



UltraFit: A speaker-friendly headset for ultrasound recordings in speech science

Lorenzo Spreafico¹, Michael Pucher², Anna Matosova¹

¹Free University of Bozen-Bolzano, Italy

²Acoustics Research Institute, Vienna, Austria

l.spreafico@unibz.it, michael.pucher@oeaw.ac.at, avosotam@inbox.ru

Abstract

UltraFit is a headset for Ultrasound Tongue Imaging (UTI) printed in Nylon; altogether, it weighs about 350 g. It was developed through an iterative process of rapid prototyping a proof of concept, asking for feedback from researchers and subjects of the experiments, and instantly incorporating changes based on their feedback into the design. We evaluated the UltraFit headset by recording a speaker using an optical marker tracking system that provides sub-millimeter tracking accuracy. We show that the overall error range of the headset movement for this speaker lies within 3mm with most errors lying in a 1-2mm range. This makes the headset potentially suitable for speech science applications. Furthermore, we analyze the superior usability of the headset compared to other existing designs and describe the headsets development process.

Index Terms: articulatory phonetics, ultrasound tongue imaging, systems for new applications.

1. Introduction

Ultrasound Tongue Imaging (UTI) is a non-invasive technique that allows researchers to observe the position, shape and movements of the tongue in real time during speech [1]. Although the UTI technique has made great strides forward in recent decades, stabilizing the ultrasound transducer under the speakers chin has remained a major challenge [2]. Many participants in experiments leave the data collection session due to the discomfort and pain they experience after a few minutes of recording. Since UTI is being increasingly used for clinical [3], instructional [4, 5] and technological [6, 7] applications that oblige operators to put less stress on the speaker than usually accepted in laboratory settings, this has become a significant problem. When collecting ultrasound tongue images, it is necessary to stabilize the ultrasound transducer under the speakers chin to avoid deviations in measurement data. A number of ideas have been proposed to solve this issue, for example by using mechanical systems of varying complexity [8, 9]; or pieces of software [10]. However, the most frequent solution to the transducer issue remains the use of a stabilizing headset. The most-used head-probe stabilization headset is probably the one designed, manufactured and distributed by Articulate Instruments [11]. Over the years, the system has been refined and produced in different shapes and materials, including polycarbonate to allow co-registering ultrasound and electromagnetic articulometry data [12]. The production of metallic headsets made of rigid aluminum and of non-metallic headsets made of polycarbonate is both cost and time consuming. Typically, head-probe stabilization headsets are made of more elements that need be cut, bent, milled, finished, glued and manually assembled. In this respect, the selective laser sintering (SLS) printing procedure [13] seems to be a more viable solution for

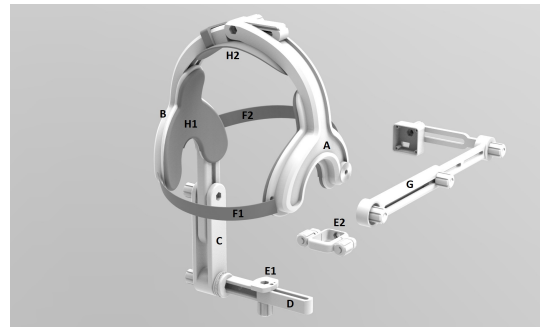


Figure 1: Exploded view of the UltraFit system.

making an easy-to-assemble three-dimensional object comprising a limited number of thermoplastic components. This additive manufacturing method permits molding functional shapes that fit better than bent aluminum sheets. Additionally, this technique allows for the implementation of a truss structure for all headset components, thus creating a headset that is both stiff and light.

2. Development of the headset

Thanks to a targeted review of head geometry studies on the anatomy of kids (over six years old) and adults from different world regions [14], as well as extensive prototyping and testing [15], we developed a SLS printed headset that could ensure fewer pain complaints than other stabilization headsets and still allow the collection of accurate UTI data. This headset, UltraFit, is an entirely new product; it is not simply an improvement on previous designs. UltraFit takes advantage of the different material used, the alternative manufacturing procedure and, most importantly, the opportunity to rapidly prototype a proof of concept, requesting feedback from researchers and experiment subjects and instantly incorporating the changes into the design. UltraFit is printed in Nylon, and altogether it weighs about 350 g. The system consists of four structural components (Figure 1A-D), a probe clamp (E1, E2), two Velcro straps (F1, F2), one optional camera bracket (G), comfort pads (H1, H2), and some plastic or metal bolts and nuts. Parts A and B constitute the rigid structure that hugs the contour of the head. Parts A and B are coupled using a sliding mechanism that allows for the width of the structure to be adjusted to fit different head sizes and shapes. All points of contact on the head mount are padded using die-cut foam cushions for increased fit and comfort. Two Velcro straps (F1, F2) help keep the head mount in place. Parts C-D constitute the transducer bracket. Arm C slides and tilts on guide B, allowing for the vertical and hori-

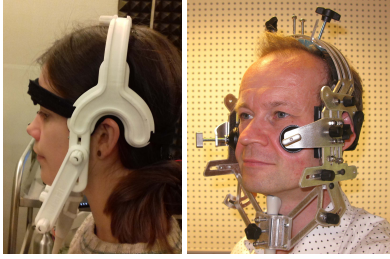


Figure 2: UltraFit headset (left) and Articulate Instruments Ltd headset (right; photo courtesy of Alan Wrench, <http://www.articulateinstruments.com/>).



Figure 3: Visual marker configuration (left). Video still from recordings of silence (middle) and speech (right).

zontal placement of the transducer to be adjusted. Arm D revolves on its axis, permitting the rotational adjustment of the transducer in the midsagittal plane. Arm D accommodates the transducer holder E and makes it possible to translate it from left to right and to align it with the tongue sulcus. The shape of the transducer holder can be either universal (E2) to accommodate various types of probes or custom fitted (as in the case of E1) to optimize the size and weight of the transducer bracket. Optionally, an adjustable camera bracket (G) can be installed on the head mount. Standard bolts and nuts are made of stainless steel, but plastic hardware can also be used to simultaneously record ultrasound and electromagnetic (EMA) data. The UltraFit headset has been designed with two requirements in mind: usability and accuracy.

3. Usability of the headset

The comfort and usability of the UltraFit headset in terms of subject and researcher satisfaction were a major design goal. With regard to the comfort of the headset for experimental subjects, we first considered the problem of physical pain in prolonged use due to the contact between the headset and the speakers head. Most headsets available on the market rest on the subjects head and potentially create pressure points on the frontal fontanelle (soft spot), the posterior fontanelle and the occipital bone, the phenosquamosal suture and the zygomaticotemporal suture. In contrast, UltraFit has been designed not to rest on sutures, nerves or external veins. This reduces the onset of external compression headaches (sometimes also known as football-headset headaches [16, 17] as well as so-called head-phone neuralgia [18]). With regard to the comfort of the experimental subjects, we also considered users acceptance of the headset, as far reaching acceptance barriers are prevalent which represent serious obstacles to technical solutions [19]. To this end, we produced a light-colored headset with organic shapes and rounded edges covered with soft and easily washable fab-

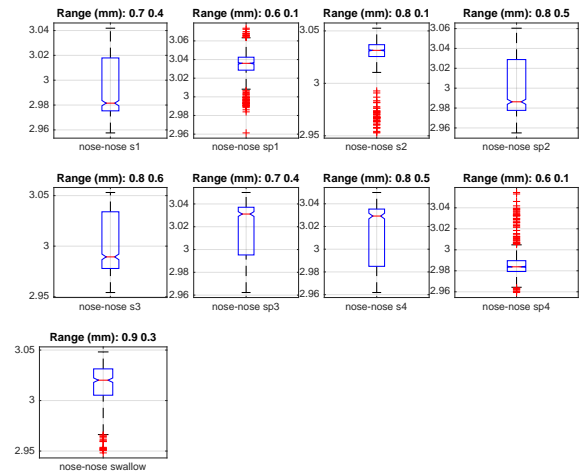


Figure 4: Distribution of Euclidean distance between nose marker 1 and 2 for silence 1 (s1) to silence 4 (s4), speech 1 (sp1) to speech 4 (sp4), and swallowing (sw).

rics that are not cold to the touch in order to make it more pleasant to use than the traditional sharp, cold, aluminum headsets. In terms of the headsets usability for the researchers, we designed UltraFit to reduce the movements necessary to fix it on the speakers head and position the probe under the speakers chin. In fact, the two operations can be performed in only a few steps that involve the placement of two Velcro strips and the tightening of a maximum of four screws. Compared to the number of operations required to set up the Articulate Instruments Ltd headset shown in Figure 2, the UltraFit system we developed is much more user friendly. Based on our laboratory experience, facilitating probe placement reduces experimental preparation time, which lessens the speakers suffering and makes he or she more likely to participate in a subsequent data collection session.

4. Accuracy of the headset

To evaluate the accuracy of the UltraFit headset, we recorded a male speaker reading a sequence of CVCV utterances. We recorded four sequences of 10 seconds each; they were always preceded by 10 seconds of silence. At the end of the four silence/speech sequences, we also recorded the speakers swallowing a sip of water.

Facial movement was recorded with a NaturalPoint OptiTrack Expression system using seven FLEX:V100R2 infrared cameras. In the past, we also used the visual optical marker tracking system for visual speech synthesis [20] and recorded a visio-articulatory speech database [21]. This system records the 3D position of reflective markers glued to the speakers face at 100 Hz. In our recordings, we glued markers to the speakers nose, the headset, and the ultrasound probe, as shown in Figure 3. Additionally, we used the four headband markers to remove head movement from the recordings. For the evaluation, we used the output of the system directly without applying any manual corrections.

By measuring the distance between different markers, we were able to measure the movement of the headset. When no movement of the headset is present, the distance between any two marker points should stay fixed with no variance. The distribution of distances between markers on the nose and the

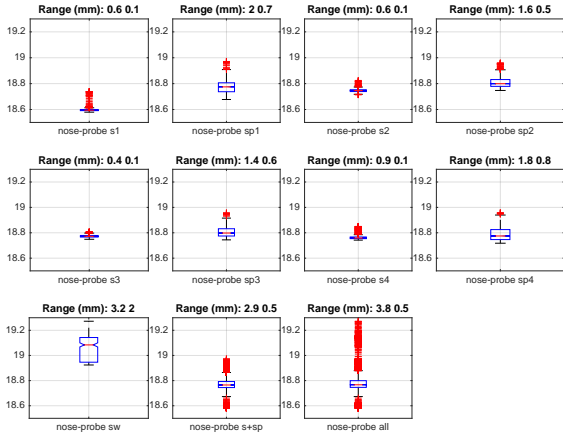


Figure 5: Distribution of Euclidean distance between nose marker 1 and probe marker 1 for silence 1 (s1) to silence 4 (s4), speech 1 (sp1) to speech 4 (sp4), swallowing (sw), silence and speech (s+sp), and all recordings.

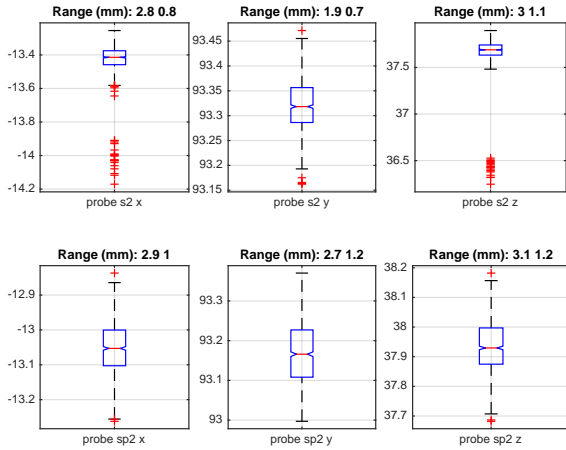


Figure 6: Distribution of individual coordinates (x, y, z) for nose marker 1 for silence 2 (s2) and speech 2 (sp2) after head movement removal.

probe is of special interest, because these show the movement of the probe relative to the speakers head. We expect the largest movement to take place at the probe.

To evaluate the accuracy of the optical marker tracking system itself, which the vendor claims is capable of measuring to the sub-millimeter level, we measured the distance between the two nose markers. This distribution is shown in Figure 4. The range in millimeter (mm) in the title of each subfigure is given for the 2nd to 97.5th percentile (first number) and for the 25th to 75th percentile of the data (second number). Outliers are shown as red crosses. The median is shown as red line, the box covers the 25th to 75th percentile and the whisker extends from the 2nd to the 97.5th percentile. As can be seen from Figure 4 the system does indeed record with sub-millimeter accuracy since the values all lie within a range of one millimeter. With this type of accuracy we can be confident to be able to measure the performance of the UltraFit headset.

Figure 5 shows the Euclidean distances between the 3D points nose marker 1 and probe marker 1 for the whole record-

ing session with the male speaker. This shows the error of the ultrasound headset during the recording session. The range in millimeters (mm) is again given for the 2nd to 97.5th percentile (first number) and for the 25th to 75th percentile of the data (second number). We can see that the distances expand after the first silence sequence (s1) from a median of 18.6 to a median of 18.8 and stay at a similar distance for the rest of the recording with another expansion when the final swallowing occurs. For the speech-silence sequence, all of the distances stay in the range of 2 millimeters, with a majority of distances being closer. This makes the headset suitable for speech science applications. More outliers are present when measuring all of the data (silence, speech, and swallowing).

Figure 6 shows the distributions for the individual coordinates (x, y, z). This shows the error of the ultrasound headset in the different dimensions. In order to measure the movement in the different coordinates, we have to remove the head movement first. This is done by using the four points of the headband; although, we did observe small movements of the headband during the recordings due to movements of the forehead. For head movement removal, we solve a system of linear equations that map the headband markers to a reference frame; then, the probe markers are transformed using this mapping. First we find X such that the current headband markers (A_c) multiplied by X give us the initial headband markers (B_i).

$$A_c X = B_i$$

Then we apply X to each row vector in homogeneous coordinates containing the probe marker coordinates p to get new probe marker coordinates \hat{p}

$$\hat{p} = pX.$$

As we can see from Figure 6 the achieved distribution shows a slightly larger variance than the measurements with the Euclidean distances. This comes from the fact that the head removal introduces some errors. In particular, the outliers in the silence recordings are due to movements of the headbands during the recordings.

Figure 6 also shows which coordinate exhibits the largest error. This is the z coordinate that points horizontally from the speaker's face, as shown in Figure 7. We see that the range of the 2nd to 97.5th percentile is maximally 3.1mm, not considering outliers that are most likely introduced by the head movement removal.

Figure 7 shows the distribution of all markers after head movement removal. The headband markers are fixed, and we can see the largest distribution in the probe markers.

5. Conclusion

The aim of the UltraFit design was to produce a headset that could be used for data collection with the ultrasound tongue imaging technique and could guarantee improved performance over already available systems. For advanced comfort of use, UltraFit has been designed not to rest on sutures, nerves, or external veins. To increase the acceptability of the headset, we decided to produce a light-colored headset with organic shapes and rounded edges that would not be cold to the touch.

The availability of such a headset would facilitate the extension of the UTI techniques field of use, making it more acceptable for application on demanding users, such as children, or for activities that require prolonged wearing of the item, such as silent speech interfaces.

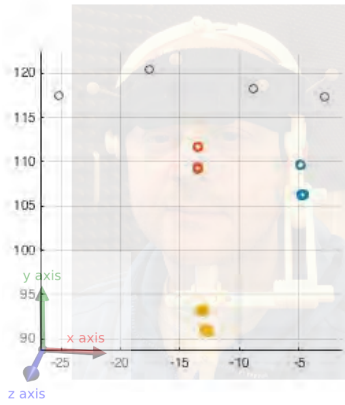


Figure 7: Distribution of markers in 3D space after head movement removal with an overlay of the picture of the speaker and the coordinate system.

To evaluate the accuracy of the headset, we recorded visual markers for a speaker on the speakers face and the headset and, thereby, measured the movement of the headset relative to the speakers head. These measurements showed that UltraFit has the potential to provide high accuracy recordings of tongue movements.

These preliminary results and the experimentation conducted thus far show that the realization of a headset characterized by better usability and high accuracy is possible. However, the limited real-use experience with the UltraFit system and the limited set of analyzed data on probe placement accuracy suggest a cautious interpretation of our results. More tests, including those sampling a greater number of speakers with different facial anatomies, are called for.

6. Acknowledgements

Thanks to Alan Wrench, Alessandro Vietti, Vincenzo Galata, and Anika Sedlaczek for helpful discussions and feedbacks with regard to designing the headset. This work was also supported by the Austrian Science Fund (FWF) project *DiÖ - Deutsch in Österreich* (I2539-G23).

7. References

- [1] M. Stone, "A guide to analyzing tongue motion from ultrasound images," *Clinical Linguistics and Phonetics*, vol. 19, no. 6-7, pp. 455–502, 2005.
- [2] L. Davidson and P. D. Decker, "Stabilization techniques for ultrasound imaging of speech articulations," *The Journal of the Acoustical Society of America*, vol. 2544, 2005.
- [3] J. Preston, M. Leece, and E. Maas, "Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia," *Neuroscience*, vol. 10, no. 240, 2016.
- [4] I. Wilson and B. Gick, "Ultrasound technology and second language acquisition research," in *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*, 2006, pp. 148–152.
- [5] S. Nakai, D. Beavan, E. Lawson, G. Leplatre, J. Scobbie, and J. Stuart-Smith, "Viewing speech in action: speech articulation videos in the public domain that demonstrate the sounds of the international phonetic alphabet (ipa)," *Innovation in language learning and teaching*, vol. 0, no. 0, 2016.
- [6] T. Hueber, E.-L. Benaroya, G. Chollet, B. Denby, G. reyfus, and M. Stone, "Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips," *Speech Communication*, vol. 52, no. 4, pp. 288–300, 2010.
- [7] D. Fabre, T. Hueber, X. Alameda-Pineda, and P. Badin, "Automatic animation of an articulatory tongue model from ultrasound images of the vocal tract," *Speech Communication*, vol. 93, no. 4, pp. 67–75, 2017.
- [8] M. Stone and E. Davis, "A head and transducer support system for making ultrasound images of tongue/jaw movement," *The Journal of the Acoustical Society of America*, vol. 98, no. 6, pp. 3107–3112, 1995.
- [9] D. Derrick, C. Best, and R. Fiasson, "Non-metallic ultrasound probe holder for co-collection and co-registration with ema," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, 2015.
- [10] D. H. Whalen, K. Iskarous, M. K. Tiede, D. J. Ostry, H. Lehnert-LeHouillier, E. Vatikiotis-Bateson, and D. S. Hailey, "The haskins optically corrected ultrasound system (hocus)," *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 3, pp. 543–553, 2005. [Online]. Available: [http://dx.doi.org/10.1044/1092-4388\(2005\)037](http://dx.doi.org/10.1044/1092-4388(2005)037)
- [11] J. M. Scobbie, A. A. Wrench, and M. L. V. D. Linden, "Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement," in *In Proceedings of the 8th Speech Production Workshop: Models and Data*, 2008, pp. 373–376.
- [12] F. Sigona, A. Stella, B. G. Fivela, F. Montagna, A. Maffezzoli, A. Wrench, and M. Grimaldi, "A new head - probe stabilization device for synchronized ultrasound and electromagnetic articulography recordings," <http://www.cril.unile.it/it/pr14.php>.
- [13] I. Gibson and D. Shi, "Material properties and fabrication parameters in selective laser sintering process," *Rapid Prototyping Journal*, vol. 3, no. 4, pp. 129–136, 1997.
- [14] A. R. Tilley and H. D. Associates, *The Measure of Man and Woman: Human Factors in Design, Revised Edition*. Wiley, 2001.
- [15] L. Spreafico, A. Matosova, A. Vietti, and V. Galat, "Two head-probe stabilization devices for speech research and applications," Poster presentation. Ultrafest VIII. Potsdam, October 4-6, 2017.
- [16] V. Krymchantowski, "Headaches due to external compression," *Current Pain and Headache Reports*, vol. 14, no. 4, pp. 321–324, 2010.
- [17] Z. Rahmani, A. Kochanek, J. J. Astrup, J. N. Poulsen, and P. Gazerani, "Helmet-induced headache among danish military personnel," *Scandinavian Journal of Public Health*, vol. 45, no. 8, pp. 818–823, 2017.
- [18] O. Bustos and A. K. Skelton, "More on headphone neuralgia," *New England Journal of Medicine*, vol. 334, no. 22, pp. 1480–1481, 1996, PMID: 8618601.
- [19] S. Gaul and M. Ziefle, "Smart home technologies: Insights into generation-specific acceptance motives," in *HCI and Usability for e-Inclusion. USAB 2009*, 2009, pp. 312–332.
- [20] D. Schabus, M. Pucher, and G. Hofer, "Joint audiovisual hidden semi-Markov model-based speech synthesis," *IEEE Journal on Selected Topics in Signal Processing*, vol. 8, no. 2, pp. 336–347, April 2014.
- [21] D. Schabus, M. Pucher, and P. Hoole, "The MMASCS multimodal annotated synchronous corpus of audio, video, facial motion and tongue motion data of normal, fast and slow speech," in *In Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, 2014, pp. 3411–3416.