Title

Loss of air sacs improved hominin speech abilities

Corresponding Author

Bart de Boer

Author affiliation

Amsterdam Center for Language and Communication

University of Amsterdam

Spuistraat 210

1012 VT Amsterdam

b.g.deboer@uva.nl

Abstract

In this paper, the acoustic-perceptual effects of air sacs are investigated. Using an adaptive hearing experiment it is shown that air sacs reduce the perceptual effect of vowel-like articulations. Air sacs are a feature of the vocal tract of all great apes, except humans. Because presence or absence of air sacs is correlated with the anatomy of the hyoid bone, a probable minimal and maximal date of the loss of air sacs can be estimated from fossil hyoid bones: *Australopithecus afarensis* did still have air sacs about 3.3 Myr ago, while *Homo heidelbergensis,* some 600 kyr ago and *Homo neandethalensis* some 60 kyr ago, did no longer. The reduced distinctiveness of articulations produced with an air sac is in line with the hypothesis that air sacs were selected against because of the evolution of complex vocal communication. This relation between complex vocal communication and fossil evidence may help to get a firmer estimate of when speech first evolved.

Keywords:

Speech

Language

Air sacs

Hyoid bone

Vocal tract

Introduction

A major problem for students of the evolution of language is to date the emergence of complex vocal communication. There is very little fossil evidence for this, but *air sacs* have been proposed as a possible source of evidence (Fitch, 2000). Air sacs are large cavities, connected to the vocal tract above the vocal folds and are present in all apes, except humans (Hewitt et al., 2002). Given the prevalence of air sacs in apes, it is likely that the lack of air sacs in humans is a derived feature (Hewitt et al., 2002). A key question in human evolution is why air sacs disappeared, and whether selection for speech abilities played a role (Fitch, 2000; Fitch and Hauser, 2002; Hewitt et al., 2002; Nishimura et al., 2007). From fossil hyoid bones, a minimal and maximal date of the loss of air sacs can be deduced: *Australopithecus afarensis* did still have air sacs about 3.3 Myr ago (Alemseged et al., 2006), while *Homo heidelbergensis,* some 600 kyr ago (Martínez et al., 2008) and *Homo neandethalensis* some 60 kyr ago (Arensburg et al., 1989), did not. Little is known about the selective pressures leading to the loss of air sacs. Here, the acoustic and perceptual effects of air sacs are studied in order to contribute to a more complete evolutionary scenario. Using an adaptive hearing experiment it is shown that air sacs reduce the perceptual effect of vowel-like articulations. This finding is in line with the hypothesis that air sacs disappeared because of complex vocal communication.

Many mammals have air sacs (Frey et al., 2007), and at least five different anatomical configurations are recognized (Bartels, 1905; Frey et al., 2007). The air sacs found in apes are of the lateral ventricular type (Hewitt et al., 2002). These are relatively large, soft-walled cavities connected to the vocal tract through the lateral ventricles. They attach

to the vocal tract just above the vocal folds and below the false vocal folds, passing between the thyroid cartilage and the hyoid bone (illustrated in Fig. 1). In all homininae (chimpanzees, gorillas, bonobos, humans) the presence or absence of air sacs coincides with the presence or absence of the hyoid bulla (Avril, 1963; Kleinschmidt, 1938; Miller, 1941; Nishimura et al., 2007): a cup-shaped extension of the hyoid bone[1]. This may serve to keep open the connection between the vocal tract and the air sac.

There has been considerable debate about the exact function of air sacs. As they can become infected and form a serious health hazard (Lawson et al., 2006) they cannot be functionless, vestigial or accidental structures: there would be strong selective pressure against their presence. Most often a direct acoustic function is assumed (Avril, 1963; Fitch and Hauser, 1995; Haimoff, 1983; Schön, 1971). However, experiments with surgically altering air sacs of monkeys have given contradictory results (Gautier, 1971; Hilloowala and Lass, 1978) and an indirect role in vocalization, prevention of hyperventilation, has also been proposed (Hewitt et al., 2002). Lieberman, drawing a parallel with nasalized speech, has proposed that air sacs would make speech less clear (Lieberman, 2011, p. 333). Most of the debate on the (acoustic) role of air sacs has been speculative, as physical models of the acoustics of air sacs were not available (de Boer, 2008, 2009; Riede et al., 2008) until very recently. This paper investigates, using these new results, the acoustic and perceptual effect of air sacs on speech-like vocalizations.

---

[1] It has been noted that orangutans do have air sacs, but do not have a hyoid bulla. However, the siamang (*Symphalangus syndactylus*) also has an air sac and a hyoid bulla. The absence of a hyoid bulla therefore appears to be a derived characteristic in orangutans, and may be related to their unusual submental anatomy Brown, B., Ward, S.C., 1988. Basicranial and facial topology in *Pongo* and *Sivapithecus*, in: Schwartz, J.H. (Ed.), Orang-utan biology. Oxford University Press, Oxford, p. 247–260.

To estimate the acoustic effects, it is assumed that the relevant vocalizations in the hominin lineage were, as in most apes, produced by vocal cord vibration, and that their perceptual quality was mainly determined by the resonance frequencies of the vocal tract. Theoretical and experimental analysis (de Boer, 2008, 2009; Riede et al., 2008) shows that when an air sac of ape-like size is connected as a side branch to an ape-like vocal tract, the original resonance pattern undergoes three main changes. Firstly, a new resonance near the resonance frequency of the isolated air sac appears and acoustic energy at this frequency is radiated effectively through the air sac wall, thus increasing acoustic output at this frequency. Secondly, the original resonances of the vocal tract without air sac are shifted up and thirdly, they are shifted closer together (Fig. 2). In addition, at higher frequencies, extra resonances and anti-resonances appear, which correspond to the higher resonances and anti-resonances of the air sac. As the lowest resonance of an ape-like air sac is below the lowest resonances of an ape-like vocal tract, these qualitative effects are independent of the articulation being produced (de Boer, 2009).

A consequence of these modifications is that the vocalization has more resonances per kilohertz, and contains more energy at low frequencies. This is known to create an impression of larger size (Fitch and Hauser, 2002), as well as allowing the sound to travel further in dense foliage (Gautier, 1971; Hombert, 2010; Marten et al., 1977). Both of these may serve adaptive functions.

But what is the effect of an air sac if it is attached to a vocal tract that is used for complex vocal communication? Would it diminish the perceptual effect of articulatory actions? As suggested by Lieberman (2011), nasalization seems to provide a parallel: it is

also caused by attaching an acoustic side branch to the vocal tract (House and Stevens, 1956). It has been shown that the same articulatory actions have smaller perceptual effects when a vowel is nasalized (Wright, 1986). However, the acoustic effect of nasalization (Dang et al., 1994; House and Stevens, 1956) is different from that of an air sac, and therefore knowledge about the effect of nasalization is not directly applicable to air sacs. It is therefore necessary to do a perceptual experiment about the effect of air sacs.

Methods

To estimate the perceptual effects, and indirectly fitness due to communicative success, it needs to be understood how air sacs modify the relation between articulatory actions permitted by a given vocal tract anatomy, and the *discriminability* of the produced sounds. The effect of two deformations (Fig. 3) of a basic vocal tract were measured in a listening experiment with 22 human subjects (13 women, 9 men, all native speakers of Dutch). The tract used approximates the human vocal tract shape for producing [a] (approximately the first vowel of English "father") and consists of a narrow tube at the glottis followed by an equally long wide tube at the mouth. The first deformation approximates [ə] (the vowel of English "the") and consists of a tube of uniform diameter. The second deformation approximates [y] (the vowel of French "tu") and consists of a wide tube followed by an equally long narrow tube. These conditions ideally suit the purposes of this investigation as they represent simple articulatory actions (raising the

tongue and closing the mouth from an open configuration) and can be implemented with two-tube models.

Generation of stimuli

Stimuli were generated on the basis of two sets of three Perspex models (all stimuli without noise are available as on-line supporting material). It was decided to use a physical model rather than a simulated computer model, because existing speech synthesis models cannot deal satisfactorily with the three dimensional nature of air sacs (the model worked out in (de Boer, 2009) only works at low frequencies and the model in (Riede et al., 2008) approximates the air sac as a one-dimensional tube, rather than a three-dimensional cavity). The model for [a] consisted of a tube of 7.8 cm length and 1.6 cm diameter connected to a tube of 7.8 cm length and 3.6 cm diameter. The model for [ə] of a tube of 15.8 cm and a diameter of 2.6 cm. The [y]-model consisted of a tube of 8.0 cm length and 3.6 cm diameter connected to a tube of 7.8 cm length and 1.6 cm diameter. For the models with air sacs, a perpendicular side tube of 4.0 cm length was connected at 1.6 cm from the start of the tubes. The air sac itself was cylindrical, 8.7 cm long and 3.6 cm diameter. Thickness of all Perspex used was 0.2 cm (photos of the models are available as online supplementary material).

Single impulses were generated from these models by tapping the models on the end representing the glottis with the bare hand. The sound was then recorded with an Altec Lansing AHS30 electret microphone at 5 cm from the model's mouth and sampled at 44.1 KHz using a Soundmax Integrated Digital HD Audio PC card. These pulses were checked for consistency, and it was found that spectra thus generated correspond very

7

well with theoretical predictions(de Boer, 2009). Stimuli were generated directly by digitally convoluting a recorded pulse with an impulse train. The Impulse frequency descended from 160 Hz to 100 Hz in a 0.25 s interval in order to model an acceptable intonation contour. For models with air sacs, this procedure resulted in unrealistically low damping, and therefore pulses for the models with air sacs were damped exponentially by multiplying with $e^{-0.02t}$ (with t in milliseconds) corresponding to an increase in bandwidth of approximately 6 Hz. For signals that were generated without an air sac this multiplication had no perceptual effect. Signals were scaled to all have the same average power. They were presented in the middle of 0.5 s of white noise. All signals were recorded and generated at 44.1 KHz.

Testing of subjects

From a pilot study it was clear that the perceptual effect of the air sac was large enough (Cohen's $d > 1$) so that only a relatively small number of subjects would need to be tested. The 22 participants were university or conservatory students in the age range 18–30 years. Informed consent was obtained from the participants according to the university of Amsterdam's guidelines. They received a brief instruction about the experiment and their classification performance was tested in a soundproof room using a AKG hi-fi headphone and a Roland Edirol USB Audiocapture UA-25 D-A converter.

The signals were presented using an adaptive threshold estimation procedure based on a 2-unforced choice task (Kaernbach, 2001). Participants had the choice between two categories (either [a] and [ə] or [a] and [y]) and a "don't know" option. The

first signal presented always had a 15 dB signal-to-noise ratio. The ratio was changed, using a diminishing step size, depending on the response of the participant: for a good response the ratio was decreased by one times the step size, for a wrong response it was increased by three times the step size, for a "don't know response" it was increased by one times the step size. Using these multipliers of step sizes, the signal-to-noise-ratio will converge to 75% accuracy in classification. The step size decreased from 5 dB after 1 dB in 40 steps using an exponential decrease.

The 75%-correct classification threshold for each subject was determined by their average signal-to-noise-ratio over the last 15 trials.

Results

The distribution of minimal signal-to-noise ratios that allowed the participants to correctly classify signals is given in Fig. 4 for the different conditions. It can be observed that the condition *with* an air sac tolerates less noise. The difference between the conditions with and without air sac is significant for both stimulus pairs with $P = 3.0 \times 10^{-6}$, $t(21) = 6.30$ for [a] versus [ə] and $P = 4.8 \times 10^{-6}$, $t(21) = 6.09$ for [a] versus [y] according to the two-tailed paired t-test ($n = 22$ for all data sets). The mean of the difference in signal-to-noise ratio is 5.51 dB ($\sigma = 4.10$ dB) for [a] versus [ə] and 4.56 dB ($\sigma = 3.51$ dB) for [a] versus [y] (the difference between these values is not significant, $P = 0.38$, $t(21) = 0.89$, for the two-tailed paired t-test). Because the differences were calculated as the average over the last 15 trials (see methods section), their distribution was expected to be normal, and the normality of all distributions could indeed not be

rejected with the Lilliefors test ($P = 0.078$, $KS = 0.17$ for [a] vs. [ə], no air sac; $P = 0.78$, $KS = 0.10$ for [a] vs. [y], no air sac; $P = 0.095$, $KS = 0.17$ for [a] vs. [ə], with air sac; $P = 0.92$, $KS = 0.089$ for [a] vs. [y], with air sac).

Furthermore, without and with an air sac, the classification of [a] versus [ə] requires a slightly higher signal-to-noise ratio than the classification of [a] versus [y] (1.25 dB ($\sigma = 2.78$ dB) mean difference, $P = 0.047$, $t(21) = 2.11$ without air sac and 2.20 dB ($\sigma = 4.15$ dB) mean difference, $P = 0.021$, t(21) = 2.49 with air sac, for the two-tailed paired t-test). This is according to expectation, as it is expected that perceptual distance between [a] and [ə] to be smaller than that between [a] and [y] (Bladon and Lindblom, 1981) (the difference between the means is the same as the one in the previous paragraph, and therefore again not significant with $P = 0.38$, $t(21) = 0.89$).

Discussion and Conclusion

These results are in line with the hypothesis that an air sac reduces the perceptual distance between articulations. This reduction cannot be compensated without spending more articulatory effort. There are two reasons for this: [a] and [y] are the articulations whose lowest-frequency resonances are maximally far apart for the dimensions of the vocal tracts used. These resonances can therefore only be moved further apart by making a more exaggerated articulation (one which has larger differences in cross-sectional area between the front and back cavities). Secondly and more importantly, the extra low-frequency resonance and anti-resonance caused by the air sac are almost constant across

different articulations (de Boer, 2009). This makes the signals more similar in a way that cannot be compensated for by articulatory actions.

A potential confounding factor is that listeners might have more familiarity with the human-like stimuli *without* the air sac, as these are more like ordinary speech sounds. Some of the difference between the two types of stimuli might be explained by this familiarity. However, the stimuli with air sac were also interpreted by subjects as speech sounds, although they were heard as [a], [ə] and [ɪ], instead of [a], [ə] and [y]. [ɪ] is also a phoneme of Dutch, the native language of the participants. Furthermore, the difference in signal-to-noise ratio between the conditions without air sacs and those with air sacs was around 5dB. In comparison, the difference in signal-to-noise ratio between the [a]/[ə] contrast and the [a]/[y] contrast was 1–2dB in both conditions. This difference reflects the larger acoustic distance between [a] and [y] compared to the distance between [a] and [ə]. If unfamiliarity alone were the cause of the 5dB decrease of distinctiveness, one would not expect the relatively subtle 1–2dB difference to be preserved. Finally, classification of stimuli was 100% correct at the initial 15dB signal-to-noise level for all stimuli.

It is hard, and perhaps impossible to control completely for the effect that humans are more familiar with acoustic stimuli without air sacs than with stimuli with air sacs. An experiment, using a same-different paradigm, which is potentially less sensitive to familiarity is in preparation. However, as conveying linguistic information requires classifying signals, rather than just telling whether they are different, a classification experiment was more appropriate. The problem of familiarity remains for every experiment that is done to compare perception of modern articulations with that of reconstructed ancestral articulations. However, doing experiments, even though they may

have drawbacks, is preferable over only putting forward theoretical arguments. This paper is, as far as the author is aware, the first effort to experimentally test the perception of reconstructed ancestral human speech sounds.

If it is assumed then that the experimental results are due to lower distinctiveness of the stimuli and that communication is more successful when one is able to produce more distinctive signals, it follows that having an air sac attached to the vocal tract is an impediment for successful communication through speech. The loss of air sacs is consequently an indication for evolutionary pressure on complex vocal communication. The presence or absence of air sacs, through its relation with the shape of the hyoid bone, is one of the very few speech-related evolutionary innovations that can be tracked in the fossil record. The results presented here thus provide a much needed route for empirical evaluation of scenarios for the evolution of speech and language.

Acknowledgments

Figure Legends

Fig. 1 Chimpanzee air sac anatomy. Shown are the upper part of the trachea, the larynx (the crycoid and thyroid cartilages), the hyoid bone, the (right) ventricular appendix, and the connecting tube and cavity of the air sac.

Fig. 2 Acoustic effect of the air sac. Shown are the power spectra of the signal predicted for the articulation [a] without (grey line) and with (black line) an air sac. The most important changes are indicated with arrows: an extra peak near the resonance frequency of the air sac, and the shifting up and shifting closer together of the original resonances of the vocal tract without the air sac.

Fig 3. Schematic representation of models used for generating the stimuli. Shown are the three tubes without air sacs, as well as the tube for [a] with the air sac attached. The tracts for [ə] and [y] have been omitted, but are entirely analogous. The black bar labeled "2 cm" indicates the scale.

Fig 4. Boxplots of signal/noise thresholds. Horizontal lines indicate the medians, boxes the extent of the second and third quartile, and whiskers that of the whole data set. Note that the presence of an air sac reduces the amount of noise that subjects can tolerate in classifying the stimuli.

References

Alemseged, Z., Spoor, F., Kimbel, W.H., Bobe, R., Geraads, D., Reed, D., Wynn, J.G., 2006. A juvenile early hominin skeleton from Dikika, Ethiopia. Nature 443, 296–301.

Arensburg, B., Tillier, A.M., Vandermeersch, B., Duday, H., Schepartz, L.A., Rak, Y., 1989. A Middle Palaeolithic human hyoid bone. Nature 338, 758–760.

Avril, C., 1963. Kehlkopf und Kehlsack des Schimpansen, Pan troglodythes (Blumenbach 1799). (Mamalia, Primates, Pongidae). Gegenbaurs Morphol. Jahrb. 105, 75–129.

Bartels, P., 1905. Über die Nebenräume der Kehlkopfhöhle. Z. Morphol. Antropol. VIII, 11–61.

Bladon, R.A.W., Lindblom, B., 1981. Modeling the judgement of vowel quality differences. J. Acoust. Soc. Am. 69, 1414-1422.

Brown, B., Ward, S.C., 1988. Basicranial and facial topology in *Pongo* and *Sivapithecus*, in: Schwartz, J.H. (Ed.), Orang-utan biology. Oxford University Press, Oxford, p. 247–260.

Dang, J., Honda, K., Suzuki, H., 1994. Morphological and acoustical analysis of the nasal and paranasal cavities. J. Acoust. Soc. Am. 96, 2088–2100.

de Boer, B., 2008. The acoustic role of supralaryngeal air sacs. J. Acoust. Soc. Am. 123, 3732–3733.

de Boer, B., 2009. Acoustic analysis of primate air sacs and their effect on vocalization. J. Acoust. Soc. Am. 126, 3329–3343.

Fitch, W.T., 2000. The evolution of speech: a comparative review. Trends Cogn. Sci. 4, 258–267.

Fitch, W.T., Hauser, M.D., 1995. Vocal Production in Nonhuman Primates: Acoustics, Physiology, and Functional Constraints on "Honest" Advertisement. Am. J. Primatol. 37, 191–219.

Fitch, W.T., Hauser, M.D., 2002. Unpacking "Honesty": Vertebrate Vocal Production and the Evolution of Acoustic Signals, in: Simmons, A.M., Fay, R.R., Popper, A.N. (Eds.), Acoustic Communication. Springer, New York, p. 65–137.

Frey, R., Gebler, A., Fritsch, G., Nygrén, K., Weissengruber, G.E., 2007. Nordic rattle: the hoarse vocalization and the inflatable laryngeal air sac of reindeer (Rangifer tarandus). J. Anat. 210, 131-159.

Gautier, J.-P., 1971. Etude morphologique et fonctionelle des annexes extra-laryngées des cercopithecinae; Liaison avec les cris d'espacement. Biol. Gabon. 7, 229-267.

Haimoff, E.F., 1983. Occurence of anti-resonance in the song of siamang (Hylobates syndactylus). Am. J. Primatol. 5, 249-256.

Hewitt, G.P., MacLarnon, A., Jones, K.E., 2002. The Functions of Laryngeal Air Sacs in Primates: A New Hypothesis. Folia Primatol. 73, 70–94.

Hilloowala, R.A., Lass, N.J., 1978. Spectrographic Analysis of Laryngeal Air Sac Resonance in Rhesus Monkey. Am. J. Phys. Anthropol. 49, 129-132.

Hombert, J.-M., 2010. Ecological and Sexual Explanations for Larynx Lowering, in: Smith, A.D.M., Schouwstra, M., de Boer, B., Smith, K. (Eds.), The Evolution of Language (EVOLANG 8). World Scientific, Hackensack (NJ), p. 421–422.

House, A.S., Stevens, K.N., 1956. Analog Studies Of The Nasalization Of Vowels. J. Speech Hear. Disord. 21, 218-232.

Kaernbach, C., 2001. Adaptive threshold estimation with unforced choice tasks. Percept. Psychophys. 63, 1377–1388.

Kleinschmidt, A., 1938. Die Schlund- und Kehlorgane des Gorillas "Bobby" unter besonderer Berücksichtigung der gleichen Organe von Mensch und Orang. Ein Beitrag zur vergleichenden Anatomie des Kehlkopfes. Gegenbaurs Morphol. Jahrb. 81, 78-157.

Lawson, B., Garriga, R., Galdikas, B.M.F., 2006. Airsacculitis in fourteen juvenile southern Bornean orangutans (*Pongo pygmaeus wurmbii*). J. Med. Primatol. 35, 149-154.

Lieberman, D.E., 2011. The Evolution of the Human Head. Harvard University Press, Cambridge, MA.

Marten, K., Quine, D., Marler, P., 1977. Sound Transmission and Its Significance for Animal Vocalization: II. Tropical Forest Habitats. Behav. Ecol. Sociobiol. 2, 291–302.

Martínez, I., Arsuaga, J.-L., Quam, R., Carretero, J.-M., Gracia, A., Rodríguez, L., 2008. Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain). J. Hum. Evol. 54, 118–124.

Miller, R.S., 1941. The laryngeal air sacs of an infant and adult gorilla. Am. J. Anat. 69, 1–17.

Nishimura, T., Mikami, A., Suzuki, J., Matsuzawa, T., 2007. Development of the Laryngeal Air Sac in Chimpanzees. Int. J. Primatol. 28, 483-492.

Riede, T., Tokuda, I.T., Munger, J.B., Thomson, S.L., 2008. Mammalian laryngseal air sacs add variability to the vocal tract impedance: Physical and computational modeling. J. Acoust. Soc. Am. 124, 634–647.

Schön, M.A., 1971. The Anatomy of the Resonating Mechanism in Howling Monkeys. Folia Primatol. 15, 117-132.

Wright, J.T., 1986. The Behavior of Nasalized Vowels in the Perceptual Vowel Space, in: Ohala, J.J., Jaeger, J.J. (Eds.), Experimental Phonology. Academic Press, Inc., Orlando, pp. 45-67.
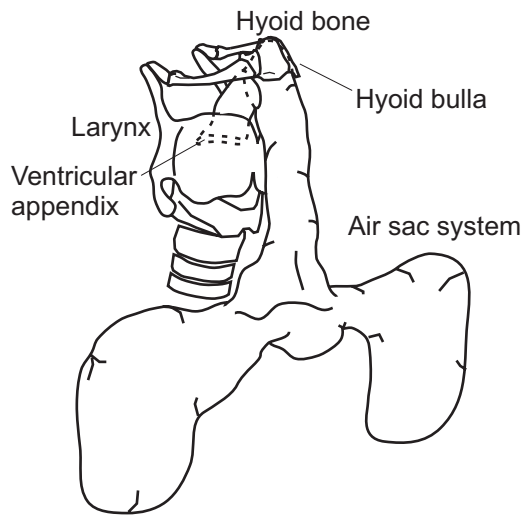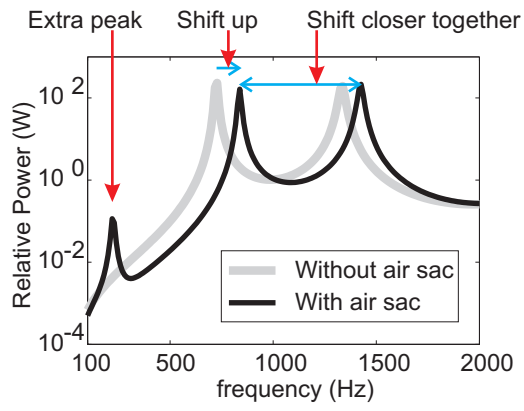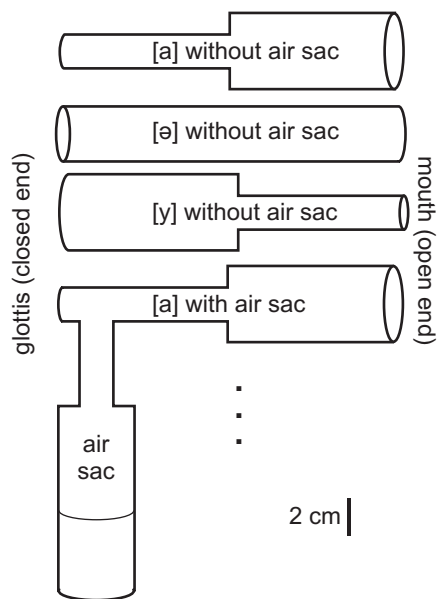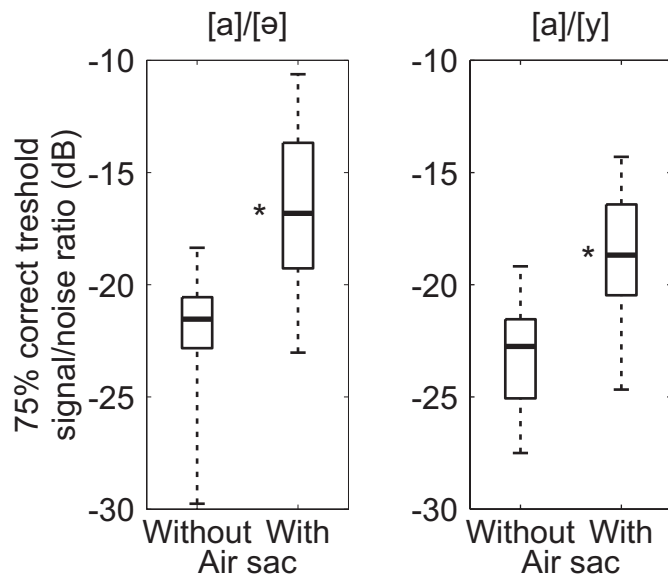
Figure 1.

Figure 2.

[a] without air sac

[ə] without air sac

[y] without air sac

[a] with air sac

glottis (closed end)

mouth (open end)

air sac

2 cm

Figure 3.

Figure 4.