Research Practicum Report: Exploring Speech Acoustics and Phonological Research

A report based on three tasks

Author: Yimeng Zhang

Student Number: 13678469

Supervisor: Dr. Titia Benders, Marloes Roosingh

Period of Internship: Oct. 2022 – Aug. 2023

Date of Submission: 31/8/2023

Educational Context: Honours programme

Enrolled Programme: BA Linguistics

Title of the Internship: Speech Lab Intern, Research Assistant

Credits: 12 ECTs

1. **What exactly has been achieved, what type of research has been carried out?**

Throughout the entire research practicum, I have participated in three different research.

1) The first research focused on the acquisition of emphatic consonants /tˤ/ and /sˤ/ of Arabic-speaking monolingual children (Alkhudidi, Demuth & Benders, 2022). By analyzing data from a speech corpus, I assisted the researcher with checking automatic formant measurements and placing temporal landmarks for the subsequent statistical analysis. Additionally, it is done by adjusting the segment boundaries that are created using a written Praat script.

This research was carried out through:

1: Online research: The researcher extracted the data from the Arabic Kuwaiti corpus online (Alqattan, 2015). In this corpus, parents interacted spontaneously with their children using toys and/or picture books.

2: Laboratory research: In addition to online data, the researcher also recruited 40 children and 10 adults from Saudi Arabia. These participants were instructed to do a single-word-repetition task on site. The collected data in this elicited experimental study was used for the data that I was going to analyze (Alkhudidi et al., 2022).

3: On-site research: As an annotator, I conducted reliability checks using a computer at the Speech Lab at University of Amsterdam due to the sensitivity of the recorded data.

4: Cross-sectional research: The researcher compared the target stimuli produced by children within a specific age range with those produced by adults. The latter is served as the control group.

Based on the researcher's reflection, my annotation was close to other annotators, and the results were reliable to draw the conclusion. Some findings from my own perspective are mentioned in the third section of this report as well.

This research aimed to provide more insights into how complex segments are acquired and what factors (both age- and language-related) have the largest impact on acquisition. Additionally, the acquisition pattern of emphatic consonants establishes a baseline information for educators and clinicians in the Arabic-speaking community to develop assessment and diagnostic tools (Alkhudidi et al., 2022).

2) The second research examined how English-speaking toddlers and pre-schoolers produce sentences with or without articles in footed and unfooted position. In this

study, I contributed by annotating data using a written Praat script as well. This study aims to shed light on language disorders that may cause difficulties in producing function words (Lovcevic, Benders, Tsuji & Fusaroli, 2023).

This research was carried out through:

1: Laboratory research: The researchers recruited 27-month-old monolingual toddlers and 3-5-year-old monolingual preschoolers to perform an elicitation task in the lab.

2: On-site research: I conducted my annotation tasks at the Speech Lab at the University of Amsterdam due to the high sensitivity of data. My analysis would be later compared with other annotators to ensure the precision of the data for drawing conclusions.

3: Cross-sectional research: The researchers compared the stimuli produced by different age groups to test potential differences in sentence production.

By inspecting the spectrograms of different speakers, I observed that they usually differ in their pronunciation. For instance, some speaker tended to omit the first syllable(s) when the subsequent one is stressed, such as *gira-* in giraffe and *ba-* in *balloon*. The vowels are barely visible in the graphs. In comparison, other speakers pronounced unstressed syllables more clearly.

3) The third research involves a meta-analysis that investigates the pattern of infant-directed and adult-directed speech across languages. To achieve this, the vowel triangles generated by Praat in related studies are used. My task is to help the researcher collect the first and second formants of the vowels [a], [i] and [u] according to the diagrams using a measurement tool called PlotDigitizer.

1: Technological applied research: This study is one of the first studies that tries to gather the data directly from vowel triangles due to the scarcity of studies with raw data. To achieve this, I searched for suitable online tools and initiated this pilot project.

2: Online research: All studies and the PlotDigitizer tool were processed online. While I was collecting the data, I input vowel diagrams into the tool which plotted them on a coordinate system, then set the ranges for x- and y-axes. Finally, I plotted the vertices of the triangle manually and extracted the data into an Excel sheet in order to run an R programme for further research.

3: Theoretical research: The researcher compiled a literature list of all the studies that the

meta-analysis will cover. This list not only includes the vowel triangles from which I collect the data from, but also provides information about the goal and basic details of each study. Moreover, I could refer to the original studies through the reference list if I have further questions. It is a useful list to be created before meta-analyses.

Based on the collected data, there is no consistency across studies of adult-directed speech in Australian English. However, in Kalashnikova & Burnham (2018), adult-directed speech shows similar formant values across age groups (7, 9, 1, 15 and 19 months). Generally, child-directed speech also lacks consistency across studies and does not show a trend of vowel triangle area decreasing with age. According to Kalashnikova & Burnham (2018) and Kalashnikova et al. (2018), child-directed Australian English still somewhat demonstrates similar results in 9- and 11-month-old groups respectively.

Compared to Australian English, child-directed speech in American English shows more consistent data both within and across studies, and three vowels exhibit a clear increase in the first formant as infants grow. However, Startling (2019) is an exception that the first formants of the vowels are lower than in younger age groups in other studies. Moreover, as Australian English, adult-directed American English shows similar formant values across age groups and studies, but the formant values of the same vowels produced in adult-directed American English differ from those in Australian English.

2. Describe the research group (or the research project) and give an overview of the research questions and your contribution to the research. What still remains to be done in the (near) future?
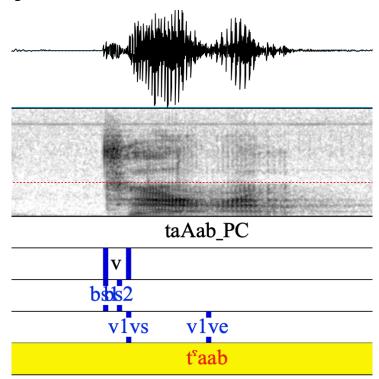

For all three research projects, most of the foundation has been provided by the supervisor and the researchers. I have been a research assistant who completed the tasks that I was assigned.

The first research was conducted by a researcher at University of Macquarie for her Ph.D. thesis. The research questions are: *what is the developmental trajectory in the acquisition of Arabic emphatic consonants? And how this is influenced by frequency, perceptual saliency, and articulatory complexity?* To address these questions, this research investigated how typically developing Arabic-speaking monolingual children at different age ranges acquire the emphatic consonants /tˤ/ and /sˤ/ compared to non-emphatic /t/ and /s/. The data was collected both from a

corpus and an elicited speech of recruited participants (Alkhudidi et al., 2022). The researchers used a programme in Praat to automatically analyze the data. Afterwards, they also recruited annotators to perform a reliability check.

My contribution to this research has been conducting the reliability checks by manually labelling 10% of the entire dataset. Specifically, I was responsible for determining what voice onset time is and where the burst(s) and vowels start for /tˤ/ and /t/; and for determining how long the consonant is and where the vowels start for /sˤ/ and /s/. To prevent position effects, all the segments were tested at word initial, medial and final position. The landmarks were initially generated by the programme, but I needed to adjust them by inspecting the spectrograms and waveforms due to some systematic errors.

Here is a brief example of labelling landmarks for emphatic /tˤ/ during my practice round before the actual coding task:

According to the spectrogram that takes the priority according to annotation criteria, there are two major bursts for this word-initial consonant, and the landmarks for bursts are thus placed. Furthermore, the start and end of the adjacent vowel /a/ are labeled based on the second vowel formant (the second horizontal dark band from bottom to top) in the spectrogram.

From the perspective of a non-native speaker, I suggest comparing the articulation of /sˤ

tˤ/ in different dialects since the constrictions of Arabic emphatic consonants remain a controversial topic (Alkhudidi et al., 2022). In future studies, the articulation of these two consonants in other dialects can be studied in order to determine if the current acquisition pattern of children from Saudi Arabia is suitable for other Arabic monolingual children.
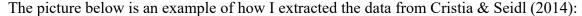
The second research is my supervisor's project at the University of Amsterdam. The research question is *do toddlers' omitted articles leave traces, and are those different between footed and unfooted position?* The researchers divided the stimuli into four categories with respect to footedness and article condition. Footedness manipulation is controlled by disyllabic and monosyllabic verbs. For instance, in the footed condition a monosyllabic verb such as "to *buy*" is used, while in the unfooted condition, a disyllabic such as "to *cover*" is employed. For the article condition, Article/Sg is when an article is included. In the sentence "Hazel buys a doll", the article is pronounced. In contrast, the Control/Generic condition does not contain an article, and the sentence would be "Hazel buys dolls." The singular article is replaced by a plural noun. Participants recruited for this study were asked to produce target SV(article)O sentences based on corresponding visual prompts after a few training sessions. After running a written Praat script with the collected data, the researcher recruited some annotators to conduct a reliability check.
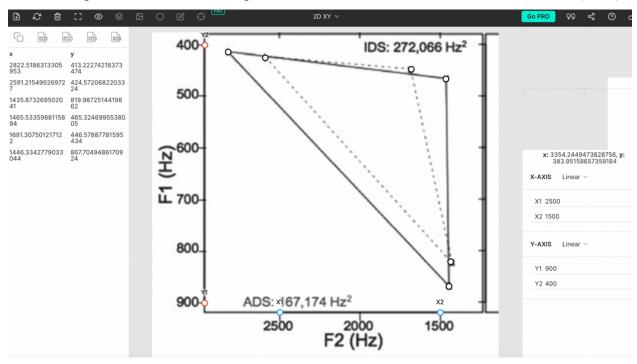
My contribution to this research is also the reliability check. Specifically, I measured the duration of various segments within a sentence, such as the vowel and the frication of the [z] in the verbs, the start and end of the articles, consonants and vowels of the nouns. However, unlike the previous study I have conducted on Arabic emphatic consonants, this one is more complex in the sense that the data I was analyzing for this study contained more collisions. I needed to consider the effect of other words to the target segments within the sentence. For instance, when a noun begins with a liquid, it becomes harder to inspect the starting point of the vowel from the spectrogram. Furthermore, when the stress of a noun is on its second syllable, the first syllable tends to be extremely close to the second which makes the vowel more difficult to be seen as well. A particular example would be *gira-* in *giraffe*, which has been almost obscured by some speakers.

In my opinion, the tokens that involve collisions could be approached differently in future

discussions. After discussing with one of the researchers, we decided to ignore relevant tokens. For instance, if the syllables *gira*-in giraffe are unclear on the spectrogram, the landmarks are thus deleted. However, I believe it is worth considering such tokens differently in future studies.

The third research aimed to compare infant-directed speech and adult-directed speech cross-linguistically by conducting a meta-analysis. The current research question is *whether F1 and F2 are (somewhat) consistently raised in IDS compared to ADS*. My task was to use an online tool to measure the values of first and second formants of the vowels in the diagrams. Before analyzing the data, it is necessary to find a tool to gather the data from the vowel diagrams since no records showed that any researchers had tried to do so. I first tried three tools, namely PlotDigitizer, WebPlotDigitizer and SketchAndCalc, respectively. After discussing with the researcher and reading some literature, such as Aydin & Yassikaya (2021), PlotDigitizer was the optimal software that we decided to initiate this pilot project with.

The picture below is an example of how I extracted the data from Cristia & Seidl (2014):

Firstly, I treated the diagrams as a coordinate system, where the horizontal scale represents the x-axis and the vertical scale represents the y-axis. Subsequently, I placed the points for defining x- and y-axis (X1, X2, Y1 and Y2) to the largest and smallest numbers that are visible on the graph. In this vowel triangle, the four points would be (2500,0), (1500,0), (0, 900) and (0, 400). After establishing these reference points, I manually plotted six points on the

vertices of the two triangles. In this graph, the F1 and F2 [i] in infant-directed speech (the bolded triangle) are about 413 Hz and 2823 Hz, respectively. I recorded data from all the vowel triangles from the studies in an Excel sheet.

Since data collection is the first step, several tasks remain to be completed in the future. For this meta-analysis, the researcher will run an R programme based on the Excel sheet that I created. Moreover, as other projects, she will recruit other interns to do the reliability check to ensure the data are precise for analysis. For future research, I hope more online measurement tools could be tested or created to determine the most accurate one.

3. What have you learnt during the research practicum? Reflect on what you have learnt.

Through this experience, I have learnt several theoretical knowledge and grasped practical skills.

1) In the first research, I learnt about Arabic emphatics and the articulation of these consonants. The production involves two constrictions, one located in the dental/alveolar region, while the other is estimated to be in the pharynx (Alkhudidi et al., 2022). By inspecting the spectrogram, I observed that sounds as emphatic consonants tend to affect their adjacent vowels by lowering the second formants of the vowels in comparison to their non-emphatic counterparts in the adult group. The researcher also mentioned this phenomenon in her work (Alkhudidi et al., 2022). Additionally, I could see from the graphs that infants tend to produce more bursts than adults when they produce plosives /t/ and /tˤ/ due to the smaller size of their vocal cavity. Although I had learnt the theory behind this phenomenon during phonetics and transcription courses, hearing how children's production differs from that of adults in practice still gives me a strong impression.

Apart from theoretical knowledge, I became more familiar with analyzing data with Praat, especially with spectrograms. I learnt how to place the boundaries for bursts, sibilants, and vowels. The actual coding was different from the exercises that I have done in class. There were usually special cases, such as the liquid /l/ and the tap /r/ which made the boundaries of vowels more challenging to determine. I needed to find solutions for these problems during the annotation process, which requires not only knowledge of the features of the sounds but also an understanding of the influence of adjacent sounds and different situations of the speakers.

Last but not least, I also gained some knowledge of conducting research on child language acquisition using Praat. For instance, I learnt that creating annotation criteria is necessary for data analysis since different criteria might lead to various results. I also learned that writing Praat scripts can help label the boundaries for the data. However, these landmarks need to be checked by multiple annotators to make sure the analysis is precise. This process is called reliability check. Although I might could also learn these concepts through coursework, being involved in an actual project allowed me to develop my skills for real-world challenges and deepens my understanding. This will be a valuable experience for me to pursue my own research in the future.

2) My tasks for the second research were similar to the first research, but I learnt something different from this experience. Compared to the first study where I was asked to label some landmarks for words, my coding task for this study involved analyzing different parts of sentences.

Additionally, I have learned some general skills, such as backward planning. During the first research, the researcher and I forgot to establish a deadline for my task. The deadline that I set for myself was different from her expectations. Therefore, I had to adjust my original plan after discussing the deadline afterwards. In a meeting with my supervisor, Dr. Titia Benders, she suggested that I could use a strategy called backward planning in the future. As a goal-oriented person, this skill would help me to keep productivity and stay on track. Therefore, before the second project began, I asked one of the researchers about the deadline to prevent the same situation from happening again. I also created a plan for myself based on the deadline. Furthermore, I recognized the importance of communication when I started with this project. When I was having my initial meeting with one of the researchers, Marloes Roosingh, she mentioned that she would upload some files including the audio and TextGrids to SURFfilesender, and she would upload other files later. Unfortunately, it led to a miscommunication because she was not aware that she forgot to send me the TextGrids with the audio files, while I believed that she would send them to me later. Neither of us noticed the problem until Marloes realized that I did not have the complete files to do the annotation. Moreover, I did not have time to download the new files before a one-week vacation, and I was unaware that the files would expire in one week. Therefore, we did not realize there was

another problem until I came back from vacation. With these two accidents, my progress has postponed for approximately a month. This highlights the importance of enhanced communication between team members.

3) For the third research, I have learned many new definitions such as the scales (mel scale and bark scale). Compared to hertz scale which is a linear scale used to measure physical frequency, the mel and bark scales are logarithmic scales specially designed for human perception of sound frequency (Chen, Zhang, Liu, & Yuan, 2010). These definitions are essential for conducting research related to perception and production. Furthermore, I realized how interdisciplinary knowledge is necessary for conducting research. This project is a great example as a combination of mathematics, computer science and linguistics is used to gather, record and process the data.

Moreover, I have acknowledged that there are many gaps that remains to be filled in linguistics. In my opinion, the vowel triangle is one of the basic schematic representations of vowels, and the graphs are easy to create with Praat. However, researchers often focus only on measuring the area of the triangles in their studies and neglecting the actual values of the formants. This has led to the absence of a uniform calculator for the coordinates. Therefore, I learned that being a researcher requires not only academic skills but also the ability to recognize the unsolved problems in the field.

4. Describe how you could expand on your experiences in this research practicum in a possible extended honours thesis.

Having worked on two projects related to infant articulation, I have developed an interest in studying the acquisition process of infants. There is an interesting phenomenon in Mandarin where children tend to mix retroflexes ($/\widehat{t\wp}^h/$, $/\widehat{t\wp}/$) with affricates ($/\widehat{ts}^h/$, $/\widehat{ts}/$, $/t\textctc^h/$, $/t\textctc/$). After discussing this phenomenon with my supervisor, Dr. Titia Benders, she suggested that this mixing might be attributed to two possible factors. One is that children may not be able to distinguish the retroflexes from affricates in their minds, the other factor is children are able to distinguish the sounds underlyingly, but they cannot yet produce the sounds as in adult speech due to biological immaturity. This topic can be investigated using similar research methods to those employed in the first two projects that I assisted with. It is possible to conduct acoustic analysis on children's production in order to determine if there is a difference between their

mispronunciation and the target sounds. However, as a bachelor's student, I am not yet allowed to conduct experiments with children. Interestingly, this linguistic phenomenon also occurs in adult speech although those who exhibit such problems are usually diagnosed with certain disorders or affected by dialects. I could at least start by studying adult speech using similar methods in my extended honours thesis.

5. Finally, evaluate the research practicum overall (what went well, what went wrong, what could have gone better?)

This research practicum provided me with an opportunity to work as a linguist. As my first working experience using the academic knowledge I have gained at university, it has given me a clearer idea of which area of linguistics I would like to specialize in for my future academic career.

Overall, the research practicum went well. It allowed me to meet more professors and researchers in the field of linguistics by attending bi-weekly BiPhon meetings and monthly Lab Linguistics meetings. During these meetings, I acquired some basic knowledge of how research methods are designed for each study and how researchers communicate to enhance their studies, such as through participation in academic conferences. Throughout this experience, I have gained an insight into the process of conducting research and a general understanding of the steps I need to follow if I would like to publish an academic paper in the future.

My supervisor, Dr. Titia Benders has been significantly helpful during the entire research practicum. She not only assigned tasks to me but also encouraged me to explore the unknown by myself. For instance, when I was working on the third project, I did not understand the terms "bark" and "mel" that are the scales used for measuring the formants. She did not tell me the answers directly. Instead, she encouraged me to find the definitions on my own. When I was conducting the research, I learned more than the meanings of these scales; I also discovered under what kind of circumstances these scales were employed. In the subsequent meeting, we discussed my findings and she explained the scales specifically relative to the studies that would be analyzed. Throughout the entire practicum, Dr. Benders has played the role of a guide, and with her assistance, I became more self-reliant and confident when facing new challenges in my academic career.

      The research practicum was planned a bit different from my initial expectation. I thought I would help with one researcher with one of their projects exclusively. However, since there were no ongoing studies that required assistants for more than three hundred hours, I worked on three separate projects consecutively. While this allowed me to experience more kinds of studies in phonetics and phonology, it was somewhat time-consuming to transition to and adapt to the new project after completing the previous one. Furthermore, I realized that I needed to develop more efficient time management skills because my schedule for going to the Speech Lab had to change every two months due to the flexibility of blocks. My plans were occasionally interrupted by exams and courses, but fortunately I managed to achieve the goals that I had set for myself for each project. I am pleased that I could apply the knowledge into practice while I am learning from the class at the same time.

Reference

Alkhudidi, A., Demuth, K., Davies, B., & Benders, T. (2022, December). The Acquisition of the Arabic Emphatic – Plain Consonant Contrasts by Arabic Speaking Children. [Poster presentation] The Australasian Speech Science and Technology Association Conference, Australia.

Alqattan, S. (2015). Early phonological acquisition by Kuwaiti Arabic children.

Aydin, O., & Yassikaya, M. Y. (2021). Validity and Reliability Analysis of the PlotDigitizer Software Program for Data Extraction from Single-Case Graphs. *Perspectives on Behavior Science*.

Chen, H., Zhang, W., Liu, J., & Yuan, Q. (2010). Objective assessment of speech quality by combining Bark- and Mel-scale frequency. *IEEE*.

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913-934.

Lovcevic, Benders, Tsuji & Fusaroli (2023). Vowel-hyperarticulation in Infant-directed Speech: A Multi-method Meta-analysis**. Manuscript submitted to: Psychological Bulletin.

Kalashnikova, M., & Burnham, D. (2018). Infant-Directed Speech from Seven to Nineteen Months Has Similar Acoustic Properties but Different Functions. *Journal of Child Language*, 1-19.

Kalashnikova, M., Goswami, U., & Burnham, D. (2018). Mothers speak differently to infants at-risk for dyslexia. *Developmental Science*, *21*(1), e12487.

Starling, G. (2019). Vowel Production in Infant-Directed Speech: an Assessment of Hyperarticulation and Distributional Learning (PhD Thesis). University of Edinburgh.