

Review of: "Are non-native speakers' English pronunciation similar?: Evidence through PRAAT recording software"

Dominic Schmitz¹

¹ Heinrich-Heine Universität Düsseldorf

Potential competing interests: No potential competing interests to declare.

The author has embarked on an interesting topic of inquiry that could have implications for the acquisition of English pronunciation by non-native speakers from Bangladesh and China. However, both the research and the paper would benefit from substantial revisions. Please carefully consider the following questions and comments as you proceed with your revisions:

Section 1

- the introduction contains quite some information that is either common knowledge for researchers in this area (e.g. the meaning of 'phoneme') or unrelated to the research at hand (e.g. Levelt's model)
- instead of having this unnecessary information, consider motivating your own research: why is it interesting? has there been previous similar research? if so, why do you need to add something?

Section 2

- again, superfluous information: why do you very briefly (and rather roughly) explain how language sounds are produced? this information is common ground and not required (it were if your study was located in, e.g., articulatory phonetics)
- why is Praat part of the literature review? Praat is software, not literature, and its explanation should thus be situated in the methodology section
- why do you mention forensic linguistics? I agree that it is an interesting field, but it is not part of the research at hand

This section definitely needs more information on similar research and the gap your study aims to fill.

Section 2.1

- again, redundant information: you do not need to summarise after a few short paragraphs

Section 2.2

- what language is "Chinese"? be specific or be knowledgeable in explaining that there are multiple potential L1s when talking about China as a whole
- why is your interim summary part of Section 2.2? it should be its own section, i.e. Section 2.3

Section 3.1

- there is no need for the first paragraph, i.e. you do not need to explain what 'research design' is - instead, focus on explaining your research design; for instance, there is no information on how you came up with your stimuli or how the recording process worked

Section 3.4

- this does not make sense: "she is fluent in English" vs. "not fluent in English speaking"

Section 3.6

- what type of microphone did you use?
- in which surroundings did you record the two speakers? was it quiet, was it loud?

Section 4

This is the section I take most issue with. I cannot follow your reasoning at all.

- Figure 1:a. There is no voicing between the burst and the onset of friction. If there was voicing, there would be quasi-periodicity visible.
- Figure 1:b. I cannot identify a burst.
- Figure 1:c. This is friction followed by a vowel; I cannot see a burst.
- Figure 2:a. This is largely nothing but probably a very silent/low frequency fricative.
- Figure 2:b. No burst visible.
- Figure 2:c. No burst visible; instead a fricative, followed by a vowel, followed by another fricative?
- Figure 3:a & Figure 4:a. You talk about stress in regard to these figures, but you cannot see stress in them. Both images show a monosyllabic word (*think*); stress can only be identified via comparison of multiple syllables. Additionally, in Figure 4:a, the coda of *think* is not fully contained within the annotation.
- Figure 3:b and Figure 4:b. Honestly, I do not understand these figures. Figure 3:b is created without IPA, Figure 4:b contains IPA - but only for 'Phonemes'. I cannot make sense of two figures I am supposed to contrast when their makeup is already different independent of their actual difference.
- Figure 5 and Figure 6. Here you talk about fundamental frequency (I assume). However, the values you report at the values of the location you clicked on (i.e. the location of the red dotted line). It is not any meaningful value related to the two speakers. Additionally, what would a difference in fundamental frequency tell us? Nothing more but that two speakers have different fundamental frequency.

In sum, please provide all figures with correctly aligned textgrid annotations. This is the only way your readers can follow your reasoning.

Sections 5 & 6

- I cannot follow your reasoning; mixing IPA with non-IPA transcriptions/descriptions of sounds does not help

A very general problem: Looking at two data points is not meaningful in any way. With one data point per speaker, you cannot even conclude whether these two speakers are different in any way. Inferring anything about the speakers' general populations, i.e. L2 speakers of English from Bangladesh and China, is out of the question.

Further, I cannot confirm your reasoning. You conclude that Speakers A and B produce different sounds for the onset of *think*. However, looking at Figures 5 and 6 (and also 3:a and 4:a) to me it looks like both speakers produce the same sounds: voiceless dental fricatives /θ/.

On top, the results of the qualitative analysis are missing completely (or at least, I cannot spot them).

Finally, once Section 2 is revised to contain information on previous similar research and the research at hand is motivated, the discussion needs to relate to previous findings and this paper's motivation. That is, while there are research questions and hypotheses given, none of them are mentioned ever again, while they definitely should be mentioned in the discussion.