FACULTY OF ARTS
Charles University

Institute of Phonetics, Faculty of Arts, Charles University

Charles University
Faculty of Arts
Institute of Phonetics
doc. Mgr. Radek Skarnitzl, Ph.D.

A Review of the Doctoral Dissertation

Czech accent in English: Linguistics and biometric speech technologies,

submitted by

Mgr. Jakub F. Bortlík

The doctoral dissertation by Mgr. Jakub Bortlík examines various aspects of foreign accentedness in English as the foreign language (L2), specifically the rating of foreign accents and the impact of foreign accentedness on language identification and speaker identification technologies. The focus of the submitted dissertation is therefore highly topical, interesting, and original, interfacing linguistics and speech technology, areas where there has been relatively little contact so far.

The two main theoretical chapters present an overview of the rating of foreign-accented speech and biometric speech technologies, respectively. In the first of these, the author introduces essential terminology and discusses studies which have investigated various factors affecting foreign accent rating, such as familiarity with the accent, the actual question given to the respondents, length of stimuli, or listening conditions. As for terminology, I must confess that I failed to understand why the author adopted (non)-native "talkers". First, the explanation (p. 4) for diverging from decades of relatively well settled terminology (the commonly used "speakers") completely eludes me; second, the author himself is not consistent in using this unusual term; and third, "talker" seems to parallel the extremely unfortunate Czech term "řečník", used by some experts in speech technology instead of the unmarked "mluvčí". I would like to ask Mr. Bortlík for a comment during the defence.

The second theoretical chapter discusses two major applications within speech technology, automatic speaker recognition (ASR) and language identification (LID). After introducing the key terminology and metrics for gauging the performance of ASR, the author briefly describes some of the main challenges ASR faces, namely language and channel mismatch, which are subjected to experimental testing in subsequent chapters. However, I would have welcome a more complete picture of challenges faced by ASR technology, such as mismatch in speakers' behaviour (e.g., loudness, Lombard effect, or voice disguise) or spoofing. On the other hand, I would like to commend the author for explaining the working of a biometric ASR system in a very simple and reader-friendly way. The ASR and LID sections are concluded with the introduction of the specific software used for hypothesis testing. I truly appreciate that the author compares two systems (Phonexia and an open-source toolkit SpeechBrain); such comparisons, using the same material and procedures, are invaluable. I only have one question concerning the description in Chapter 3: should the LLR relationship mentioned on p. 17 not be the other way around?

The experimental part of the dissertation is divided into three chapters. In a rather unorthodox manner, Jakub Bortlík decided to describe the foreign accent rating study and the ASR and LID experiments in two subsequent chapters which introduce the bare minimum of methodology, and then the results and discussion; the last of these chapters is then dedicated to a detailed description of methodological aspects. While I can understand the appeal of not overwhelming readers with methodological minutiae, I believe that a dissertation represents a specific genre which has certain requirements, and these should be, if possible, adhered to. The structure, as it now stands, would be much more suitable for a book; and if Jakub Bortlík plans to publish the dissertation as a book, he will have less work with it. Importantly, Chapter 6, called Data collection and experiments, is crucial in understanding the validity and generalizability of the experiments conducted in the previous two chapters. It is fair to say that, in chapters 4 and 5, the author does refer the reader to relevant sections of chapter 6 where necessary.

The first experimental chapter on foreign accent rating returns to some of the question posed in chapter 2. Specifically, the author tests how the formulation of the rating task affects the rating, how it changes with recordings processed by imitating a landline telephone transmission, how foreign accent rating correlates with speech rate and raters' familiarity with the accent, and how ratings are affected by language mismatch. The second experimental chapter tests the performance of the above-mentioned tools in conditions of language and/or channel mismatch. LID technology is also tested for its sensitivity to the amount of speech material available. Both experimental chapters are very well conceptualized and written. The author justifies each partial experiment and provides detailed predictions and hypotheses for it. The results are analyzed using current statistical methods, visualized with suitable plots, thoroughly commented, and carefully discussed.

Returning to the methodological chapter 6, here the author describes the details which are crucial for judging the contributions of the submitted dissertation, such as the makeup of the questionnaires and speaking tasks, details about the speakers and listeners, or the administration of the listening experiment using PsyToolkit. There is one issue I would like to discuss during the defence, namely the length of the training stage of the listening experiment. On p. 72, the author mentions that some respondents found training too long and exhausting, and I can only agree. The explanation provided for this choice does not seem adequate to me: one typically aims to include such a selection in a training session which provides respondents with an idea of the range of accentedness they can expect. Why did it seem necessary to the author to present "a balanced representation of all talkers" (p. 72)? Is there any support for a training session conceived in this manner in literature?

The dissertation is concluded with a discussion where Jakub Bortlík summarizes the results of his experiments, relates them to existing studies and discusses their limitations, as well as implications.

Overall, I regard the submitted dissertation as very good research which spans linguistics and speech technology. One of its aspects which I appreciate is the author's honesty in admitting shortcomings and limitations of the experiments, whether they were due to time restrictions, the ongoing pandemic or other reasons. As for formal aspects, the dissertation is of an exceptionally high standard – both in terms of the excellent language with very few mistakes and typos (e.g., there is a piece of text missing at the top of p. 7; *How srong* in Fig. 6) and crediting sources (I only noticed one page entry missing with a direct citation on p. 6). I truly appreciate that, in his dissertation, Jakub Bortlík is asking questions and discussing options relating to those questions which go beyond what speech engineers typically do, and that he is trying to suggest linguistically and phonetically interpretable answers to these questions. Thanks to this dissertation, the gap between linguistics and phonetics on the one hand and speech technology on the other has become a little bit smaller.

Based on the review presented above, I conclude that Mgr. Jakub Bortlík has convincingly demonstrated his capacity to conduct independent scientific research and that the submitted dissertation meets the requirements of a doctoral dissertation. I recommend that it be accepted for defence.

Prague, January 10, 2022

doc. Mgr. Radek Skarnitzl, Ph.D.