

Time travelling to the mother tongue

July 19 2016



Spectrogram showing the shape of the sound of a word. Credit: John Aston

The sounds of languages that died thousands of years ago have been brought to life again through technology that uses statistics in a revolutionary new way.

No matter whether you speak English or Urdu, Waloon or Waziri, Portuguese or Persian, the roots of your <u>language</u> are the same. Proto-Indo-European (PIE) is the mother tongue – shared by several hundred contemporary languages, as well as many now extinct, and spoken by people who lived from about 6,000 to 3,500 BC on the steppes to the



north of the Caspian Sea.

They left no written texts and although historical linguists have, since the 19th century, painstakingly reconstructed the language from daughter languages, the question of how it actually sounded was assumed to be permanently out of reach.

Now, researchers at the Universities of Cambridge and Oxford have developed a sound-based method to move back through the family tree of languages that stem from PIE. They can simulate how certain words would have sounded when they were spoken 8,000 years ago.

Remarkably, at the heart of the technology is the statistics of shape.

"Sounds have shape," explains Professor John Aston, from Cambridge's Statistical Laboratory. "As a word is uttered it vibrates air, and the shape of this soundwave can be measured and turned into a series of numbers. Once we have these stats, and the stats of another spoken word, we can start asking how similar they are and what it would take to shift from one to another."

A word said in a certain language will have a different shape to the same word in another language, or an earlier language. The researchers can shift from one shape to another through a series of small changes in the statistics. "It's more than an averaging process, it's a continuum from one sound to the other," adds Aston, who is funded by the Engineering and Physical Sciences Research Council (EPSRC). "At each stage, we can turn the shape back into sound to hear how the word has changed."

Rather than reconstructing written forms of ancient words, the researchers triangulate backwards from contemporary and archival audio recordings to regenerate audible spoken forms from earlier points in the evolutionary tree. Using a relatively new field of shape-based



mathematics, the researchers take the soundwave and visualise it as a spectrogram – basically an undulating three-dimensional surface that represents the shape of that sound – and then reshape the spectrogram along a trajectory 'signposted' by known sounds.

While Aston leads the team of statistician 'shape-shifters' in Cambridge, the acoustic-phonetic and linguistic expertise is provided by Professor John Coleman's group in Oxford.

The researchers are working on the words for numbers as these have the same meaning in any language. The longest path of development simulated so far goes backwards 8,000 years from English one to its PIE ancestor oinos, and likewise for other numerals. They have also 'gone forwards' from the PIE penkwe to the modern Greek pente, modern Welsh pimp and modern English five, as well as simulating change from Modern English to Anglo-Saxon (or vice versa), and from Modern Romance languages back to Latin.

"We've explicitly focused on reproducing sound changes and etymologies that the established analyses already suggest, rather than seeking to overturn them," says Coleman, whose research was funded by the Arts and Humanities Research Council.

They have discovered words that appear to correctly 'fall out' of the continuum. "It's pleasing, not because it overturns the received wisdom, but because it encourages us that we are getting something right, some of the time at least. And along the way there have also been a few surprises!" The method sometimes follows paths that do not seem to be etymologically correct, demonstrating that the method is scientifically testable and pointing to areas in which refinements are needed.





Spectrograms showing how the shape of the sound of a word in one language can be morphed into the sound of the same word in another language. Credit: John Aston

Remarkably, because the statistics describe the sound of an individual saying the word, the researchers are able to keep the characteristics of pitch and delivery the same. They can effectively turn the word spoken by someone in one language into what it would sound like if they were speaking fluently in another.

They can also extrapolate into the future, although with caveats, as



Coleman describes: "If you just extrapolate linearly, you'll reach a point at which the sound change hits the limit of what is a humanly reasonable sound. This has happened in some languages in the past with certain vowel sounds. But if you asked me what English will sound like in 300 years, my educated guess is that it will be hardly any different from today!"

For the team, the excitement of the research includes unearthing some gems of archival recordings of various languages that had been given up for dead, including an Old Prussian word last spoken by people in the early 1700s but 'borrowed' into Low Prussian and discovered in a German audio archive.

Their work has applications in automatic translation and film dubbing, as well as medical imaging (see panel), but the principal aim is for the technology to be used alongside traditional methods used by historical linguists to understand the process of language change over thousands of years.

"From my point of view, it's amazing that we can turn exciting yet highly abstract statistical theory into something that really helps explain the roots of modern language," says Aston.

"Now that we've developed many of the necessary technical methods for realising the extraordinary ambition of hearing ancient sounds once more," adds Coleman, "these early successes are opening up a wide range of new questions, one of the central being how far back in time can we really go?"

Medical imaging reshaped

The statistics of shape are not just being used to show how different languages relate to each – they are also being used to improve the



analysis of medical images.

Just as soundwaves have a shape that can be analysed using statistics, so do the patterns of neurons interacting with each other or the dimensions of the surface of a tumour. Now a new research Centre will develop tools that use the mathematics of the shapes found in medical images to improve diagnosis, prognosis and treatment planning for patients.

The EPSRC Centre for Mathematical and Statistical Analysis of Multimodal Clinical Imaging, one of five 'maths' centres recently funded by £10 million from EPSRC, is co-led by Aston and Dr Carola-Bibiane Schönlieb from the Department of Applied Mathematics and Theoretical Physics in Cambridge.

"The new methodologies will allow clinical medicine to move beyond one person reading single scans, to automated systems capable of analysing populations of images," explains Schönlieb. "As a result, clinicians will have far greater scope to ask complex questions of the medical image."

It's already possible to extract statistical information from an image of a patient's thigh bone, turn the data into a template for comparison with those from other people in the population, and then ask whether a particular shape of bone is more prone to being broken than others in the elderly.

Most organ scans split the image into many elements, which are then analysed voxel by voxel. "But complex structures like the heart and the brain should be analysed holistically," explains Dr James Rudd, from the Department of Medicine, who leads the clinical interaction with the Centre. "The tools we are developing will enable the analysis of organs like the brain as single objects with millions of connections."



More information: Audio demonstrations are available here: <u>www.phon.ox.ac.uk/jcoleman/anc ... nt-sounds-audio.html</u>

Provided by University of Cambridge

Citation: Time travelling to the mother tongue (2016, July 19) retrieved 10 May 2024 from https://phys.org/news/2016-07-mother-tongue.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.