

The process of initiating speech  
and  
The search for good analysis tools

Pertti Palo

10 Mar 2025

## Land acknowledgement

I would like to respectfully acknowledge that we are located on Treaty 6 territory, a traditional gathering place for diverse Indigenous peoples including the Cree, Blackfoot, Métis, Nakota Sioux, Iroquois, Dene, Ojibway/Saulteaux/Anishinaabe, Inuit, and many others. I am a very recent uninvited guest and I have only started learning about their histories, languages, and cultures and of their historical and continued contribution to our community.

# Outline

- ▶ Introductions - Me and the topic
- ▶ Some naming experiments
- ▶ Some principles
- ▶ The tool: Pixel Difference (PD)
- ▶ Some quick results
- ▶ Conclusions

You can find these slides at:

<https://github.com/giuthas-talks/McEwan2025/>

# Who's this Pertti Palo?



- ▶ I grew up in Finland, became me in Scotland and have since worked in Indiana and here in Edmonton.
- ▶ I have a couple of degrees in engineering and a PhD in Phonetics.
- ▶ I also have a life in folk music, folk dancing, oral storytelling, wandering (hiking and long distance skiing), role-playing games, crafts (knitting, terrain crafting, and other things).

# The topic

- ▶ In my thesis I concentrated on timing of utterance onset in both acoustics and articulation (Palo 2019).
- ▶ The data was high-speed tongue ultrasound from a delayed naming experiment – specifically one using the Rastle instructions (Rastle et al. 2005).
- ▶ For data analysis I need some new tools, but lets first talk about the data itself.

Classical	Stimulus (word) perception	Lexical etc processing	Movement initiation	Movement	Acoustic speech
Delayed	Lexical etc processing	Stimulus (beep) perception	Movement initiation	Movement	Acoustic speech

# Some naming experiments

## Demo of classical naming

- ▶ Let's try some versions of naming experiments.
- ▶ After the slide changes read out loud the word on the next slide as soon as you can.

red

## Demo of classical naming

- ▶ Let's try that a second time.

green

## Demo of delayed naming

- ▶ After the slide changes wait for me to snap my fingers.
- ▶ After you hear the finger snap, read the word out loud as soon as you can.

orange

## Demo of delayed naming

- ▶ Rinse and repeat

purple

## Demo of delayed naming with Rastle instructions

- ▶ After the slide changes wait **at rest** for me to snap my fingers.
- ▶ After you hear the finger snap, read the word out loud as soon as you can.

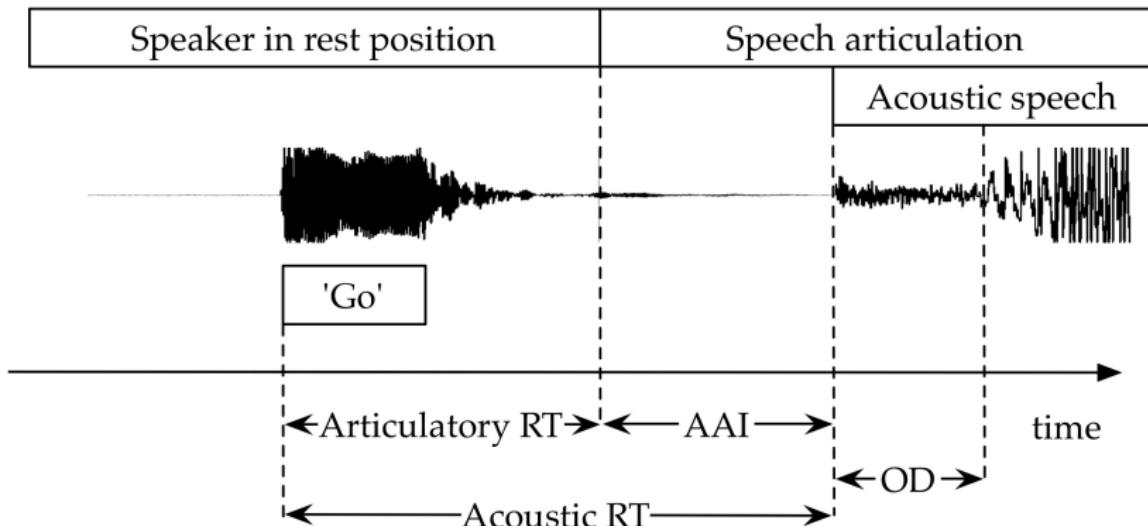
yellow

## Demo of delayed naming with Rastle instructions

- ▶ Once more

blue

# What happened in the last one?



- ▶ AAI = Articulatory onset to Acoustic onset Interval
- ▶ OD = Onset (or obstruent) Duration

## So what is the data actually like?

- ▶ Now imagine that final experiment was recorded with tongue ultrasound imaging.
- ▶ And our job was to find the articulatory onset in the resulting greyscale videos.
- ▶ Here, let's try it. [external slide set coming up]

## And why new tools?

- ▶ When trying to identify movement onset in greyscale videos with a lot of speckle 'noise', it doesn't take long to grow a desire for an easier way.
- ▶ The speckle 'noise' maybe caused by a number of factors including bubbles in the acoustic gel between the chin and the probe, noise sources in the equipment, and more interestingly changes in internal structures of tissues – such as muscle fibres tensing and relaxing.

# Some principles

## Some principles I

- ▶ When we are holding a hammer, everything looks like a nail.
- ▶ Every tool is a hammer...
  - ▶ ...at least until the first time we use it as a hammer...
  - ▶ ...which may lead to the item no longer being a tool at all or even break the nail (if it was a nail).

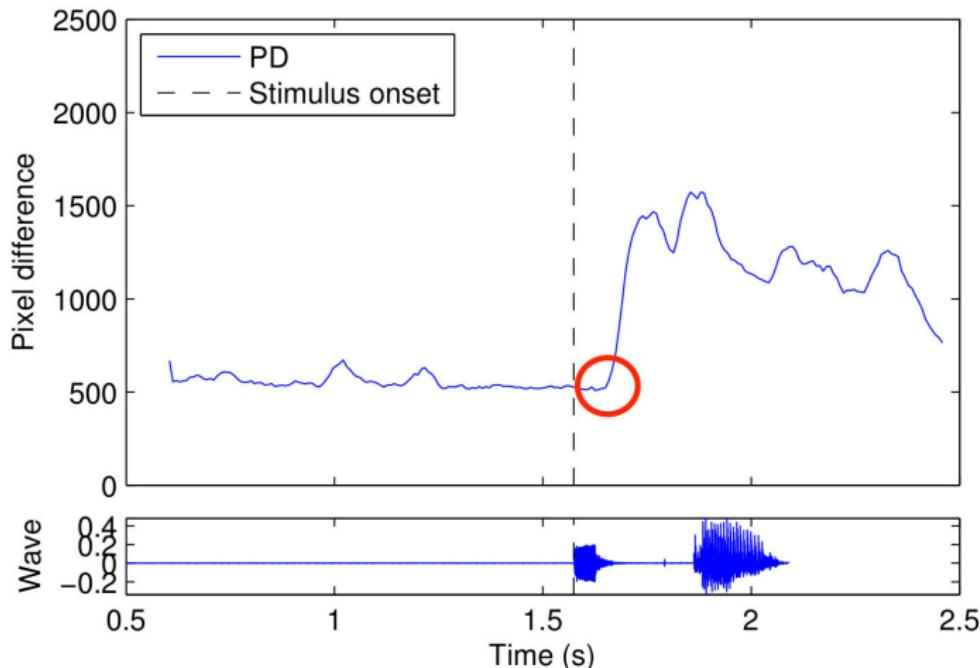
## Some principles II

- ▶ The first thing to do is to figure out if the tools we have match the question and data we have.
- ▶ And then get the right tool for the right data, if we aren't lucky to begin with.
- ▶ Keeping in mind that 'good' is usually better than 'best'.

## The tool: Pixel Difference (PD)

# Pixel Difference (PD)

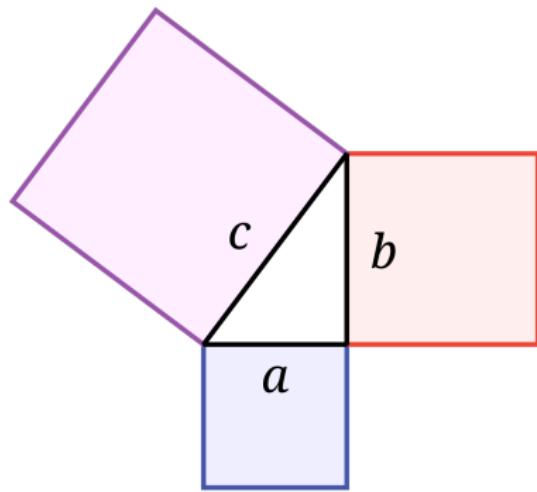
- ▶ The first tool out of the box was manual video segmentation.
- ▶ The second tool out of the box happened to work much better.



## Pixel Difference (PD): Background

- ▶ The analysis methods presented here are similar to methods developed by
  - ▶ McMillan and Corley (2010) and Drake et al. (2013) who used Euclidean distance on ultrasound frames and
  - ▶ Raeesy et al. (2011) who used a similar method on MRI data.
- ▶ The way I have used it, it is actually just the Pythagorean theorem applied in a space with a lot more dimensions than 2.

# Pixel Difference (PD): Pythagoras (wasn't first)



Pythagorean Theorem:

$$a^2 + b^2 = c^2$$

(Euclid 2006)

By en:User:Wapcaplet - Transwikied from en:. Originally created by en:User:Michael Hardy, then scaled, with colour and labels being added by en:User:Wapcaplet, transformed in svg format by fr:Utilisateur:Steff, changed colors and font by de:Leo2004, CC BY-SA 3.0,

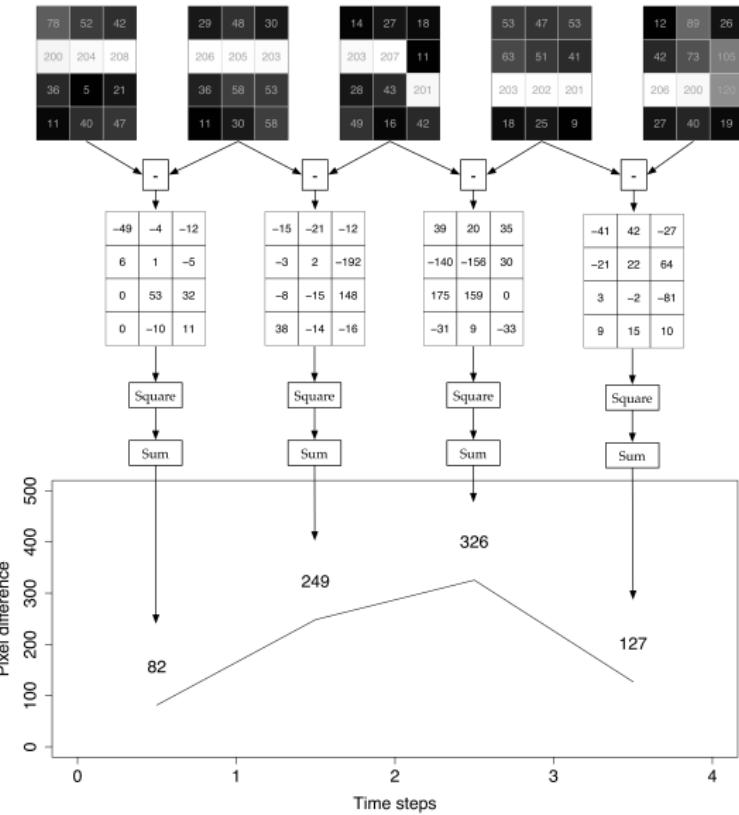
<https://commons.wikimedia.org/w/index.php?curid=640875>

## Pixel Difference (PD): The maths

$$I_2(t + 0.5) = \sqrt{\sum_{i,j} (x(i,j, t + 1) - x(i,j, t))^2}$$

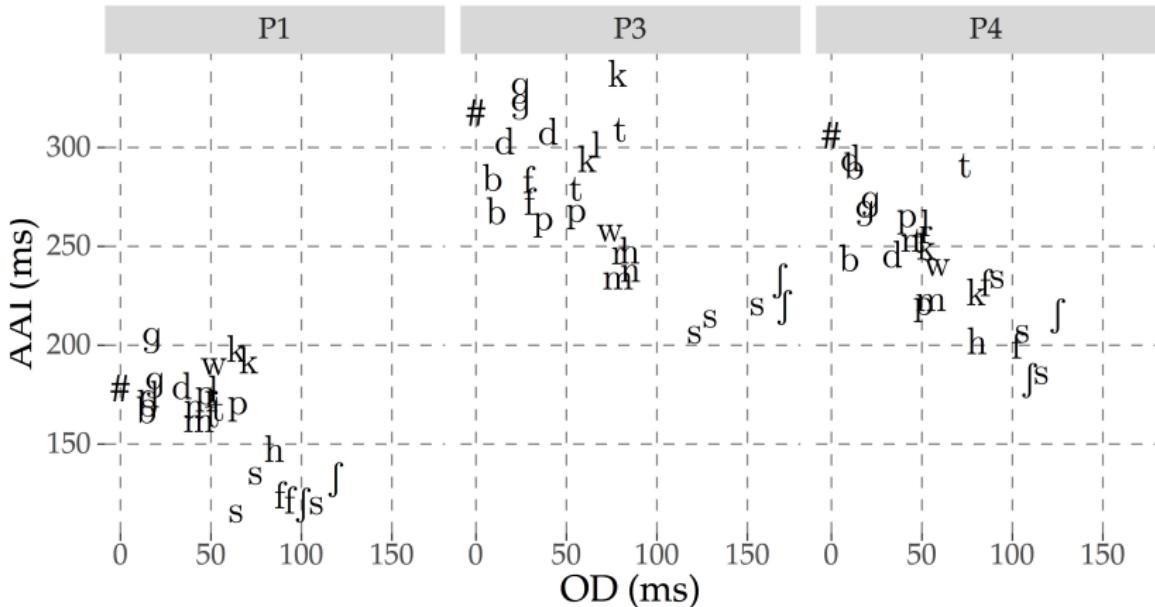
- ▶  $i$  and  $j$  are indices that span the width and height of the image,  $t$  is the time index.
- ▶ Like said, this is actually just the Pythagorean theorem applied in a space with a lot more dimensions than 2.

# Pixel Difference (PD): The maths visually



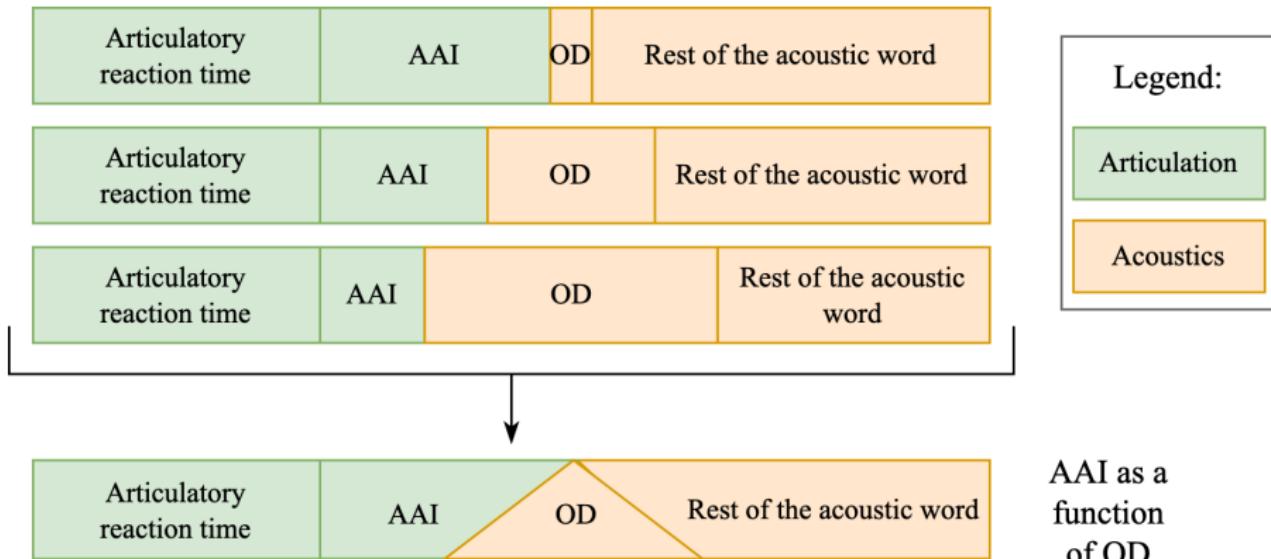
Some quick results

# Delayed naming results: Articulatory to Acoustic Interval

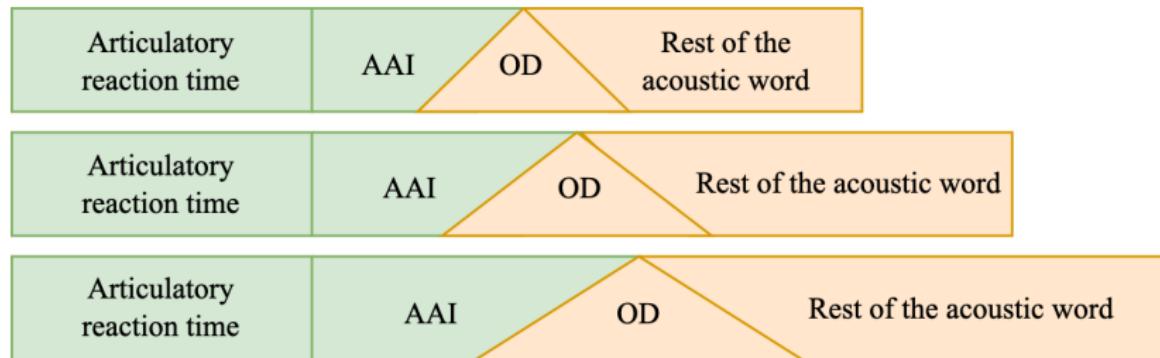


Medianised within participant, over several repetitions and over the vowels /a,i,ɔ/. Over all analysable n = 1386: 439 from P1, 672 from P3, and 275 from P4.

# Theory: Effect of OD on AAI

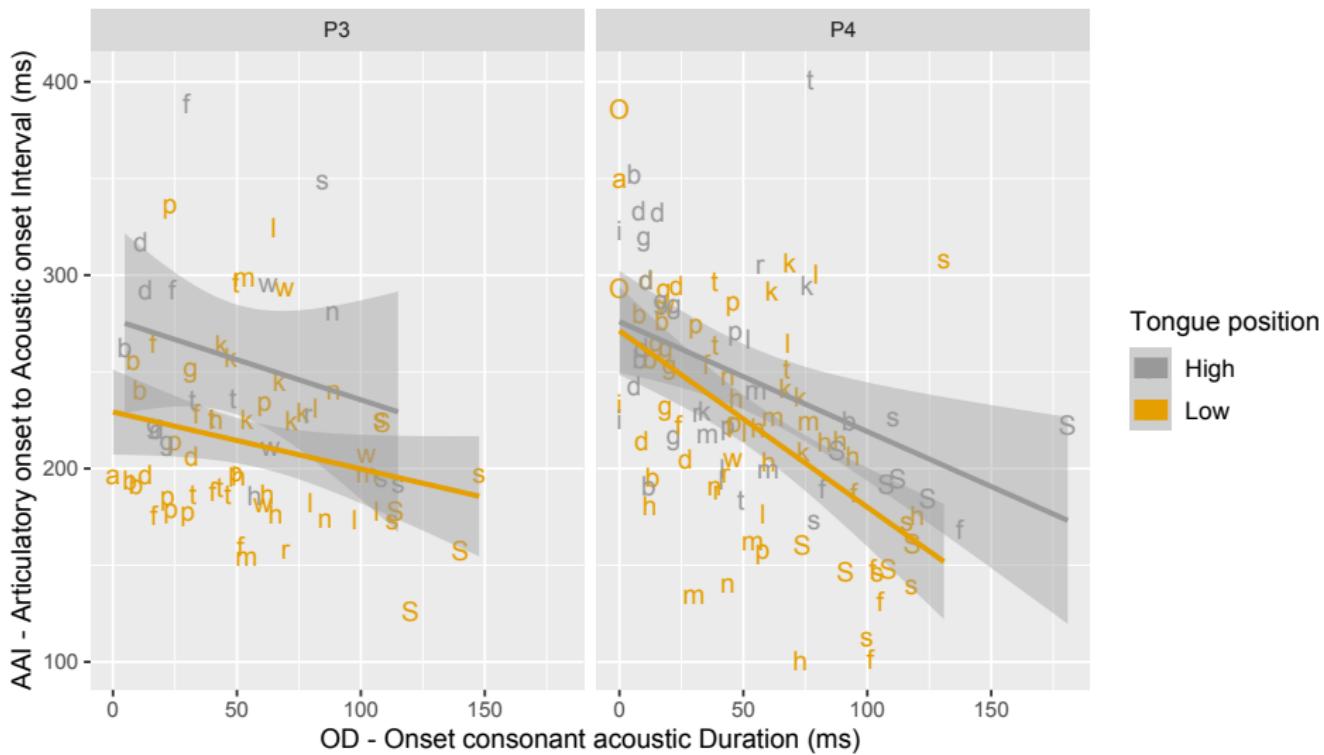


# Theory: Effect of articulatory rate on AAI



# Starting position

- 'Remain at rest' does not define what 'rest' means.



# Conclusions

# Conclusions

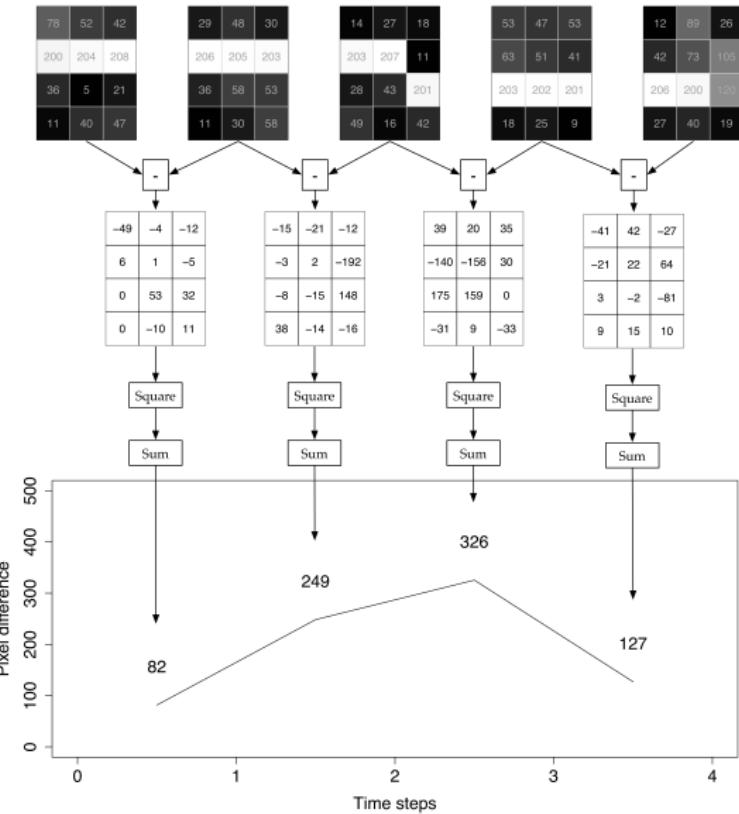
## Speech initiation:

- ▶ It's a lot more complex than I would have thought almost 12 years ago.
- ▶ The results in this talk rest on samples that aren't terribly big and don't span too many languages (but aren't based in just English).

## Methods:

- ▶ We don't always need to look to super complicated tools to actually make a drastic improvement in a workflow.
- ▶ We should always spare at least a short moment to think about methods – arguably even when doing a 1:1 replication of a previous study.
- ▶ We should try to find a balance between doubting our methods and just getting on with the work.

# Pixel Difference (PD): The maths visually



Thank you! Questions?

# References

- Drake, E., Schaeffler, S., and Corley, M. (2013). ARTICULATORY EVIDENCE FOR THE INVOLVEMENT OF THE SPEECH PRODUCTION SYSTEM IN THE GENERATION OF PREDICTIONS DURING COMPREHENSION. In *Architectures and Mechanisms for Language Processing (AMLaP)*, Marseille.
- Euclid (2006). *The Elements: Books I-XIII - Complete and Unabridged*. Barnes & Noble. Translated by Sir Thomas Heath.
- McMillan, C. T. and Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, 117(3):243–260.
- Palo, P. (2019). *Measuring Pre-Speech Articulation*. PhD thesis, Queen Margaret University, Edinburgh.
- Raeesy, Z., Baghai-Ravary, L., and Coleman, J. (2011). Parametrising Degree of Articulator Movement from Dynamic MRI Data. In *12th Interspeech*, pages 2853–2856.
- Rastle, K., Harrington, J. M., Croot, K. P., and Coltheart, M. (2005). Characterizing the Motor Execution Stage of Speech Production: Consonantal Effects on Delayed Naming Latency and Onset Duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):1083–1095.