<table>
<thead>
<tr>
<th>Title</th>
<th>Testing AutoTrace: A Machine-learning Approach to Automated Tongue Contour Data Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Sung, JH; Berry, J; Cooper, M; Hahn-Powell, G; Archangeli, DB</td>
</tr>
<tr>
<td>Citation</td>
<td>The 6th Ultrafest Meeting, Edinburgh, UK., 6-8 November 2013. In the Programme and Abstract booklet of the 6th Ultrafest Meeting, 2013, p. 9-10</td>
</tr>
<tr>
<td>Issued Date</td>
<td>2013</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/202123">http://hdl.handle.net/10722/202123</a></td>
</tr>
<tr>
<td>Rights</td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Testing AutoTrace: A Machine-learning Approach to Automated Tongue Contour Data Extraction

Jae-Hyun Sung\textsuperscript{1}, Jeff Berry\textsuperscript{2}, Marissa Cooper\textsuperscript{1}, Gustave Hahn-Powell\textsuperscript{1}, and Diana Archangeli\textsuperscript{3}

\textsuperscript{1}Department of Linguistics, University of Arizona, USA
\textsuperscript{2}InsideSales.com
\textsuperscript{3}Linguistics, University of Hong Kong, HK

Articulatory imaging is important for analyzing the rules of speech and can be utilized for many purposes such as analyzing sounds in different dialects, learning a second language, and speech therapy (Archangeli and Mielke (2005), Adler-Bock et al. (2007), Gick et al. (2008), Scobbie et al. (2008)). When analyzing phonological data, researchers can implement various experimental methods concerning articulation - EMMA, MRI, Palatography, and ultrasound. Although ultrasound is inexpensive, non-toxic, and portable, it does have a significant drawback. After the data is collected, someone must then trace the tongue surface contours, which creates a bottleneck for analyzing the results. Several different approaches to this problem have been proposed (Li et al. (2005), Fasel and Berry (2010), Tang et al. (2012)), with promising results. In this paper, we analyze the performance of the Deep Neural Network approach of Fasel and Berry (2010) at scale, and announce an open source project to further develop this method, named AutoTrace.

AutoTrace automates the process of extracting the data from ultrasound images, greatly reducing the amount of time necessary for tracing images, as shown in Berry et al. (2012) on a small data set. This paper reports on our tests of the efficacy of AutoTrace on a much larger data set consisting of approximately 40,600 ultrasound images taken from Harvard sentences read by 12 American English speakers. This ensured a wide variety of tongue shapes, due both to different speakers and to different types of sounds. For training data sets, we selected a combination of most and least diverse images based on their deviation from pixel averages, using the heuristic proposed by Berry (2012).

AutoTrace used training data sets of different sizes to learn networks. Each network was tested against the same set of 100 randomly selected images. These traces were hand-corrected by human expert tracers, and each network was retrained. The automatically traced contours were compared to traces made by human experts to gauge how well the program performed.

Four separate tests are considered:

a. Most diverse images only vs. most + least diverse images: the combination most & least gave better results.

b. \( n \) images vs. \( n \) images (ranging from 250 to 1056 images) taken at intervals of \( y \): the larger the training set, the better the results.

c. No retraining vs. retraining: retraining produced a “41% decrease in the number of images needing hand correction” (Berry, 2012, p. 50).

d. \( n \) images from the most diverse set vs. \( r \) randomly selected images from the whole (\( n = r \)): the \( n \) most diverse images performed better.

Currently, the comparison of two human expert tracers shows a pixel-by-pixel average difference of 2.467 pixels per image. Comparison of machine vs. human tracers shows a pixel-by-pixel average difference of 5.656 pixels per image. We are currently examining the types of errors AutoTrace is making to see whether we can improve the training technology for better results.
References


Scobbie, J. M., Stuart-Smith, J., and Lawson, E. 2008. Looking variation and change in the mouth: developing the sociolinguistic potential of ultrasound tongue imaging. Queen Margaret University.