Auditory capacities of human fossils: a new approach to
the origin of speech

I. Martínez\textsuperscript{a}, R.M. Quam\textsuperscript{b}, M. Rosa\textsuperscript{c}, P. Jarabo\textsuperscript{c}, C. Lorenzo\textsuperscript{d} and J.L. Arsuaga\textsuperscript{e}

\textsuperscript{a}(1) Universidad de Alcalá de Henares; (2) Centro UCM-ISCIII de Evolución y Comportamiento Humanos, C/ Sinesio Delgado, Nº 4, Pabellón 14, 28029 Madrid, Spain
\textsuperscript{b}(1) American Museum of Natural History; (2) Centro UCM-ISCIII de Evolución y Comportamiento Humanos, Division of Anthropology, American Museum of Natural History, Central Park West @ 79th St., New York, NY 10024, USA
\textsuperscript{c}Universidad de Alcalá de Henares, Edificio de Ciencias, Campus Universitario, 28871 Alcalá de Henares, Spain
\textsuperscript{d}(1) Universitat Rovira i Virgili (2) Centro UCM-ISCIII de Evolución y Comportamiento Humanos, Universitat Rovira i Virgili, Facultat de Lletres Institut de Paleoecologia Humana, Area de Prehistoria, i Evolució Social, Imperial Tarraco, 1, 43005 Tarragona, Spain
\textsuperscript{e}(1) Universidad Complutense de Madrid; (2) Centro UCM-ISCIII de Evolución y Comportamiento Humanos, Centro UCM-ISCIII de Evolución y Comportamiento Humanos, C/ Sinesio delgado, Nº 4, Pabellón 14, 28029 Madrid, Spain
imartinezm@isciii.es
The origin and evolution of human language has mainly dealt with the reconstruction of the upper respiratory tract of human fossils. After decades of controversy, no clear results have arisen from these studies. We propose a new approach to this issue based on the possibility to reconstruct the sound power transmission, through the external and middle ear, in fossil specimens. The results thus obtained in the more than 500 kyr old fossils from the Sima de los Huesos site (Sierra de Atapuerca, Spain) show that these hominins had the same auditory capacities as modern humans. Specifically, they show a widened bandwidth of heightened sensitivity in the midrange frequencies compared with chimpanzees. Relying on the theory of communication as developed by Shannon, this widened bandwidth suggests a greater channel capacity characterized the Atapuerca (SH) hominins and is consistent with other recent suggestions for an ancient origin for human speech capacity.

1 Introduction

Studies of sensory perception in our fossil human ancestors would greatly enhance our understanding of their lifeways and adaptations. Of all the human special senses, audition is the most readily accessible in skeletal remains since it is based on physical properties that can be approached through their skeletal structures [1,2]. Studying auditory capacities in fossil species is a major challenge, but has become feasible since the advent of CT-based analyses.

Auditory sensitivity across a wide range of frequencies has been measured in a relatively large number of primate species [3–18]. These studies have revealed general patterns in the audiogram data which appear to differ following the major taxonomic divisions within the Primate order (Figure 1), and the chimpanzee is characterized by an audiogram which is similar to most anthropoid species tested to date. The chimpanzee audiogram shows two peaks of heightened sensitivity at around 1 and 8 kHz, separated by a region of lower sensitivity at around 4 kHz. In contrast, the human audiogram differs from other primate species in showing a broad region of heightened sensitivity between approximately 1-6 kHz (Figure 1) [5, 12, 13, 14, 15, 19, 20]. Thus, humans appear to have widened the bandwidth of maximum sensitivity in these midrange frequencies. Given this difference in the audiogram between chimpanzees and humans, the study of auditory capacities in our fossil human relatives has the potential to reveal when the modern human auditory pattern first emerged in the fossil record.

Recently, an electrical circuit model of external and middle ear function developed on modern humans [21-23] was applied to 3D CT reconstructions of the outer and middle ears of the Middle Pleistocene human fossils from the site of the Sima de los Huesos (SH) at Atapuerca. The results of this innovative approach demonstrated that the Atapuerca hominins had auditory capacities similar to our own [24], and show a region of heightened sensitivity between 2-5 kHz. This represents the first time that an aspect of sensory perception has been measured in our fossil human ancestors. It is important to point out that while much of the acoustic information in human spoken language (e.g., the first two formant frequencies of the vowels) is concentrated at frequencies below about 2.5 kHz [25-27], the region between 2.5-5 kHz, where humans and the Atapuerca hominins maintain a higher auditory sensitivity than chimpanzees, also contains relevant acoustic information in human speech [26, 28, 29].

Although the study of audition is an indirect approach to the question of speech capacity in fossil specimens, prior attempts to reconstruct the linguistic capacities in fossil hominins have led to conflicting results and have often not been based on sound anatomical relationships between skeletal structures and language production [27, 30-32]. Interestingly, the recent discovery that Neandertals share with modern humans two derived substitutions in the FOXP2 gene offers tantalizing new evidence for the possible presence of spoken language in Neandertals [33]. At the same time, the beginnings of human speech have recently been suggested to date to the origin of the genus Homo [34, 35]. It is important to point out that the precise nature of this verbal communication in fossil hominins is not known, but need not be equivalent in syntax, linguistic flexibility or cognitive complexity to present day human spoken language. Thus, the implications drawn from the study of auditory capacities based on the Atapuerca hominins are consistent with the results of these recent studies.

Given the strong genetic component to the development of the ear structures and the vocal flexibility exhibited by modern human spoken language, it is not unreasonable to conclude that the mechanisms of speech production have probably adapted to the audible hearing range, rather than vice versa [36]. In addition, given the intuitive, but difficult to quantify, link between sound perception and vocal production in animals [5,29,37,38,39], study of the auditory capacities in our fossil human ancestors may have implications for the emergence of language during the course of our evolutionary history.

The present contribution reviews the reconstruction of the auditory capacities in the Atapuerca (SH) hominins and analyzes the capacity of the outer and middle ears in these fossil human ancestors.
hominins as communication channels, as defined by Shannon [40] in the Theory of Communication. According to [40], a channel is a transmission medium in which signals travel to transmit information from the source to the receiver. For communication to be efficient, the channel should be adapted to the kind of signal to be transmitted. Since the capacity of the channel is an indication of the amount of information that could be transmitted without any loss, we believe that this approach can help place the relationship between sound production and sound perception in our fossil human ancestors on a firmer theoretical basis.

2 Modeling the outer and middle ear sound power transmission in modern humans, chimpanzees and middle Pleistocene humans

The use of electrical circuits to model sound power transmission through the outer and middle ears is a common practice in auditory research [23,41,42]. A slightly modified version of the electrical circuit model published in [23] was used to estimate the sound power transmission through the outer and middle ears in the Atapuerca (SH) hominins, modern humans and one chimpanzee (Pan troglodytes) individual [24]. The modification introduced into the model refers to the cochlear input impedance (Zc) which was directly measured in 11 human cadaver ears by Aibara et al. [43]. These authors found a flat, resistive cochlear input impedance of about 21.1 Ω from 0.1-5.0 kHz, and this empirical value was used in the model. The model was compared with measurements on humans cadavers made by Aibara et al. [43], finding no statistically significant difference.

While some fundamental aspects of hearing are determined by properties of the inner ear [44], the anatomical structures of the outer and middle ear are responsible for producing an acoustic filtering of the audible sound energy in the environment and delivering this energy to the hearing organ (Organ of Corti) housed in the cochlea [21]. This acoustic filtering is not uniform across frequencies, with some frequencies transmitted better than others, and the outer and middle ear play important roles in shaping the audiogram.

This model incorporates nearly 30 variables related to head size and the dimensions and physical properties of the anatomical structures of the outer and middle ear, and provides reliable results up to at least 5 kHz. Seventeen of the model variables are related to soft tissue structures (e.g. cartilage, ligaments) which cannot be studied in fossils (or bony skulls of extant primates), limiting the measurements in fossil specimens to those related to the skeletal structures of the outer and middle ear. For the remaining soft-tissue variables, the modern human values must be used in fossil specimens since nothing is known of the soft-tissue properties of the ear structures in chimpanzees. Nevertheless, a sensitivity study of the individual soft-tissue variables [24] revealed that most of these do not have an appreciable effect on the model results above 2 kHz. Thus, any difference from modern humans in the model results for the fossil specimens will be primarily due to variation in the skeletal anatomy of the outer and middle ear.

2.1 Calculus of sound power transmission through the external and middle ear

The effective area at the oval window is defined by Rosowski [21] as the parameter that accounts for the power loss in the external and middle ears, according to the following expression:

$$E_{AW}^{(\phi,\theta)} = \frac{\text{Power into the Inner Ear}}{\text{Intensity of the Incident Plane Wave}} = E_{ATM}^{(\phi,\theta)} \cdot \text{MEE}$$

(1)

Where $E_{AW}^{(\phi,\theta)}$ defines the effective area at the tympanic membrane as the ratio between the power into the middle ear and the intensity of the incident plane wave.

$$E_{ATM}^{(\phi,\theta)} = \frac{\text{Power into the Middle Ear}}{\text{Intensity of the Incident Plane Wave}} = \frac{\rho_t c}{\rho_0 c} \frac{\text{Re} \{Z_c\}}{\text{Re} \{Z_T\}}$$

(2)

On the other hand, MEE is the middle-ear efficiency that quantifies how much of the power entering the middle ear actually reaches the inner ear.

$$E_{MEE}^{(\phi,\theta)} = \frac{\text{MEE}}{\text{Power into the Cochlea}} = \frac{\text{Power into Middle Ear}}{\text{Power into Middle Ear}} = \frac{\text{Re} \{Z_c\}}{\text{Re} \{Z_T\}}$$

(3)

By using Eqs. (1), (2), and (3), the effective area of the oval window can be calculated according to the following expression:

$$MEE = \frac{\text{Power into the Cochlea}}{\text{Power into Middle Ear}} = \frac{\text{Power into Middle Ear}}{\text{Power into Middle Ear}} = \frac{\text{Re} \{Z_c\}}{\text{Re} \{Z_T\}}$$

(4)

If the gain produced by diffraction and scattering about the head and the radiation impedance looking out from the ear opening into the environment are both insignificant, the effective area of the oval window can be calculated as follows:

$$E_{AW}^{(\phi,\theta)} = \frac{U_{S}}{P_{PW}} \frac{\text{Re} \{Z_c\}}{\text{Re} \{Z_T\}} \rho_0 c$$

(5)

In the expressions above, $Z_T$ is the impedance at the tympanic membrane, $Z_C$ is the impedance at the cochlea, $\rho_0$ is the static density of air, $c$ is the propagation velocity of sound, A and B are transmission parameters of the external and middle ears equivalent circuit, $P_T$ is the sound pressure at the tympanic membrane, $P_{PW}$ is the sound pressure of the incident plane wave, and $U_S$ is the volume velocity at the stapes.

2.2 The Atapuerca (SH) human fossils

The Sierra de Atapuerca is well known for the extraordinarily large sample of Middle Pleistocene human fossils recovered from the site of the Sima de los Huesos [45]. Prior to the 2006 field season, the SH site has yielded more than 5,500 human fossils, belonging to at least 28 individuals [46], which have been assigned to the species Homo heidelbergensis and are considered to represent the ancestral European population that evolved into the Neandertals [47–49]. The SH site has a firm minimum radiometric age of 530 kyr [50].

We have analyzed the sound power transmission through the outer and middle ears in five individuals from the SH site. To measure the necessary variables in these
individuals, we have relied on 3D CT reconstructions of the temporal bone, as well as the presence of well-preserved auditory ossicles within the collection [24]. Finally, to evaluate the influence of the skeletal variables on the interspecific difference in the acoustic filtering patterns, we have measured, through 3D CT reconstruction, the skeletal variables in one chimpanzee individual (Pan troglodytes), and we have modeled it by using the modern human values [23, 43] for the remaining soft-tissue-related variables, as we have done in the SH specimens.

3 The Theory of Communication

In his classic paper, C.E. Shannon [40] states: “The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning”. In the same paper, he defines what is a communication system. It is composed of essentially five parts:

1. An information source which produces a message or sequence of messages to be communicated.
2. A transmitter which produces a signal suitable for transmission over the channel.
3. The channel, that is the medium used to transmit the signal from the transmitter to the receiver.
4. The receiver, that is the inverse part of the transmitter.
5. The destination, that is the person or thing for which the message is intended.

Human communication conforms to this definition of communication system, and the five components can be identified. The information source is the brain of the speaker, which produces a message with some meaning to be communicated to the destination. Using the lungs, larynx and mouth, speech is produced that should be suitable for transmission over the channel. The channel is composed of the air and auditory organs. The receiver can be considered the inner ear and the brain of the person to whom information is transmitted, that is the destination.

It is not possible to study either the source or the transmitter in fossils, since they are formed of soft tissues that are not preserved. In contrast, data from the channel are preserved, and channel capacity, as an indication of the amount of information that could be transmitted without any loss, can be estimated. Channel capacity of a noisy channel is defined in the following theorem [40]:

**Theorem:** The capacity of a channel of band W perturbed by an arbitrary noise is bounded by the inequalities:

$$W \cdot \log_2 \left( \frac{S + N_i}{N} \right) \leq C \leq W \cdot \log_2 \left( \frac{S + N}{N_i} \right)$$

Where S is average transmitter power, N is the average noise power, and N \(_i\) is the entropy power of the noise. If N=N\(_i\) for white noise, it gives the well known formula:

$$C = W \cdot \log_2 \left( 1 + \frac{S}{N} \right)$$

The channel capacity is measured in bits per second (bps).

If we assume that the signal to noise ratio is the same for all individuals, we can focus on channel bandwidth, which then becomes a proxy for channel capacity. A wider channel bandwidth would correspond to a greater channel capacity and would allow for the transmission of a larger quantity of information. Although a number of definitions for channel bandwidth can be considered, we have used the **occupied bandwidth** [51], defined as the bandwidth such as under the lower cutoff frequency and over the upper cutoff frequency, the average power is equal to a specified percentage, \(\beta/2\), of the total average power. In this paper we have considered \(\beta/2\) equal to 5%.

4 Results

The sound power transmission curves obtained for the SH hominins, chimpanzee and modern human individuals are shown in Fig. 2. Above 3 kHz the chimpanzee curve shows a sharp drop in sound power transmission, while the modern human curve maintains higher values for sound power transmission. Between 3-5 kHz, the human and chimpanzee curves are separated by approximately 10 dB or more, coinciding with previous results based on audiograms [6, 13]. We interpret this agreement as evidence that the differences in skeletal anatomy can explain much of the interspecific differences in the sound power transmission between these closely related species, since the values for the soft-tissue variables were held constant. Consequently, these differences in skeletal anatomy can also be used to validly infer sound power transmission patterns in fossil hominins.

![Fig. 2. Sound power (dB) at the entrance to the cochlea relative to \(P_0=10^{-12}\) W for an incident plane wave intensity of \(10^{-12}\) W/m\(^2\). All individuals have been modeled by using the model defined by Rosowski [23] and the cochlear input impedance (Zc) of Aibara et al. [43]. Solid blue line, modern human; solid green line, chimpanzee 3D CT; solid red line, AT-84; dashed red line, AT-4103; dashed-dotted red line, Cranium 5; solid magenta line, AT-421; dashed magenta AT-1907.](image-url)

The sound power transmission curves obtained for the SH hominins fall near both the chimpanzee and modern human curves up to around 2 kHz. At higher frequencies, the results for the SH individuals are more similar to modern...
humans, and lack the sharp dropoff in sound power transmission seen in chimpanzees.

Based on the model results, the occupied bandwidth can be measured (Table 1). This bandwidth is defined as the range of frequencies in which the 90% of sound power is transmitted through the outer and middle ear [51]. Modern humans show a widened bandwidth compared with chimpanzees. Most of the SH fossils have the widened bandwidth, with the exception of AT94, which has an intermediate bandwidth between modern humans and chimpanzees.

<table>
<thead>
<tr>
<th>INDIVIDUAL</th>
<th>BANDWIDTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimpanzee</td>
<td>2365 Hz</td>
</tr>
<tr>
<td>Humans</td>
<td>3120 Hz</td>
</tr>
<tr>
<td>AT1907</td>
<td>3260 Hz</td>
</tr>
<tr>
<td>AT4103</td>
<td>3090 Hz</td>
</tr>
<tr>
<td>AT421</td>
<td>3130 Hz</td>
</tr>
<tr>
<td>AT84</td>
<td>2640 Hz</td>
</tr>
<tr>
<td>Cranium 5</td>
<td>3110 Hz</td>
</tr>
</tbody>
</table>

Table1. Bandwidth for 90% power of the different individuals.

5 Conclusion

Our analysis shows that the skeletal anatomy of the outer and middle ear in the SH hominins is compatible with a human-like sound power transmission pattern. Although much of the acoustic information in spoken language is concentrated in the region up to around 2.5 kHz (e.g. the first two formant frequencies of the vowels), the region between 2-4 kHz also contains relevant acoustic information in human speech. Indeed, the obtained bandwidth for humans and most of the SH fossils fit the bandwidth used in modern telephone communication (300Hz to 3.4kHz). This bandwidth is necessary to ensure the intelligibility of the communication because human speech contains much acoustic information in this band. The presence of a widened bandwidth of heightened sensitivity in the midrange frequencies in the SH hominins suggests that these fossil humans already possessed the anatomical features of the outer and middle ear that support the perception of human spoken language. Since the Sima de los Huesos hominins are not on the direct evolutionary line which gave rise to our own species, but form part of the Neandertal evolutionary lineage [47-49], it is conceivable that this condition was already present in the last common ancestor of modern humans and Neandertals. Thus, our results based on auditory capacities reinforce recent suggestions [33-34] for an ancient origin for human spoken language. Further research into the auditory capacities in even earlier hominin taxa may shed new light on precisely when this uniquely human feature first began to emerge during the course of our evolutionary history.

Acknowledgments

We wish to thank the Atapuerca Research and Excavation Team for their work in the field, especially to A. Aramburu, A. Bonmati, J. M. Carretero, A. Esquível, N. García and A. Gracia for their work at the Sima de los Huesos site. The excavations at the Atapuerca sites are funded by the Junta de Castilla y León. This research was supported by the Ministerio de Ciencia y Tecnologia of the Government of Spain, Project No. CGL2006-13532-C03-02.

References
