4pSC5

Computational modeling of human isolated auditory word recognition using DIANA

UNIVERSITY OF ALBERTA
DEPARTMENT OF LINGUISTICS

Filip Nenadić, Louis ten Bosch, and Benjamin V. Tucker

Alberta Phonetics Laboratory

Introduction

DIANA
An end-to-end computational model of spoken word processing, see **Figure 1** (ten Bosch, Ernestus, & Boves, *in preparation*)

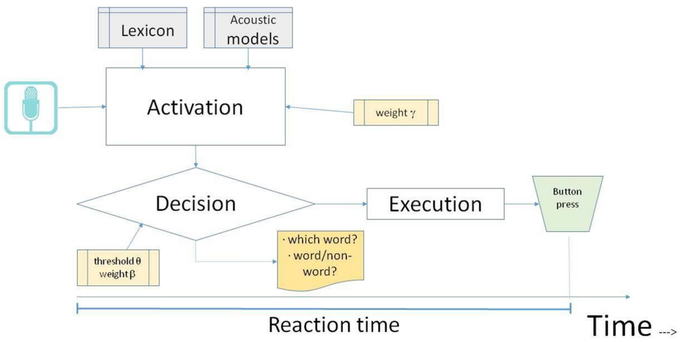

Figure 1. DIANA uses actual speech signal as input and outputs word or not word decisions and response time estimates.

DIANA is an activation and competition model:
- Acoustic input is converted into Mel-frequency cepstral coefficients, which preserve tonotopical information
- Subword units (currently phone-like) are represented by Gaussian mixture models
- Activation calculated by matching feature vectors of 10ms segments using a Bayesian framework

MASSIVE AUDITORY LEXICAL DECISION (MALD)
A database of word and pseudoword audio recordings and participant responses in an auditory lexical decision task (Tucker et al, *under revision*)

Recordings from a single male speaker
- 26,793 words organized into 67 lists of 400 items
- 9,592 pseudowords organized into 24 lists of 400 items

A single session consists of one word and one pseudoword list (800 items total)

Responses from 231 monolingual native listeners of English from a total of 284 separate sessions

PRESENT STUDY
Use MALD to create the first North American English model of spoken word processing using DIANA

Acoustic model training and testing

ACOUSTIC MODEL TRAINING WAS PERFORMED USING ASR-BASED TRAINING IN THREE STEPS:

Initial models trained on two unpublished spontaneous speech corpora (9 hours of recordings from 27 speakers)

Models were corrected to account for silent pauses, and the number of Gaussian mixtures per subword unit state was increased to 32

Speaker adaptation performed in 400-word increments up to 8000 words (up to 20 MALD lists; under 4 minutes each)

PERFORMANCE TESTED ON THREE LISTS WITH THE ENTIRE MALD WORD LIST AS THE LEXICON:

In a free word recognition task (the model adapted on 10 lists was used for model simulations)

By observing the top competitors (N-best lists), and competition as the signal unfolds

In a lexical decision simulation which compares activation of words to phonotactically licensed sequences

By comparing response latency estimates to actual participant response latencies (4 or 5 participants per list)

Results and discussion

Figure 2. Free word recognition accuracy in the three lists. Further analyses used the model adapted on 10 MALD lists.

Figure 3. Difference between the worst and other competitors at each time step.

Target word	Competitor	Activation
BROWSE (correct)	BROWSE	-2,890.86
	BROWS	-2,890.86
	BROWNS	-2,938.98
	ROUNDS	-2,941.75
ASSURED (incorrect)	USHERED	-4,475.29
	ASSURED	-4,485.90
	ISSUED	-4,522.81
	PRESSURED	-4,549.67

Table 1. Top competitors for a correctly and an incorrectly recognized word. In both cases, top competitors are similar to the target word.

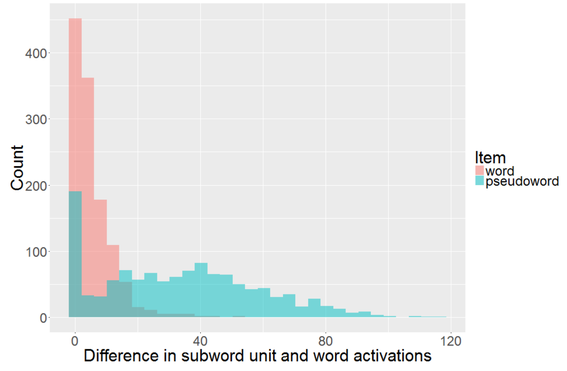

Figure 4. Difference in subword unit free loop activation and word activation divided by the total number of units. In a balanced response regime, error rates are between 17% and 21%.

DISCUSSION
- Acoustic models need improving (see **Figure 1**, **Figure 4**)
- Even with current models, between-word competition successfully simulated (see **Table 1**, **Figure 3**)
- Correlation between DIANA estimated RT and average participant logged RT was 0.18 for list 67, 0.29 for list 66, and -0.49 for list 65

FUTURE DIRECTIONS
- Add leading and trailing silence to MALD items to prevent initial stop recognition issues
- Increase speaker adaptation training set size
- Test on more lists
- Vary DIANA parameters to increase fit (e.g. the difference between top candidates needed to select the winner)

ten Bosch, L., Ernestus, M., and Boves, L. (in preparation). DIANA: An end-to-end computational model of human speech processing.
Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić F., and Sims, M. (submitted). Massive auditory lexical decision: Toward reliable, generalizable speech research.

174th meeting of the Acoustical Society of America (New Orleans, LA) December 2017

contact: nenadic@ualberta.ca

This project funded by the Social Sciences and Humanities Research Council: Grant #435-2014-0678