolution (although the VPL trait is not shared by all branches of the vertebrate tree), because among animals a huge variety of sound production mechanisms and VPL techniques is employed, and most of them do not require a direct control over the larynx, but a simple control over the respiratory system. These mechanisms can alter frequency parameters as in amplitude modulations, adding side bands to signals or increased source levels, leading to subtle increases in fundamental frequency ($F_o$), but do not imply the influence of the filter system, as studied in human voice by the foundational work of Gunnar Fant (Fant 1960, 1966). For instance, birds use a syrinx capable of producing two sounds at the same time; in mammals, Odontocetes produce sounds with specifically evolved phonic lips in their nasal passages, and in elephants the trunk may be used as a sound source; in primates, lip smacking or unvoiced speech sounds are created by using parts of the mouth (Janik, Knörnschild 2021).

As for vocal mimicry (VM), it refers to the animal ability to learn and imitate a sound from another species or from environmental noises (e.g., water dripping, leaves rustling). It differs from VPL, as VM does not concern the members of the same species (conspecifics). For instance, parrots are the most renowned mimics (Benedict *et al.* 2022). VM has also been heard occasionally from bottlenose dolphins (*Tursiops truncatus*) (Reiss, McCowan 1993), harbour seals (*Phoca vitulina*) (Ralls *et al.* 1985), killer whales (*Orcinus orca*) (Abramson *et al.* 2018), orangutans (*Pongo* species) (Wich *et al.* 2008), and African savannah elephants (*Loxodonta africana*) (Stoeger, Manger 2014). The emergence of vocal mimicry is necessarily tied to the evolution of vocal learning, as mimicry requires the ability to acquire sounds through learning. Therefore, we assume that vocal mimicry could not have evolved prior to vocal learning.

Phylogenetically, VM is thus more recent than VPL, but it still remains an archaic trait, given that it is common to species much older than humankind. In these evolutionary stages the human species did not yet possess an anatomical apparatus that would allow it to articulate sounds efficiently, as Lieberman (1992, 1993) and others have demonstrated (see section 6).

As for the distribution of VPL and VM in the animal world, according to the literature, bats exhibit VPL (Knörnschild 2014); cetaceans exhibit both VM and VPL (Ridgway *et al.* 2012; Janik, Knörnschild 2021); primates do not exhibit VPL

(Zuberbühler *et al.* 2022);[13] and humans exhibit VPL (Janik, Knörnschild 2021; Vernes *et al.* 2021).

## 8. Further human evolution of vocal signal

We support the hypothesis that the acoustic characteristics of the human speech signal (used for communication purposes) derive from an adaptive evolution of the cetaceans' echolocation signals. Nevertheless, the human speech signal is far more complex than that of animal echolocation.

How did the dynamics of biosonar harmonics further evolve into the specialized forms necessary for the complexity of human speech and the variability of languages? Subsequent to the hypothetical human acquisition of the spectral characteristics of animal ultrasound, human beings then shared this "gift" with their conspecifics. And from this sharing came a subsequent evolution, which led to a refinement of the articulatory possibilities of the human vocal tract. The description of the main trends of this evolution can be summarized as follows.

The evolution of the pattern CF + FM into Formants + Transitions assigned CF and FM to functions no longer related to spatial localization. And this evolution induced some changes in the range of frequencies and their configurations. As a consequence, not only – as is obvious – has the human mammal not exploited the ultrasonic frequency range, but it has also adapted its brain to process and recognize more than one CF (in fact the timbre of the vowels is characterized by more than one steady formant).

Moreover, the human mammal adapted FM both with falling frequency and with rising frequency, in order to make it possible to use non-sonorant consonants from a wider range of places of articulation. In this way, the cetaceans' modulated harmonics became the human's formant transitions.

As regards the evolution from a single CF to several CFs (called formants), it should be noted that in the case of the whale, the CF of the call signal overlaps the CF of the echo signal. Therefore, in the passage from the system of mammals equipped with echolocation to that of humans, the two CFs of the whale have simply become the two formants (or more) necessary to discriminate the human's vowel timbres, while the co-evolutionary passage occurred by admitting that the two formants could be synchronous, rather than asynchronous: they could belong to the same signal (the human voice signal) instead of two signals (the cetacean's call signal and the echo signal).

Furthermore, co-evolution has also worked by translating the spatial "proximity" code of the whale into that of the vowel timbre of human speech. In the whale, the indication of the proximity of the target is given by the delay, the temporal distance

[13] "For production learning, that is, the ability to control the structure of sounds, primates rank at the other extreme end of flexibility in animal communication. Numerous studies have found that species, including great apes, are simply unable to mould their vocal output in any meaningful way, incapable of producing recognisable phonetic units or any similar properties of human speech […]. The only consistent kind of documented production learning is in terms of subtle modifications of existing call types that are already part of the vocal repertoire (Lemasson *et al.* 2011). Such accommodation or convergence of calls are typically responses to relationship variables, sometimes at the expense of individual recognition (Zurcher *et al.* 2021) and appear to be fairly widespread in primates, the result of sensory–motor integration also seen in humans (Janik, Knörnschild 2021; Ruch *et al.* 2018; Zurcher *et al.* 2021; Fischer *et al.* 2020). Vocal production learning, in short, is very modest in primates, a likely consequence of poorly evolved motor control of the sound-production apparatus (Lameira *et al.* 2014)" (Zuberbühler *et al.* 2022, p. 2).

between the call CF and the echo CF, and by the increase in frequency of the echo CF compared to the calling CF, mechanically generated by the Doppler effect; the distance is minimal if the delay is short and if the frequency increase is minimal (depending on the relative approaching velocity between the target and the whale). Similarly, the first two formants of the human vowels (or of the sonorant consonants) are distant in frequency if the vowel is front (e.g. [i] or [e]), i.e. farther from the glottis, but close in frequency if the vowel is back (e.g. [u] or [o]), i.e. closer to the speaker's glottis: the proximity in CF frequency of the whale's harmonics co-evolves in terms of proximity of human formants, endowing both mammals with the same function of indicating a spatial distance with respect to the emitting body.

Consequently, the human availability of more synchronous formants in the same acoustic signal triggers a further evolution: the whale's modulated harmonics (FM) become the humans' formant transitions, so that a very low F2 (second formant: for example that of a [u]) can also be adjacent to a consonant locus high in frequency (and therefore the transition of F2 can be rising), as in the case of a sequence [tu] or [ku], or a very high F2 (for example that of [i]) can be adjacent to the same consonant but the transition of F2 is falling, as in the case of a sequence [ti].[14]

In short, comparing the whale and *Homo sapiens* enables us to describe the acoustic features of human speech as the result of a co-evolution of the system of acoustic localization of objects in the common space inherited from mammals and specialized only by humans in order to allow them more sophisticated uses of their sensory apparatus.

## 9. Conclusions

In this paper we do not deal with the evolution of language, but only with the evolution of the human vocal signal.

We began with a review of studies concerning the main functions of hearing in the animal world. Then we described the ultrasonic echolocation in cetaceans and its spectral characteristics. In parallel, we described the human vocal signal and its spectral characteristics, and observed their remarkable acoustic similarities. Lastly, we have argued that it is reasonable to suppose the existence of a possible "natural" context in which the ultrasonic sounds of cetaceans have become "audible" to humans, and that at the time of this possible "encounter", humans had not yet developed a vocal signal with formants and transitions, whereas the cetacean echolocation was fully developed; in addition, humans were able to perform both VM and VPL. If we wanted to locate this "encounter" chronologically, it should be dated to the time of *Homo erectus*, that is, to the moment in which, according to Everett and to some recent archaeological discoveries, "language" was born.[15]

[14] Spectral analyses of human speech sounds "exhibit three basic acoustic patterns or components: CF components (formants), FM components (transitions), and noise bursts (fills). Vowels are identified mainly by the first and second formants (F1 and F2), although the third formant (F3) has some influence on recognition. Vowels are thus expressed in coordinates of F1 versus F2 frequencies and, to some extent, F1 versus F3 frequencies. Plosives, some of the fricatives, and combinations of phonemes are largely identified by combinations of transitions as well as by voice-onset times" (Suga *et al.* 1983, p. 1623).

[15] Lawrence Barham's recent discovery of the world's oldest wooden structure - by *H. erectus* - around Kalambo Falls in Zambia provides an additional support to the thesis that language was invented by *H. erectus* more than one million years ago (Barham *et al.* 2023; Barham, Everett 2021).

We have also added some further considerations regarding the later evolution of human vocal sounds, once they had become the shared heritage of human groups. This evolution can be considered the basis from which the great variety of vocal sounds, which differentiates the linguistic communities of the planet, was then generated.

In closing, we wish to answer a possible objection. Stable and modulated harmonics are not only found in the ultrasounds of cetaceans: they are also found in the so-called 'song' of cetaceans, performed in the "audible" acoustic range – which, however, is intended not for echolocation, but for "communication" between cetaceans of the same *pod* or is conspecific in any case. Why then would hominins have imitated the ultrasounds of cetaceans and not – more simply – the "audible" sounds of the cetaceans?

To answer, we must remember that our hypothesis is that the ultrasounds imitated by hominins were produced by cetaceans while hunting. Thus, these ultrasounds are related to a crucial biological function: the search for food – and this function is also important for humans. Instead, their 'song' has no biological function of human interest, as it serves to "communicate" only with other cetaceans, with other conspecifics.

Human beings would have no biological advantages in "talking" to cetaceans, while they would have if they were able to hunt *like* whales, in acquiring their same "technical" ability to procure food.

Therefore, while the imitation of the 'song' of cetaceans would not be useful for humans, it would be if humans were able to imitate the ultrasounds emitted by cetaceans during foraging.

**Bionotes:** Amedeo De Dominicis is full professor of Phonetics and Phonology at the University of Tuscia (Viterbo, Italy). He founded and directs the University Phonetics Laboratory (https://labfonetica.unitus.it). He has organized numerous international conferences and a summer school of Phonetics. He was visiting professor at the universities of Paris VII and Aix-en-Provence. He has been supervisor of PhD thesis both in Italy and abroad. He is the author of numerous publications listed on the website: https://labfonetica.unitus.it/dedominicis.html.

Alberto Preti received both master's degree and PhD at Sapienza University of Rome. After some years at Italian National Energy Agency and at the CNR Institute of Acoustics, he moved to the CNR Institute for Complex Systems where he is presently senior researcher. After working on laser optics, acoustics and elasticity of disordered systems, he is currently active on topics of statistical physics, including granular media and liquid crystals. AP has coauthored more than hundred peer reviewed papers, is associate editor of two ranked international scientific journals, and collaborates intensively with universities in Sao Paulo (BR) thorough frequent long-term visits.

**Authors' addresses:** dedomini@unitus.it; alberto.petri@isc.cnr.it

## References


Abramson J.Z., Hernández-Lloreda M.V., García L., Colmenares F., Aboitiz F. and Call J. 2018, *Imitation of novel conspecific and human speech sounds in the killer whale (*Orcinus orca*)*, in "Proceedings of the Royal Society B: Biological Sciences" 285 20172171, pp. 1-10.

Accetturo M., Creanza T.M., Santoro C., Tria G., Giordano A., Battagliero S., Vaccina A., Scioscia G. and Leo P. 2010, *Finding new genes for non-syndromic hearing loss through an in silico prioritization study*, in "PLoS ONE" 5 (9): e12742, pp. 1-16.

Alippi A., Shkerdin G., Bettucci A., Craciun F., Molinari E. and Petri A. 1992, *Low-Threshold Subharmonic Generation in Composite Structures with Cantor-Like Code*, in "Physical Review Letters" 69, pp. 3318-3321.

Barham L., Duller G.A.T., Candy I., Scott C., Cartwright C.R., Peterson J.R., Kabucku C., Chapot M.S., Melia F., Rots V., George N., Taipale N., Gethin P. and Nkombwe P 2023, *Evidence for the earliest structural use of wood at least 476,000 years ago*, in "Nature" 622, pp. 1-26.

Barham L. and Everett D. 2021, *Semiotics and the Origin of Language in the Lower Palaeolithic*, in "Journal of Archaeological Method and Theory" 28, pp. 535-579.

Bednarik R.G. 1999, *Maritime navigation in the Lower and Middle Palaeolithic*, in "Comptes Rendus de l'Académie des Sciences – Series IIA – Earth and Planetary Science" 328 [8], pp. 559-563.

Bednarik R.G. 2003, *Seafaring in the Pleistocene*, in "Cambridge Archaeological Journal" 13 (1), pp. 41-66.

Benedict L., Charles A., Brockington A. and Dahlin C.R. 2022, *A survey of vocal mimicry in companion parrots*, "Scientific Reports" 12, 20271.

Blumstein S.E. 1980, *Speech Perception: An Overview*, in Yeni-Komshian G., Kavanagh J.F. and Ferguson C.A. (eds.), *Child Phonology*, vol. ii, *Perception*, Academic Press, New York, pp. 9-21.

Bok E., Jin P.J., Choi H., Kyu H.C., Wright O.B. and Lee S.H. 2018, *Metasurface for Water-to-Air Sound Transmission*, in "Physical Review Letters" 120, 044302, pp. 1-4.

Boyd R., Silk J.B. 2000, *How Humans Evolved*, W. W. Norton & Company, Inc., New York.

Bretagne A., Tourin A. and Leroy V. 2011, *Enhanced and reduced transmission of acoustic waves with bubble meta-screens*, in "Applied Physics Letters" 99, 221906, pp. 1-3.

Businness.com 2022, https://www.businessresearchinsights.com/market-reports/bone-conduction-headphones-market-100857.

Capasso L., Michetti E. and D'Anastasio R. 2008, *A Homo erectus Hyoid Bone: Possible Implications for the Origins of the Human Capability for Speech*, in "Collegium antropologicum" 32 [4], pp. 1007-1011.

Carter T., Contreras D.A., Holcomb J., Mihailović D.D., Karkanas P., Guérin G., Taffin N., Athanasoulis D. and Lahaye C. 2019, *Earliest occupation of the Central Aegean (Naxos), Greece: Implications for hominin and Homo sapiens' behavior and dispersals*, in "Science Advances" 5 [10], pp. 1-9.

Clark G. 1947, *Whales as an economic factor in prehistoric Europe*, in "Antiquity" 21, pp. 84-104.

Cooper F.S., Delattre P.C., Liberman A.M., Borst J.M. and Gerstman L.J. 1967, *Some Experiments on the Perception of Synthetic Speech Sounds*, in Lehiste I. (ed.), *Readings in Acoustic Phonetics*, The MIT Press, Cambridge, MA, pp. 273-282.

Cranford T.W. 1999, *The sperm whale's nose: sexual selection on a grand scale?*, in "Marine Mammal Science" 15, pp. 1133-1157.

Cranford T.W., Amundin Mats and Norris Kenneth S. 1996, *Functional morphology and homology in the odontocete nasal complex: implications for sound generation*, in "Journal of Morphology" 228, pp. 223-285.

Davies K.T. J., Cotton J.A., Kirwan J.D. and Teeling E.C. 2012, *Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence*, in "Heredity" 108 [5], pp. 480-489.

Deacon T.W. 1997, *The Symbolic Species: The Co-evolution of Language and the Brain*, W.W. Norton & Company, New York.

Dediu D., Levinson S.C. 2013, *On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences*, in "Frontiers in Psychology" 4, pp. 1-17

Dediu D., Levinson S.C. 2018, *Neanderthal language revisited: not only us*, in "Current Opinion in Behavioral Sciences" 21, pp. 49-55.

Delattre P. 1970, *Des indices acoustiques aux traits pertinents*, in Hala B., Romportl M. and Janota P. (eds.), *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 6-13 September 1967, Academia Publishing House of the Czechoslovak Academy, pp. 35-47.

Delattre P., Liberman A.M. and Cooper F.S. 1955, *Acoustic Loci and Transitional Cues for Consonants*, in

"The Journal of the Acoustical Society of America" 27 [4], pp. 769-773.
Dror A.A., Avraham K.B. 2010, *Hearing impairment: a panoply of genes and functions*, in "Neuron" 68 [2], pp. 293-308.
Druzhinin O.A., Ostrovsky L.A. and Prosperetti A. 1996, *Low-frequency acoustic wave generation in a resonant bubble layer*, in "The Journal of the Acoustical Society of America" 100, pp. 3570-3580.
Durand M. 1954, *La perception des consonnes occlusives. Problèmes de palatalisation et de changements consonantiques*, in "Studia Linguistica: A Journal of General Linguistics" 8, pp. 110-122.
Eastman E.R. 1995, *Petrosal and inner ear of a squalodontoid whale: implications for evolution of hearing in odontocetes*, in "Journal of Vertebrate Paleontology" 15, pp. 431-442.
Eller A.I. 1984, *Subharmonic response of bubbles to underwater sound*, in "The Journal of Acoustical Society of America" 55, pp. 871-873.
Eller A., Flynn H.G. 1968, *Generation of Subharmonics of Order One-Half by Bubbles in a Sound Field*, in "The Journal of the Acoustical Society of America" 46, pp. 722-727.
Esteve M. 2020, *El estrecho vínculo vital entre las orcas y los neandertales*, in "Conversaciones Aquae | Podcast". https://www.fundacionaquae.org/neandertales-orcas-naturaleza/. (26.02.2023)
Everett D.L. 2017, *How Language Began: The Story of Humanity's Greatest Invention*, Liveright, New York.
Fant G. 1960, *Acoustic Theory of Speech Production*, Mouton, The Hague.
Fant G. 1966, *A note on vocal tract size factors and non-uniform f-pattern scalings*, in "Speech Transmission Laboratory Quarterly Progress and Status Report" 1, pp. 22-30.
Fischer J., Wegdell F., Trede F., Dal Pesco F. and Hammerschmidt K. 2020, *Vocal convergence in a multi-level primate society: insights into the evolution of vocal learning*, in "Proceedings of the Royal Society B: Biological Sciences" 287, Article 20202531.
Fitch T.W., Reby D. 2001, *The descended larynx is not uniquely human*, in "Proceedings of the Royal Society B: Biological Sciences" 268 [1477], pp. 1669-1675.
Fitzhugh W.W. and Chaussonnet V. (eds.) 1994, *Anthropology of the North Pacific*, Smithsonian Institution Press, Washington.
Fleischer G. 1976, *Über Beziehungen zwischen Hörvermögen und Schädelbau bei Walen*, in "Säugetierkundliche Mitteilungen" 24, pp. 48-59.
Fordyce R.E. 1980, *Whale evolution and Oligocene Southern Ocean environments*, in "Palaeogeography, Palaeoclimatology, Palaeoecology" 31, pp. 319-336.
Fordyce R.E. 1992, *Cetacean evolution and Eocene/Oligocene environments*, in Prothero D. and Berggren W. (eds.), *Eocene-Oligocene climatic and biotic evolution*, Princeton University Press, Princeton, NJ, pp. 368-381.
Fordyce R.E. 1994, *Waipatia maerewhenua, new genus and new species (Waipatiidae, new family), an archaic Late Oligocene dolphin (Cetacea: Odontoceti: Platanistoidea) from New Zealand*, in "Proceedings of the San Diego Museum of Natural History" 29, pp. 147-176.
Fordyce R.E. and de Muizon C. 2001, *Evolutionary history of cetaceans: a review*, in Mazin J.-M. and de Buffrénil V- (eds.), *Secondary Adaptation of Tetrapods to Life in Water*, Verlag Dr. Friedrich Pfeil, München, pp. 169-233.
Griffin D. 1959, *Echoes of bats and men*. Anchor Books, Garden City, N.Y.
Hemilä S., Nummela S. and Reuter T. 1999, *A model of the odontocete middle ear*, in "Hearing Research" 133, pp. 82-97.
Heyning J.E. 1989, *Comparative facial anatomy of beaked whales (Ziphiidae) and a systematic revision among the families of extant Odontoceti*, in "Contributions in science, Natural History Museum of Los Angeles County" 405, pp. 1-64.
Heyning J.E. and Mead J.G. 1990, *Evolution of the nasal anatomy of cetaceans*, in Thomas J.A. and Kastelein R.A. (eds.), *Sensory abilities of cetaceans*, Plenum, New York, pp. 67-79.
Janik V.M. 2014, *Cetacean vocal learning and communication*, in "Current Opinion in Neurobiology" 28, pp. 60-65.
Janik V.M. and Knörnschild M. 2021, *Vocal production learning in mammals revisited*, in "Philosophical Transactions of the Royal Society B" 376, pp. 1-10.
Kandel E.R., Schwartz J.H., Jessell T.M., Siegelbaum S.A. and Hudspeth A.J. (eds.) 2013, *Principles of Neural Science*, McGraw-Hill, New York.
Karpov S., Prosperetti A. and Ostrovsky L.A. 2003, *Nonlinear wave interactions in bubble layers*, in "The Journal of Acoustical Society of America" 113, pp. 1304-1316.
Kellogg R. 1936, *A review of the Archaeoceti*, in "Carnegie Institute of Washington Publication" 482, pp. 1-366.
Ketten D.R. 1991, *The marine mammal ear: specializations for aquatic audition and echolocation*, in

Douglas B.W., Popper A.N. and Fay R.R. (eds.), *The Evolutionary Biology of Hearing*, Springer-Verlag, New York, pp. 717-754.
Ketten D.R. 1992, *The cetacean ear: Form frequency and evolution*, in Thomas J.A., Kastelein R.A. and Supin A.Y. (eds.), *Marine Mammal Sensory Systems*, Plenum Press, New York, pp. 53-75.
Knörnschild M. 2014, *Vocal production learning in bats*, in "Current Opinion in Neurobiology" 28, pp. 80-85.
Kossl M., Foeller E., Drexl M., Vater M., Mora E.C., Coro F. and Russell I.J. 2003, *Postnatal development of cochlear function in the mustached bat, Pteronotus parnellii*, in "Journal of Neurophysiology" 90 [4], pp. 2261-2273.
Lameira A.R., Maddieson I. and Zuberbuehler K. 2014, *Primate feedstock for the evolution of consonants*, in "Trends in Cognitive Sciences" 18 [2], pp. 60-62.
Lee K.M., Isakson G.A. and Wilson P.S. 2018, *Improved object detection sonar using nonlinear acoustical effects in bubbly media*, in "Proceedings of Meetings on Acoustics" 29.
Lee T. and Iizuka H. 2020, *Sound propagation across the air/water interface by a critically coupled resonant bubble*, in "Physical Review B" 102, pp. 1-8.
Leighton T.G., Lingard R.J., Walton A.J. and Field J.E. 1991, *Acoustic bubble sizing by combination of subharmonic emissions with imaging frequency*, in "Ultrasonics" 29, pp. 319-323.
Leighton T.G., Richards S.D. and White P.R. 2004, *Trapped within a 'wall of sound'. A possible mechanism for the bubble nets of humpback whales*, in "Acoustics Bulletin" 29 [1], pp. 24-29.
Leighton T.G., White P.R. and Finfer D.C. 2005, *Possible applications of bubble acoustics in Nature*, in *Conference: Proceedings of the 28th Scandinavian Symposium on Physical Acoustics*, Ustaoset, Norway, 23-26 January 2005.
Lemasson A., Ouattara K., Petit E.J. and Zuberbuhler K. 2011, *Social learning of vocal structure in a nonhuman primate?*, in "BMC Evolutionary Biology" 11. https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-11-362.
Liberman A.M., Delattre P. and Cooper F.S. 1952, *The Role of Selected Stimulus-Variables in the Perception of Unvoiced Stop Consonants*, in "American Journal of Psychology" 65 [4], pp. 497-516.
Lieberman P. 1992, *On Neanderthal Speech and Neanderthal Extinction*, in "Current Anthropology" 33 [4], pp. 409-410.
Lieberman P. 1993, *On the Kebra KMH 2 Hyoid and Neanderthal Speech*, in "Current Anthropology" 34 [2], pp. 172-175.
Lieberman P. 2013, *The Unpredictable Species: What Makes Humans Unique*, Princeton University Press, Princeton.
Lieberman P. and Crelin E. S. 1971, *On the Speech of Neanderthal Man*, in "Linguistic Inquiry" 2 [2], pp. 203-222.
Li G., Wang J.H., Rossiter S.J., Jones G., Cotton J.A. and Zhang S.Y. 2008, *The hearing gene Prestin reunites echolocating bats*, in "Proceedings of the National Academy of Sciences" 105 [37], pp. 13959-13964.
Liu Y., Cotton J.A., Shen B., Han X., Rossiter S.J. and Zhang S.Y. 2010, *Convergent sequence evolution between echolocating bats and dolphins*, in "Current Biology" 20 [2], pp. R53-R54.
Luo Z., Li Y., Liu Z., Shi P. and Zhang J. 2010, *The hearing gene Prestin unites echolocating bats and whales*, in "Current Biology" 20 [2], pp. R55-R56.
Luo Z. and Gingerich P.D. 1999, *Terrestrial Mesonychia to aquatic Cetacea: transformation of the basicranium and evolution of hearing in whales*, in "University of Michigan papers on paleontology" 31, pp. 1-98.
MacNeilage P.F. 2008, *The Origin of Speech*. *Studies in the Evolution of Language*, Oxford University Press, Oxford/New York.
Mead J.G. 1975, *Anatomy of the external nasal passages and facial complex in the Delphinidae (Mammalia: Cetacea)*, in "Smithsonian contributions to zoology" 207, pp. 1-72.
Miller G.S. 1923, *The telescoping of the cetacean skull*, in "Smithsonian Miscellaneous Collections" 76, pp. 1-70.
Milo R.G., Quiatt D., Aiello L.C., Burling R., Frayer D.W., Gargett R.H., Gibson K.R., Jessee S., Kien J., Krantz G.S., Peters E.H., Ragir S., Wallace R., Wescott R.W., Wilson L., Wolpoff M.H. and Wynn T. 1993, *Glottogenesis and Anatomically Modern Homo Sapiens. The Evidence for and Implications of a Late Origin of Vocal Language*, in "Current Anthropology" 34 [5], pp. 569-598.
Monks G.G., Mcmillan A.D. and St Claire D.E. 2001, *Nuu-Chah-Nulth whaling: archaeological insights into antiquity, species preferences, and cultural importance*, in "Arctic Anthropology" 38, pp. 60-81.
Nishimura T. 2006, *Descent of the Larynx in Chimpanzees: Mosaic and Multiple-Step Evolution of the Foundations for Human Speech*, in Matsuzawa T., Tomonaga M. and Tanaka M. (eds.), *Cognitive*

*Development in Chimpanzees*, Springer, Tokyo, pp. 75-95.
Norris Ke.S. 1968, *The evolution of acoustic mechanisms in odontocete cetaceans*, in Drake E.T. (ed.), *Evolution and environment: a symposium presented on the occasion of the one hundredth anniversary of the foundation of Peabody Museum of Natural History at Yale University*, Yale University Press, New Haven, pp. 297-324.
Nummela S., Reuter T., Hemila S., Holmberg P. and Paukku P. 1999a, *The anatomy of the killer whale middle ear (*Orcinus orca*)*, in "Hearing Research" 133, pp. 61-70.
Nummela S., Wägar T., Hemila S. and Reuter T. 1999b, *Scaling of the cetacean middle ear*, in "Hearing Research" 133, pp. 71-81.
Ostrovski L.A. 2003, *Nonlinear scattering of acoustic waves by natural and artificially generated subsurface bubble layers in sea*, in "The Journal of the Acoustical Society of America" 111, pp. 741-749.
Ostrovsky L.A., Sutin A.M., Soustova I.A., Matveyev A.I. and Potapov A.I. 1998, *Nonlinear, low-frequency sound generation in a bubble layer: Theory and laboratory experiment*, in "The Journal of the Acoustical Society of America" 104 [2], pp. 722-726.
Payne R.S. and McVay S. 1971, *Songs of Humpback Whales*, in "Science" 173 [3997], pp. 585-597.
Peregrine D.H. 1983, *Breaking waves on beaches*, in "Annual Review of Fluids Mechanics" 15, pp. 149-178.
Pompeckj J.F. 1922, *Das Ohrskelett von Zeuglodon*, in "Senekenbergiana" 4, pp. 43-100.
Purves P.E. and Pilleri G.E. 1983, *Echolocation in whales and dolphins*, Academic press, London.
Ralls K., Fiorelli P. and Gish S. 1985, *Vocalizations and vocal mimicry in captive harbor seals,* Phoca vitulina, in "Canadian Journal of Zoology" 63 [5], pp. 1050-1056.
Reichmuth C. and Casey C. 2014, *Vocal learning in seals, sea lions, and walruses*, in "Current Opinion in Neurobiology" 28, pp. 66-71.
Reidenberg J.S. and Laitman J.T. 1988, *Existence of vocal folds in the larynx of Odontoceti (toothed whales)*, in "The Anatomical Record" 221, pp. 884-891.
Reiss D. and McCowan B. 1993, *Spontaneous vocal mimicry and production by bottlenose dolphins (Tursiops truncatus): evidence for vocal learning*, in "Journal of Comparative Psychology" 107 [3], pp. 301-312.
Ridgway S., Carder D., Jeffries M. and Todd M. 2012, *Spontaneous human speech mimicry by a cetacean*, in "Current Biology" 22 [20], pp. R860-R861.
Rosen J. and Gothard L.Q. 2009, *Encyclopedia of Physical Science*, Infobase Publishing, New York.
Ruch H., Zürcher Y. and Burkart J.M. 2018, *The function and mechanism of vocal accommodation in humans and other primates*, in "Biological Reviews" 93 [2], pp. 996-1013.
Sales G. and Pye D. 1974, *Ultrasonic communication by Animals*, Chapman and Hall, London.
Samarra F.I.P., Deecke V.B., Vinding K., Rasmussen M.H., Swift R.J. and Miller P.J.O. 2010, *Killer whales (*Orcinus orca*) produce ultrasonic whistles*, in "The Journal of the Acoustical Society of America" 128 [5], pp. EL205–EL210.
Savelle J.M. and Kishigami N. 2013, *Anthropological Research on Whaling: Prehistoric, Historic and Current Contexts*, in Kishigami N., Hamaguchi H. and Savelle J.M. (eds.), *Anthropological Studies of Whaling.* Senri ethnological studies 84, National Museum of Ethnology, Osaka, pp. 1-48.
Seersholm F.V., Pedersen M.W., Søe M.J., Shokry H., Mak S.S.T., Ruter A., Raghavan M., Fitzhugh W., Kjær K.H., Willerslev E., Meldgaard M., Kapel C.M.O. and Hansen A.J. 2016, *DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4,000 years ago*, in "Nature Communications" 7. https://www.nature.com/articles/ncomms13389.
Smith A.B. and Kinahan J. 1984, *The invisible whale*, in "World Archaeology" 16, pp. 89-97.
Stoeger A.S. and Manger P. 2014, *Vocal learning in elephants: neural bases and adaptive context*, in "Current Opinion in Neurobiology" 28, pp. 101-107.
Suga N., O'Neill W.E., Kujirai K. and Manabe T. 1983, *Specificity of combination-sensitive neurons for processing of complex biosonar signals in auditory cortex of the mustached bat*, in "Journal of Neurophysiology" 49 [6], pp. 1573-1626.
Tejedor S.M.T., Louisnard O. and Vanhille C. 2022, *Generation of subharmonics in acoustic resonators containing bubbly liquids: A numerical study of the excitation threshold and hysteretic behavior*, in "Ultrasonics Sonochemistry" 88, pp. 1-9.
Tobias P. 1998, *Evidence for the Early Beginnings of Spoken Language*, in "Cambridge Archaeological Journal" 8, pp. 72-78.
Tyson R.B., Nowacek D.P. and Miller P.J.O. 2007, *Nonlinear phenomena in the vocalizations of North Atlantic right whales* (Eubalaena glacialis) *and killer whales* (Orcinus orca), in "The Journal of the Acoustical Society of America" 122, pp. 1365-1373.

van den Bergh G.D., Kaifu Y., Kurniawan I., Kono R.T., Brumm A., Setiyabudi E., Aziz F. and Morwood M.J. 2016, *Homo floresiensis-like fossils from the early Middle Pleistocene of Flores*, in "Nature" 534, pp. 245-248.

Vater M. and Kössl M. 2004, *Introduction: The ears of whales and bats*, in Thomas J.A., Moss C.F. and Vater M. (eds.), *Echolocation in Bats and Dolphins*, The University of Chicago Press, Chicago/London, pp. 89-98.

Vernes S.C., Janik V.M., Fitch W.T. and Slater P.J.B. 2021, *Vocal learning in animals and humans*, in "Philosophical Transactions of the Royal Society B" 376 [1836]: 20200234.

Wahlberg M. and Surlykke A. 2014, *Sound Intensities of Biosonar Signals from Bats and Toothed Whales*, in Surlykke A., Nachtigall P.E., Fay R.R. and Popper A.N. (eds.), *Biosonar*, Springer Handbook of Auditory Research 51, Springer-Verlag, New York, pp. 107-141.

Wich S.A., Swartz K.B., Hardus M.E., Lameira A.R., Stromberg E. and Shumaker R.W. 2008, *A case of spontaneous acquisition of a human sound by an orangutan*, in "Primates" 50 [1], pp. 56-64.

Wood F.G. and Evans W.E. 1980, *Adaptiveness and ecology of echolocation in toothed whales*, in Busnel R.-G. and Fish J.F. (eds.), *Animal sonar systems*, Plenum, New York, pp. 381-425.

Wynn T. 1998, *Did* Homo erectus *Speak?*, in "Cambridge Archaeological Journal" 8, pp. 78-81.

Yamaura K. 1980, *On the relationships of the toggle harpoon heads discovered on the northwestern shores of the Pacific*, in "Material Culture" 35, pp. 1-19.

Yen N.-C. 1971, *Subharmonic generation in acoustic systems*. Memorandum No. 85, Harvard University, Cambridge, Massachusetts.

Zuberbühler K., León J., Deshpande A. and Quintero F. 2022, *Socially scripted vocal learning in primates*, in "Current Opinion in Behavioral Sciences" 46, 1-5. https://www.sciencedirect.com/science/article/pii/S2352154622000596?.

Zürcher Y., Willems E.P. and Burkart J.M. 2021, *Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations*, in "Scientific Reports" 11. https://www.nature.com/articles/s41598-021-95101-8.

# APPENDIX

## The hypothesis of air bubbles by whales

A possible mechanism of ultrasound conversion is constituted by air bubbles in water. Bubbles are efficient nonlinear resonators (se Druzhinin *et al.* 1996, Lee *et al.* 2018) and can be generated in many ways. They are produced by breaking waves in the oceanic subsurface layer (Ostrovsky 2003) and at the seaside (Peregrine 1983). In addition, during the hunt, cetaceans produce a net of bubbles to trap prey (see Leighton *et al.* 2004). Bubbly liquids can easily generate acoustic nonlinearities also at low bubble concentration and for low-level excitation (Karpov *et al.* 2003).

An acoustic wave propagating in an elastic medium usually determines its linear response, thus producing deformations in the medium modulated in time at the same frequency of the incoming wave.

If, however, the wave intensity is large and/or the medium is compliant, the response *R(t)* can be nonlinear and can in general be expressed as a power series of the incident wave intensity or pressure *P* (e.g., Leighton *et al.* 1991):

$$R(t) = \alpha_1 P(t) + \alpha_2 P^2(t) + \alpha_3 P^3(t) + \alpha_4 P^4(t) + \ldots, \quad (1)$$

with $\alpha_n$ constant. This can generate supplementary waves with frequencies different from the incident wave. The large compressibility of gas bubbles makes them very efficient converters. In the conversion, a primary role is played by the bubble oscillation frequency. For small amplitude pulsations and neglecting the surface tension, the natural frequency of oscillation is (Eller, Flynn 1968):

$$f \approx \frac{1}{2\pi r_n}\sqrt{\frac{3\gamma P_0}{\rho}}$$

where $r_n$ is the bubble radius, $P_0$ is the liquid pressure, $\varrho$ the liquid density, and $\gamma$ the specific heat ratio $\cong 1 \div 1.4$.

Bubble radius varies from millimeters to microns; hence at normal pressure and water density, bubbles naturally resonate at frequencies ranging from ≈ 1 kHz to beyond 1 MHz. Another interesting fact is that in most situations bubble size can be found distributed according to power laws (Leighton *et al.* 2005). This implies that bubbles of any size are present, hence supplying a broad spectrum of frequencies for conversion.

Nonlinear elasticity can give rise to down conversion of frequency by different mechanisms. One mechanism is subharmonics excitation. If the bubble is invested by a wave with a frequency $\omega$ which is close to a multiple $n$ of its resonating (angular) frequency $\omega_0$ the nonlinear response (1) will contain the terms $cos^n(n\omega_0 t)$. So, for instance, from the quadratic term in (1) one has $cos^2(2\omega_0) = (1 + 2\ cos\ (\omega_0 t))/2$; i.e., the bubble can resonate at its proper frequency. With somewhat more involved dynamics, explained by the Rayleight-Plesset equation (see e.g., Leighton *et al.* 1991) or its modification, it can resonate at a submultiple frequency (subharmonics ordinarily appear in pairs such that the sum of each frequency pair equals the frequency of the driving signal). These phenomena have been known for a long time (see e.g., Yen 1971, Eller 1984 and refs. therein) and have been widely investigated both theoretically and experimentally. It is interesting that layers of a bubbly liquid surrounded by pure liquid can give rise to strongly nonlinear behavior even for relatively low-level excitation

(Karpov *et al.* 2003). Bubble layers can act as resonators enhancing the nonlinear bubble response. In a recent study, amplitude thresholds for the generation of subharmonics down to ¼ of the driving frequency have been achieved in bubble resonators (Tejedor Sastre *et al.* 2022).

Another mechanism of down conversion facilitated by the large compressibility of gas bubbles is frequency mixing. Layers of bubbles surrounded by pure liquid exhibit many resonances, hence a bubbly layer excited in correspondence of two resonant modes at frequency $f_1$ and $f_2$ can efficiently generate a signal at low-frequency $f$ corresponding to the difference frequency $f_1$- $f_2$, when this also corresponds to a resonant mode (see Druzhinin *et al.* 1996). In addition to having been predicted theoretically, this parametric mechanism has been observed and calculated in several specific experiments obtaining signals with frequencies around 1 kHz from signals around 30 or 60 kHz (see Ostrovsky *et al.* 1998, Ostrovsky 2003). Naturally produced bubbles in the ocean show an extremely wide range of size and thus of resonant frequencies (see Leighton *et al.* 2005). Hence, parametric down conversion can occur for a very large number of frequencies. Moreover, self-similarity can facilitate subharmonic generation (Alippi *et al.* 1992); therefore, the power law distribution of bubble sizes mentioned above constitutes a favorable environment for this process, also triggering a nonlinear effect of frequency down conversion. The efficiency of this process depends on the bubble size distribution function and on the frequencies of the interacting waves (see Wahlberg *et al.* 2014). Moreover, small non-resonant bubbles decrease the sound speed so that a jump in acoustic impedance occurs at the layer boundaries. The boundary then becomes a planar resonator (Ostrovsky *et al.* 1998).

If bubble layers in a surrounding pure liquid can act as resonators enhancing down conversion efficiency, as recent studies have highlighted, then sound can be trapped within them because of the high acoustic impedance difference, with bubble-free water causing strong reflection at the bubble layer boundaries (Ostrovsky *et al.* 1998). For the same reason, sound wave transmission from water to air and vice versa is very poor. To increase it, the water-air interface requires a solid transducer decreasing the difference of impedance and allowing the transmission of sound. This is the case for the bones of the inner ear, where a liquid interacts with air by means of the three ossicles: hammer, stirrup, and anvil. Thus, a simple way to hear sounds travelling in the sea is to put one's ear on the surface of a body also in contact with water. An example is the gunwale of a pirogue, where sounds incident from water on the hull can be heard. Another example is Arctic ice, on which hunters make a hole to access fish.

Let us consider a wave at the water-ice or water-wood interface. Say $\rho_1$ and $c_1$ density and sound speed for water, and $\rho_2$ and $c_2$ for the solid. In the simple case of incidence perpendicular to a flat interface, the transmission coefficient in terms of transmitted and incident wave amplitude, $A_t$ and $A_i$, is

$$T = \frac{A_t}{A_i} = \frac{2\rho_2 c_2}{\rho_1 c_1 + \rho_2 c_2}$$

For water $\rho_1 c_1 \cong 1.5\ 10^6$, kg m$^{-1}$ s$^{-2}$ whereas $\rho_2 c_2 \cong 3.5\ 10^6$kg m$^{-1}$ s$^{-2}$ for ice, yielding T≈ 3/5. For wood, both density and sound speed can vary depending on the species and the direction with respect to the wood grains. For not too heavy woods like pine or spruce, $\rho_2 \approx 450 \div 550$ kg m$^3$ and the sound speed $c_2 \cong$ (3.3 ÷ 5) $10^3$ m s$^{-1}$, yielding $0.7 \leq T \leq 1$.

The above example shows how efficient a solid body to collect sound from water can be. In short, wave transmission depends on the specific acoustic impedance $z = \rho c$ of the media in contact. The more they are similar, the more the wave is transmitted. If a solid body presents two surfaces, one in water and the other in air, such as a floating ice layer or a boat, sound can be heard by putting one's head in close contact with the surface. In this way bone conduction bypasses eardrum and sound is effectively transmitted to the inner ear through the skull (this is the principle behind bone conduction headphones, a market in constant linear growth expected to reach about 3,000 million dollars in 2028 (Businness.com 2022).

Coupling between water and solid is not the only way to collect sound from water. Bubble layers unveil particular properties in this case as well. Indeed, several studies have shown that their presence can enhance air-water transmission (see e.g., Bretagne *et al.* 2011, Lee, Iizuka 2020, and refs. therein). This appears to be a recent field of research, and very few results are available regarding the reverse, water-air transmission, although the acoustic wave equation obeys the principle of reciprocity and must yield the same reflection and transmission coefficients. One study (Bok *et al.* 2018) has investigated water-air transmission through an array of individual elements containing membranes and an air-filled cavity. Each of these individual systems, called "meta-atoms", consists of a tube, in part filled with water and in part with air, placed at the air-water interface to achieve impedance matching. The work experimentally demonstrates that the transmission of sound at ≈ 700 Hz is increased even by 2 orders of magnitude, allowing about 30% of the incident acoustic power from water to be transmitted into air (Bok *et al.* 2018). In addition to the aforementioned reciprocity, the similarity of this system to a bubble layer suggests that the latter could also increase water-air sound transmission.

In summary, transmission of sonic waves can be accomplished in an effective manner through a solid interface both in ultrasonic and audible range. The latter was considered in Sections 5-6, while down conversion by air bubbles can be considered as one possible channel of sound modulation spillover aside from audible frequencies.