



ECONOMIC AND SOCIAL RESEARCH COUNCIL

REFERENCE NUMBER
RES 000-23-1094
TITLE
Articulation and Coarticulation in the Lower Vocal Tract
INVESTIGATORS
Greg P. Kochanski and John Coleman
INSTITUTION
University of Oxford

Speech is a complex activity where muscles work in tight coordination. Speech is also a largely hidden activity: much of the relevant action happens inside the body where it cannot easily be studied.

In the past, linguists have been short of data and this has pushed the field towards its current heavily theoretical emphasis. Recently, techniques have become available to observe the vocal tract, but they have limits: some require X-ray exposure, some require sensors to be glued to the tongue, and none have the desired combination of speed, safety, and data quality.

A goal of our project was to test how MRI (Magnetic Resonance Imaging) can be used to observe speech. For MRI, the limiting factor is speed. Good quality images take seconds to produce, but speech is a rapid activity. People normally produce about four syllables per second, so an image must be captured in less than 1/10th of a second: most MRI techniques are too slow.

However, there are specialized MRI sequences that are designed to image a beating heart. These "gated" sequences work on the principle of the stroboscope. An MRI image is computed from about

200 pulses, each carrying a fraction of the information. In these gated sequences, pulses are collected over the course of 20 heart-beats, and then -- after collection -- sorted by the phase of the heart-beat in which they occurred.

We applied the idea to speech. We took short phrases like "enough sauce" spoken to a metronome and treated the tongue like a heart beat. The MRI scanner collected all the pulses that occurred near the "s" in any of the 20 repetitions of "sauce" and constructed an image showing how the "s" is pronounced. Likewise for the other sounds.

The heart is a very repeatable muscle. Each beat is much the same as the next and gated MRI sequences depend on that. This raised some unexpected but important questions: Could our volunteers speak all 20 repetitions precisely the same way? If not, the technique would fail. And, even if the technique could work, there were scientific risks: What kind of phrases could we use? If the technique turned out to be too restrictive it wouldn't be valuable for studying language. We had to find out if it would work with a broad selection of words and rhythmic patterns. Finally, we had to check that repetitive speech was similar to normal speech. Our results would only be valuable to the extent that speech in our experiment matched the real world.

We found that we could make the technique work. Most volunteers could speak any phrase repeatably if it was four syllables or shorter; longer than six was too difficult. More surprising, the rhythmic pattern of the phrases made no difference. We expected that people would find alternating strong and weak syllables easier than more complex patterns. Instead, it seems that people sometimes modify the dictionary rhythmic patterns for many of the phrases to fit the beat.

Another finding, confirming a single previous result, was that some people are dramatically better than others at speaking rhythmically to a metronome. This is a simple and (one might think) natural task, closely related to singing. Possibly this relates to people's attitudes toward music or poetry; the rhythmic beat may be stronger to some people than others.

The remainder of the project used the methods we developed to understand how the brain "drives" the tongue around. The goal was to build mathematical models of articulation in speech: to predict the shape of the tongue based on the phonological properties (i.e. dictionary pronunciation) of words.

More broadly, we set off to test linguistic theories. We built mathematical models of the tongue shape based on several theories

of how language is represented in the mind. To do this, tongue images were measured, which we compared to the models' predictions for the words that were spoken. If a model based on a certain linguistic theory can accurately predict the tongue shape, then that theory can be used with more confidence. If not, perhaps another theory will do a better job. (Our images emphasized the back and root of the tongue, areas that are particularly hard to study with other techniques.)

Because the linguistic theories are often not specified in enough detail for our purposes, we try all reasonable implementations of each theory ending with 93 models in all. By this approach, we are able to test a theoretical accounts thoroughly and can hope to eliminate theories. We can give theories the benefit of the doubt, but if none of the mathematical models that descend from a linguistic theory behave well, then the finger of suspicion would point at the theory more than at any individual model.

Our initial results show that two things. First, that the context of a speech sound is critically important for determining the tongue shape. The context is almost as important as the sound itself.

Second, we show that there is a variant of the standard Chomsky and Halle linguistic description that is substantially better than what is currently assumed. (Modern linguistics commonly describes sounds as a collection of features. Features are binary, either present or not, and the correspond roughly to motions of the tongue, so the [+HIGH] is either present or not [+HIGH] is present, the tongue is physically raised.) We show that both the preceding and following context is approximately equally important. This means that the tongue is not just lazily held in a position until a change is imperative, but that the brain is actively anticipating the next sounds and positioning the tongue in advance of the need.

We were able to exclude phoneme models and variants of Chomsky and Halle theories that do not include unspecified features and feature spreading. These theories that do not include context do a much less effective job of predicting the tongue shape. The results should affect linguistic theory, and they are also the first extensive competitive test of theories in the field.

Another, unexpected outcome was our work on speech timing, triggered by the discovery that MRI image quality was determined by the repeatability of the volunteer's speech. This work resulted in a journal paper (JASA) and a conference paper (ICPhS 2007). We also found that fairly complex voluntary muscle motions can be imaged with gated MRI sequences. This opens up the possibility of dynamically imaging motions of the hands and arms via MRI.

Although we achieved most of our project goals, we had our share of difficulties (listed in rough order of importance). These are mostly the run-of-the-mill problems that you expect on a research project.

Murphy's Law (or Sod's Law) operates. Things go wrong; things turn out to be more complex than expected. This kind of thing is really unavoidable in research, because when you're doing research you are (almost by definition) doing something that no one has ever done before. Consequently, one cannot depend upon standard procedures and best practices.

Planning and good management practices help, of course. By taking time to think and try to anticipate problems, one can certainly avoid some of them. However, it is a modern myth that a well-managed project runs smoothly. Or, perhaps it is a modern myth that one can effectively manage research.

Some of these problems were unanticipated (e.g. the first point). Others were anticipated as a possibility (e.g. the second and third points) but often there's nothing that can be done in advance to avoid the impact. Here's the list of things that had an impact of about a person-month or more:

- The unexpected discovery that MRI image quality depended primarily on subject behaviour rather than technical settings of the scanner. (This is discussed in section 7, below). This introduced an entire new task to the project.
- Some of the computations we planned to do (specifically Wolpert's model of motor control) do not scale well to the more complicated, large-scale modelling we needed to represent our data. Computing such models would require either massive supercomputer access or rather more cleverness than was available. This was a problem that only gradually became obvious after several person-months of work.
- Our named researcher was offered another post, and (quite reasonably) accepted it, and we had to conduct an unexpected search for an RA.
- Our newly hired researcher had delays in finishing her D. Phil. thesis.
- NHS Ethics review took longer than expected. The ethics review is very detailed and painstaking, really designed more for invasive experiments (for instance involving new surgical procedures, new drugs, that kind of thing). For this experiment, where the risks were small, some of the procedure was a waste of time.

At one point, the waste reached the level of a farce. The ethics review needs to allow for large projects spread across

many hospitals. That means it imagines a separate overall project leader (the “Chief Investigator”) along with group leaders at each hospital. By those standards, we had a small project, with one site and only a handful of people. So, I filled both roles, both Chief Investigator and the local group leader.

Sensibly enough, the ethics review needs to see the CV of the researchers. So, I put my CV in with the rest of the application and mailed it off. Two weeks later, we get a letter (not even an e-mail!) saying that they could not process our application because we hadn't included a CV for the local group leader (me).

So, what can one do? I sent the application in again, this time with two copies of my CV and a sarcastic letter. This time, it was accepted. This particular little farce wasn't too expensive: it only cost a few pounds of postage and a few person-hours, but it delayed the actual experiments for two weeks and protected absolutely nobody.

- Our department I.T. Officer left to become an airline pilot. This led to a 3-month gap where computer support was spotty.
- We needed to install cabling in the MRI machine to synchronize it with stimuli and allow audio recordings. We found that we would only be allowed to do so when the machine was shut down for its quarterly service day. This led to some delays of the initial pilot experiments which led to some inefficiencies because data was available later than planned.