AnTon: Investigating Human Language Learning from an Energetics Point of View

Robin Hofe **Roger K. Moore**
Department of Computer Science
University of Sheffield
Sheffield, UK
R.Hofe@sheffield.ac.uk

Abstract

AnTon is an animatronic model of a human tongue and vocal tract that will be used to investigate human speech and articulation strategies with respect to their energetics. AnTon will learn to use its articulators effectively by evaluating sensory feedback and the energy requirements associated with specific articulatory gestures.

1. Introduction

The use of robots to investigate human behaviour is common practice in research today. AnTon, the animatronic tongue and vocal tract model, will be used to clarify the energetics that underlie human speech production. Energy efficiency is widely assumed to be a constraint for speech production in humans, but it is extremely difficult to collect meaningful and precise data about energy use from humans.

2. Speech variation and articulatory effort

A large amount of variation occurs naturally in speech, causing problems for technical applications (Moore, 2007). It is of interest for the development of robust speech and language applications to find suitable models of such variation.

2.1 The H&H theory

The theory of hypo- and hyperspeech (H&H) (Lindblom, 1990) claims that speakers vary their speech output on a continuum from hypo- to hyperarticulated speech. The parameters that govern the process are information throughput and energy consumption. What makes the H&H theory interesting for technical applications is the fact that it does not concentrate on a specific type of variation (e.g. emotional speech), but that it describes a general principle that all types of variation should follow.

A drawback of the H&H theory is that it has not yet been formulated and evaluated in quantitative terms. AnTon is being developed to serve as a model of speech articulation whose energy use can be monitored in detail.

2.2 The Lombard reflex

There are many different sources of variation that cause alterations in the speech signal. A source of variation that is very well known in language behaviour science is the Lombard reflex (Junqua, 1996). It describes the natural reaction of human speakers to adapt their speech to noise in their environment. Both increases in loudness and augmentation of articulatory movements have been observed as reactions to noise levels that speakers experienced in controlled experiments.

There are a number of reasons to choose the Lombard reflex as a basis for experiments with AnTon:

- the source of variation is measurable in physical terms and is easy to control;
- adaptation strategies of human subjects are well documented and can be used to evaluate AnTon's behaviour;
- experiments can be carried out with single speech sounds instead of long, complicated utterances.

3. Design principle

The construction guideline of AnTon is that the functionality of the model should result solely from copying human anatomy. The human vocal tract serves a range of functions, of which speech is not the most important one. It is probably less energy efficient for speech production than it could be, were speech its primary function. In order to investigate human speech energetics, it is essential to implement the same topology as in the biological antetype.

The imitation of systems evolved by nature in technical applications is called 'biomimetics' (Bar-Cohen, 2006). AnTon will have access to auditory feedback as well as information about its articulatory effort.

The only other contemporary robot that simulates vocal articulators is the 'Waseda Talker' (Fukui et al., 2007). It differs from AnTon in its functional rather than biomimetic design. The tongue, for example, is actuated by pistons that change the shape of the tongue surface. This design method renders the Waseda Talker less useful for the investigation of human speech energetics.

4. State of development

AnTon currently consists of a soft silicone tongue attached to a movable jaw and a fixed hyoid bone. Tongue and jaw are actuated by servo motors, the muscles are represented by filaments that run parallel to muscle fibre orientation in the tongue (Takemoto, 2001) and along realistic lines of action for the jaw (Koolstra et al., 1990). There are currently eleven muscle filaments that represent four of the major tongue muscles and seven filaments for the four jaw muscles associated with speech movements (Vatikiotis-Bateson and Ostry, 1999).

The current model produces realistic jaw opening gestures, in which the condyle of the mandible rotates and slides forward in the jaw joint, as it does naturally in humans (Norton, 2007). The tongue presents a range of forward, backward, and downward movements; upward gestures are limited by the current lack of a movable hyoid bone and the muscles connecting the tongue to the soft palate.

The next step in the development process will be to complete the oral cavity by adding soft palate and pharynx models. The vocal tract will then be sealed off by a facial skin, and be ready for first vocalisation experiments. The long-term vision for AnTon includes artificial lungs and vocal chords to make sound production more realistic and include energy costs involved in breathing and air stream control.

5. Teaching AnTon to speak

AnTon will explore its own vocalisation abilities in three main stages: babbling, mimicking, and adaptation. During the babbling stage, AnTon will establish a motor-commands-to-sound mapping, including information about articulatory energetics.

In the mimicking stage, speech sounds will be presented to AnTon's auditory sytem. It will try to mimick those sounds within a specified tolerance, using an evolutionary algorithm. Articulatory effort will be an important factor in the cost function.

The last stage will be to add noise to the auditory feedback that AnTon receives. This should reduce the perceived closeness of the sounds it produces to the targets. New solutions will have to be found that lie within the specified range of tolerance. According to H&H theory, these should require more articulatory effort.

6. Conclusion

An animatronic model of a human tongue and vocal tract is currently being developed. It will be used to evaluate and quantify the relationship between speech variation and energetics. This paper has given a brief overview over the project's scope for the immediate future and how the model will be used in speech and language research.

The project's website is:
http://www.dcs.shef.ac.uk/~robin/anton/anton.html.

References

Bar-Cohen, Y. (2006). Biomimetics - using nature to inspire human innovation. *Bioinspiration & Biomimetics*, 1:P1–P12.

Fukui, K., Ishikawa, Y., Sawa, T., Shintaku, E., Honda, M., and Takanishi, A. (2007). New anthropomorphic talking robot having a three-dimensional articulation mechanism and improved pitch range. In *Proceedings of the 2007 IEEE Conference on Robotics and Automation, Roma, Italy.*

Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the lombard reflex. *Speech Communication*, 20:13–22.

Koolstra, J. H., von Eijden, T. M. G. J., van Spronsen, P. H., Weijs, W. A., and Valk, J. (1990). Computer-assisted estimation of lines of action of human masticatory muscles reconstructed in vivo by means of magnetic resonance imaging of parallel sections. *Archs oral Biol.*, 35(7):549–556.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the h&h theory. In Hardcastle and Marchal, (Eds.), *Speech Production and Speech Modelling*, pages 403–439. Kluwer Academic Publishers.

Moore, R. K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication*, 49:418–435.

Norton, N. S. (2007). *Netter's Head and Neck Anatomy for Dentistry.* Saunders Elsevier.

Takemoto, H. (2001). Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech, Language, and Hearing Research*, 44:95–107.

Vatikiotis-Bateson, E. and Ostry, D. J. (1999). Analysis and modeling of 3d jaw motion in speech and mastication. In *1999 IEEE International Conference on Systems, Man, and Cybernetics*, volume 2, pages 442–447.