Research Practicum Report: Internship at UvA Speech Lab

Name: Yuying Zhu
Student Number: 14452685
Supervisor's Name: Dr. A.T. (Titia) Benders, M.H. (Marloes) Roosingh
Period of Internship: Sep. 2023 - Aug. 2024
Date of Submission: 11-09-24
Educational Context: Honours programme
Enrolled Programms: BA Linguistics
Title of Internship: Speech Lab Intern, Research Assistant
Credits: 12 ECTs

Table of Contents

1. Internship Overview

The internship (research practicum) was conducted at the UvA Speech Lab, under the supervision of Dr. A. T. Benders, Assistant Professor of Linguistics at the UvA, and Marloes Roosingh, Lab Manager of the UvA Speech Lab. The internship spanned over a full year from September 2023 to August 2024.

As a Research Assistant at the Speech Lab, my internship mainly involved assisting with several ongoing research projects, participating in lab meetings and linguistics-related seminars, as well as engaging in self-directed learning and exploration.

2. Research Projects

During my internship, I primarily contributed to two research projects. The first project was a study on prosody production and perception of English-speaking children in Australia, while the second project is a study on children's acquisition of Bassa phonemes. The two projects will be called *Project 1* and *Project 2* in this report.

2.1 Project 1: Perception-Production Link in Child Language

2.1.1 Overview

The first project, which spanned the entire duration of my internship, was a pilot study of a research project under the title "Perception-Production Link in Child Language" led by my supervisor, Dr. Benders. The study aims to investigate the perception and production of speech prosody of English-speaking children in Australia.

My assistance with this project involved two main stages. In the first stage, which lasted from the beginning of my internship until early December 2023, I annotated the audio files collected for the trial phase of the study. In the second stage, from late December until the conclusion of my internship, I designed and conducted a reliability analysis based on the existing data.

2.1.2 Theoretical background

Previous studies (Wheeldon & Lahiri, 1997; Yuen et al., 2022; Wynne et al., 2018)

have found that in speech production, reaction times are influenced by the number of Prosodic Words (PW) rather than Orthographic Words or syllables, emphasizing the role of PW and function word cliticization in speech planning. Research on English-speaking two-year-olds shows that, despite omitting unstressed syllables, prosodic traces like extended durations remain (Carter & Gerken, 2002). However, the cliticization of English articles in children's speech is still unclear regarding its direction (Cai et al., 2024). Therefore, analyzing segment durations in cliticized and uncliticized utterances could provide valuable insights into their speech prosody development.

For more detailed information of the theoretical background of the current study, see *Section 1.2* of the *Reliability Report* in the *Appendix*.

2.1.3 Stage 1: annotation

The initial stage of my engagement in *Project 1* was to make annotations to audio files recorded from adult speakers using *Praat* (Boersma & Weenink, 2024). Verb phrases (VP) produced by the participants that consist of a verb, a noun, and occasionally an article constitutes are the primary focus of this study. In particular, the durations of various segments of the constituents within the VP are of interest. These durations are marked by specific landmarks that indicate the start and end points of the relevant speech segments. For example, the landmark “ssv” marks the beginning of the stressed vowel in the verb, while “su.v” and “eu.v” denote the start and end points of the unstressed vowel found either in the noun or the article. An example of an annotated utterance is illustrated in *Figure 2*. The landmarks (“points” in *Praat*) on tiers 2 to 4 were those involved in the annotation task.

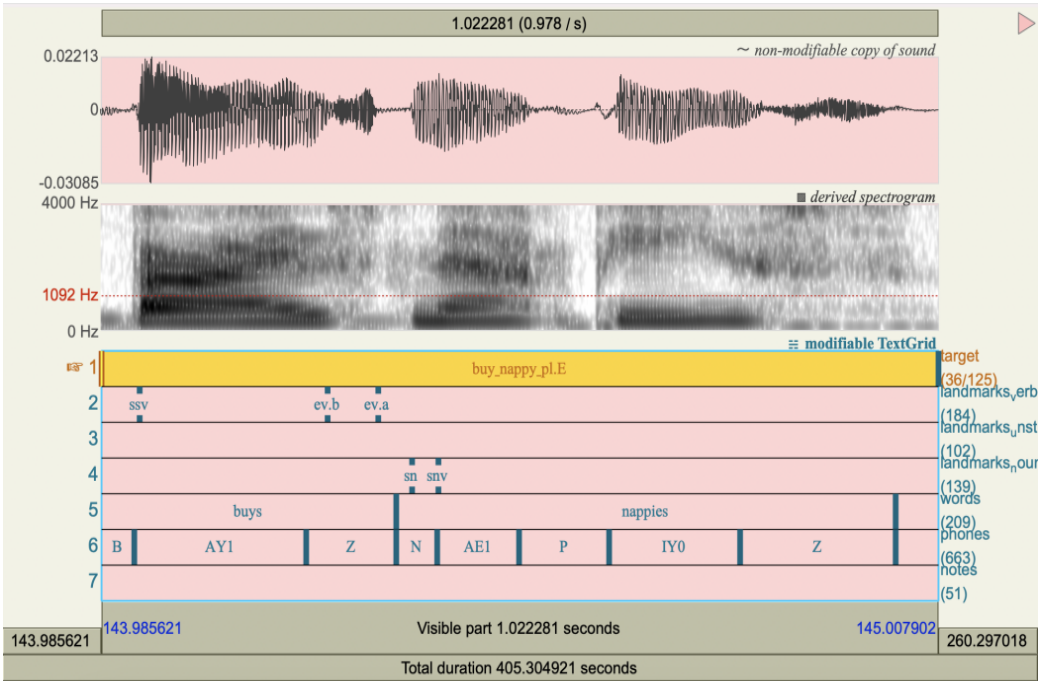


Figure 1: An annotated utterance "buys nappies"

Initially, these landmarks were generated and aligned automatically using *Montreal Forced Aligner* (McAuliffe et al., 2017). However, the timing accuracy of these landmarks was insufficient for the study's requirements, which necessitated precision at the millisecond level. Therefore, to accurately locate the speech segments of interest, it was essential to manually adjust the positions of the automatically generated landmarks, using spectrograms as the primary reference and occasionally consulting the waveform for additional clarity.

Over the course of three months during the annotation phase, I coded a total of 12 audio files involving 6 adult female speakers. To ensure confidentiality regarding the speech data, all annotation work was conducted onsite at the UvA Speech Lab. I remained largely unaware of the study's objectives, serving as a "naive coder" throughout the entire annotation process. Once the annotations were completed, I was provided with relevant literature on the research topic to facilitate my progression to the next stage of my tasks.

For more details on the annotation task, see *Sections 1.3* and *6.3* of the *Reliability Report*.

2.1.4 Stage 2: reliability analysis

(1) Overview

The second stage of my engagement in *Project 1* was to conduct a preliminary reliability analysis based on the annotations made by a former intern at the Speech Lab on 8 recordings of 4 female adult speakers. The analysis aimed to evaluate the reliability of the existing annotation schemes and identify potential areas for enhancement. The ultimate goal was to improve the quality and effectiveness of future annotations.

(2) Designing the reliability analysis

One of the biggest challenges I encountered in this stage was the lack of literature to refer to. Since most studies I encountered typically do not provide detailed information on how the reliability of data annotation schemes was verified, I had to take the initiative to independently develop the entire reliability analysis method, encompassing everything from data collection and processing to analysis and interpretation.

(3) Data collection: re-coding

Data collection for the reliability analysis was conducted through re-coding approximately 20% of the utterances annotated by the original coder in *Praat*. To ensure variation in the utterances involved in the reliability analysis and an even distribution of the utterances among speakers, a combination of controlled and randomized selection methods was used to choose the utterances for re-coding. For a more detailed description of the re-coding scheme, see *Appendix 1* of the *Reliability Report*. Ultimately, 79 utterances were re-coded for the reliability analysis.

(4) Data processing

Since the data set was reasonably small, data processing for the reliability analysis was carried out manually in the first place. Specifically, comparable

landmarks in the original codings and the re-codings were filtered out. 554 data points (i.e., annotated landmarks) were involved in the reliability analysis. The consistency between the two coders were calculated based on the criterion that the annotations to a specific landmark by the two coders should be deemed consistent if the two annotations are less than 5ms (=0.005s) apart from each other.

However, further analysis was necessary to evaluate the reliability of the existing annotation scheme, particularly regarding the consistency of the length of certain speech segments. This required more in-depth data analysis, which proved to be too complex to carry out manually. Therefore, I have done a self-oriented exploration on processing data with *Python* scripts (Van Rossum & Drake, 1995). In the end, I produced a *Python* script that in principle can automatically process any raw data in a particular format, including selecting and paring up comparable landmarks, calculating inter-coder differences, and judging their consistency. More detailed explanations on the scripts used for the data processing can be found in *Appendix 3* of the *Reliability Report*.

(5) Preliminary data analysis

For a preliminary reliability analysis, the descriptive statistics for the absolute differences and the percentages of inter-coder agreement was calculated. Upon suggestion from my supervisor, the quantiles of differences and absolute differences between coders as well as visualizations of the distributions and quantiles were also added. The visualization is as shown in *Figure 2*. For better visualization, 29 data points (5% of the total) with extreme values exceeding ±20ms were trimmed.

For the results of the preliminary data analysis, see the *Reliability Report*.

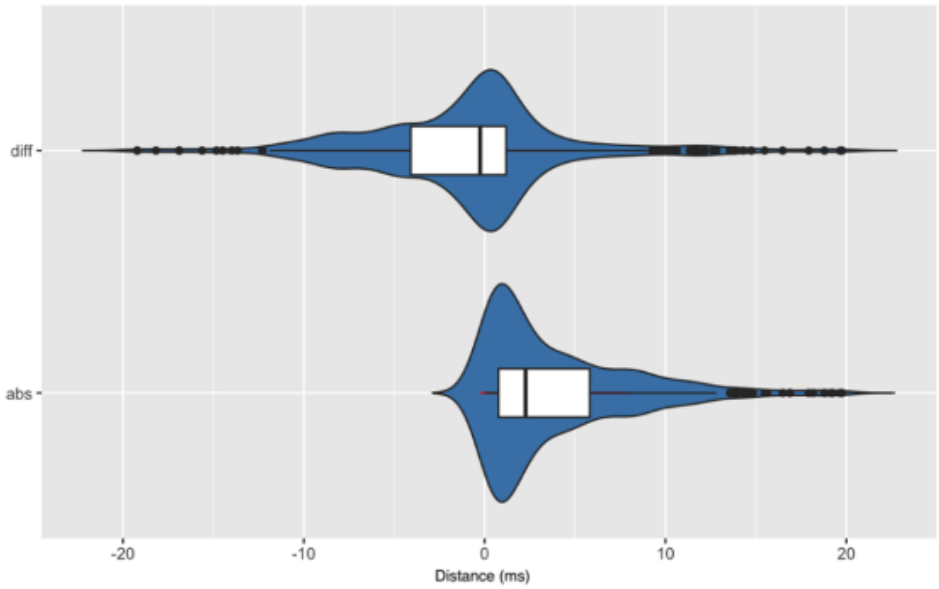


Figure 2: An overview of the inter-coder consistency by percentages

(6) Cohen's kappa

However, as pointed out by Cohen (1960), merely stating the overall agreement percentage is insufficient for assessing reliability (Lane et al., 2023). Therefore, more reliable statistical methods for consistency analysis were considered.

Following my supervisor's suggestion, the Cohen's kappa coefficient (Cohen, 1960) was used to achieve a more accurate assessment of the inter-coder consistency. Cohen's kappa, as shown in the equation in *Figure 3*, is a specialized measurement of reliability that accounts for chance agreement. In this context, P_O represents the observed agreement, while P_C denotes the expected agreement due to random chance.

$$\text{kappa} = \frac{P_O - P_C}{1 - P_C}$$

Figure 3: Cohen's kappa (Cohen, 1960, as cited in Lane et al., 2023)

The data processing and calculation of Cohen' s kappa was performed using *Python* scripts (see *Appendix 3* of the *Reliability Report*). Since Cohen' s kappa is only applicable to categorical data, the continuous outcomes obtained from the initial data processing were converted into categorical data by assigning each data point a

binary value, for instance, either 0 or 1. Specifically, for each annotation made by Coder 1, a value of 0 or 1 was assigned randomly. For the corresponding annotation made by Coder 2, the same value was assigned if the difference between the two annotations was less than 5ms; otherwise, the opposite value was assigned. The scheme for converting continuous data into categorical data is outlined in *Table 1*.

	Coder 1	Coder 2
consistent	0	0
	1	1
inconsistent	0	1
	1	0

Table 1: Scheme of converting continuous data into categorical data

To assess the effectiveness of this data processing scheme, Cohen's kappa was calculated on the same data set 50 times, each time using a distinct set of randomized binary values. It was observed that while the kappa results exhibited significant variability for smaller datasets, the kappa values derived from larger sample sizes, those with more than 50 tokens, were notably more stable. This suggests that the data processing method exhibits a level of effectiveness when applied to sufficiently large datasets.

To note, the criteria for interpreting the Cohen's kappa coefficient vary quite significantly across different studies. Two interpretations of Cohen's kappa were utilized in the reliability analysis (see *Appendix 5* of the *Reliability Report*). In general, a kappa value above 0.60 indicates a moderate to strong level of agreement.

(7) Other statistical methods explored

Due to my limited mathematical competence, the method used to calculate Cohen's kappa in the reliability analysis may be of low effectiveness. Moreover, since the categories applied in the Cohen's kappa calculation were determined by a simple

conditional judgment regarding whether the annotations made by the two coders are more than 5ms apart, it failed to capture more subtle differences in the original continuous data, namely, the extent of the disparity between the coders. Therefore, other statistical methods for calculating the reliabilities of continuous data like *Intraclass Correlation Coefficient* (Shrout and Fleiss, 1979, cited in Lane et al., 2023, see *Table 2*) and *Cronbach's alpha* (Cronbach, 1951, cited in Lane et al., 2023), as well as consistency analysis based on *Bland-Altman* plot (see *Figure 4*) were tested on the existing data set using *R* (R Core Team, 2013. For the *R* scripts, see *Appendix 3* of the *Reliability Report*). However, no valid results were yielded. It was suggested that other more reliable statistical methods should be considered for the reliability analysis.

Versions of Intraclass Correlation Statistics for Various Reliability Designs

Type of reliability study design	Raters fixed or random?	Version of intraclass correlation
Part A. Reliability of single rater		
Nested: *n* subjects rated by *k* different raters	Random	$ICC(1,1) = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_w^2}$
Subject by rater crossed design	Random	$ICC(2,1) = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_J^2 + \hat{\sigma}_e^2}$
Subject by rater crossed design	Fixed	$ICC(3,1) = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_e^2}$
Part B. Reliability of an average of *k* raters		
Nested: *n* subjects rated by *k* different raters	Random	$ICC(1,k) = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_w^2/k}$
Subject by rater crossed design	Random	$ICC(2,k) = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + (\hat{\sigma}_J^2 + \hat{\sigma}_e^2)/k}$
Subject by rater crossed design	Fixed	$ICC(3,k) = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_e^2/k}$

Table 2: Intraclass Correlation Coefficient (Shrout and Fleiss, 1979, cited in Lane et al., 2023)

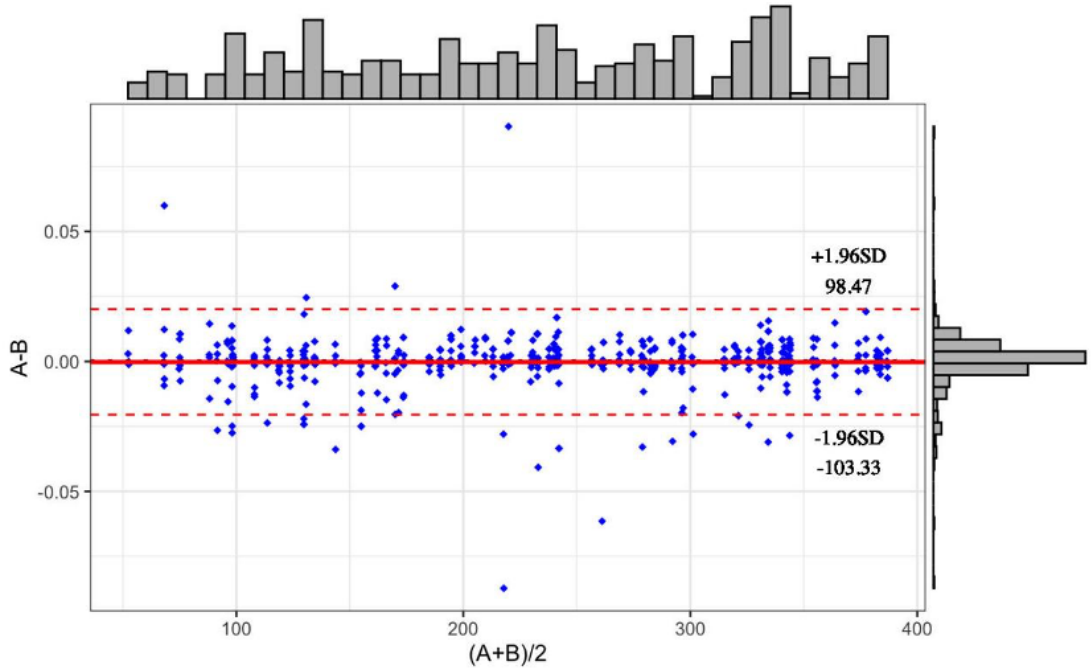


Figure 4: Bland-Altman plot for the data set obtained for the reliability analysis

(8) Results and discussion

For the detailed results of the kappa analysis, see *Sections 4.3.2* and *4.3.3* of the *Reliability Report*.

The results indicated that the overall inter-coder consistency was of a minimal to fair level, and the two coders only reached agreement on the annotations to 3 out of the 12 landmarks. The two coders' annotations to certain landmarks were of high disagreement (e.g., landmark "sn"). Some suggestions for improving the annotation schemes for landmarks that showed low inter-coder consistency were proposed. Additionally, neither the presence of an article (clitic) nor the prosodic structure of the noun exhibits significant effect on annotation accuracy, and no further conclusions can be drawn on the reliability of coding schemes regarding these features. The inter-coder consistency between durations of certain speech segments was also negligible.

Several limitations were identified in the reliability analysis:

A general issue with the current analysis is the limited sample size. Since the current kappa analysis method is only applicable to sufficiently large datasets, the results for certain landmarks may not accurately reflect the actual consistency

between coders in their annotations. It was suggested that the reliability of the annotations for these landmarks should be tested with a larger data set. Additionally, the overall sample size might also be limited, given that there are only two coders and eight files involved in the reliability analysis.

Another significant limitation was that the reliability design was finalized before I gained exposure to prosody in *Phonology*, which rendered the final data set unsuitable for a more in-depth analysis of prosodic features. It was recommended that a more systematic and theoretically grounded design for utterance selection for the re-coding should be proposed.

2.1.5 What remains to be done?

The research project was still in a pilot stage by the end of my participation. The study will continue with the recruitment of young children aged 2 to 5, and similar annotations and data analyses will be performed on the data collected from them. Since young children's speech is still under development, it tends to be messier compared to adult speech. This will present additional challenges in identifying the relevant speech segments, and new annotation as well as analysis schemes will have to be developed. I look forward to following the future progress of this study.

2.1.6 What did I learn

In addition to acquiring extensive theoretical knowledge about the prosodic aspect of children's perception and production of speech, one of the most important things I learned during the annotation stage of *Project 1* was gaining proficiency in speech analysis using *Praat*, especially with reading spectrograms and waveforms. I have also gained some firsthand insights on conducting experiments for collecting linguistic data. Furthermore, I have learned to develop a good work rhythm of lab-based work.

During the reliability analysis stage, I independently researched and studied statistics and, for the first time, applied my programming knowledge in a practical context, which significantly improved my proficiency in R and Python. More

importantly, this was my first experience working on a largely self-directed project. Throughout the process, I learned a great deal on how to work independently, including how to manage my time effectively, how to tackle challenges when things became difficult, and how to cope with the constant insecurity that I might be underperforming.

2.2 Task 2: Children's acquisition of Bassa phonemes

2.2.1 Overview

The second project, which I joined in the latter half of the academic year, was a collaborative study between my supervisor Dr. Benders and Dr. Rodolphe Prosper Mah, a Lecturer in Psycholinguistics and Speech Pathologies at the University of Douala, Cameroon. Their study focused on children's acquisition of certain phonemes in Bassa, a Niger-Congo language.

My participation in this project mainly involved looking for visual stimuli for collecting linguistic data. Towards the end of my internship, I briefly took on a literature review task for this project, but did not make noteworthy progress before the internship concluded.

2.2.2 Theoretical background

Basaa is a narrow Bantu language in the Niger-Congo language family spoken in South, Centre and Littoral regions of Cameroon (Makasso & Lee, 2015). The language exhibits distinct characteristics in its consonant inventory (see *Table 3*) and phonotactics. For example, it features implosives, prenasalized plosives and affricates, and labialized velars.

	Bilabial	Alveolar	Post alveolar	Palatal	Velar	Labialized velar	Uvular	Glottal
Plosive	p	t			k	k^w g^w		
Affricate			ʧ ʤ					
Implosive	ɓ							
Prenasalized	^{m}b	^{n}d	^{n}ʤ		^{ŋ}g			
Nasal	m	n		ɲ	ŋ	$ŋ^w$		
Tap		ɾ̥ ɾ						
Fricative	ɸ β	s			x ɣ		χ	h ɦ
Approximant	w			j				
Lateral approximant		l						

Table 3: Consonant inventory of Bassa (Makasso & Lee, 2015)

As proposed by Jacobson (1968), oppositions that are relatively rare in the world's languages tend to be acquired by children later in their phonological development. Extensive previous studies, including research on Igbo, a Niger-Congo language that shares certain characteristics with Bassa, have documented the sequential acquisition of phonemes and typical substitution patterns. However, significant disagreements persist regarding the specifics of these sequences and patterns (Ferguson & Macken, 1980; Menn, 1975; cited in Nwokah, 1986). Thus, investigating the acquisition patterns of specific phonemes in Bassa could help establish "a consistent line of development" in children's speech (Blache, 1978; cited in Nwokah, 1986).

2.2.3 Stage 1: collecting visual stimuli

My participation in *Project 2* mainly involved collecting visual stimuli to gather linguistic data from Bassa-speaking children in Cameroon.

For this task, I was given a list of stimulus words containing the target phonemes, primarily nouns, with occasional verbs and adjectives. Since the participants in this study were children, the images were intentionally chosen to be cartoon-like, simple, and easy to understand.

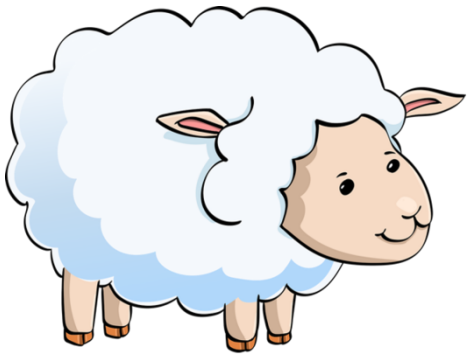
These images were used in two separate research papers. The first paper examines the contrast between labialized and non-labialized phonemes, such as [ŋg^w] versus [ŋ g]. The second paper focused on labial implosives and explosives, as well as prenasalized and syllabic nasals, including [ɓ], [^{n}ʤ], and [n̩k]. Additionally, the stimuli

I collected were utilized for investigating a 5-way contrast involving voicing and nasalization (see *Table 4*).

	/p/	**/ɓ/**	**/ᵐb/**	**/m̀-ɓ/**	**/m/**
Low : a	páá	ɓàχ	ᵐbàs	m̀ -ɓàŋ	man -> ǒ tone
Front : ɛ/e/i	péé	ɓép	ᵐbìŋ	m̀-ɓɛ̀	mèl
Back : ɔ/o/u	pós	ɓɔ̀t	ᵐbɔ́ɾ	m̀-ɓòŋ	mɔ̌

Table 4: 5-way voicing/nasalisation in Bassa

The most significant challenge in this task was the confusion that arose due to cultural differences. For example, when selecting an image for the word ǹtómbá "sheep", I initially used a picture of a typical sheep to my knowledge (see *Figure 5.1*). However, the variety of sheep that is common in Cameroon is quite different from the one I am familiar with (see *Figure 5.2*). During the trial experiment, it was noted that while some children familiar with cartoons could recognize the image, those who were not used to watching cartoons had difficulty identifying it.



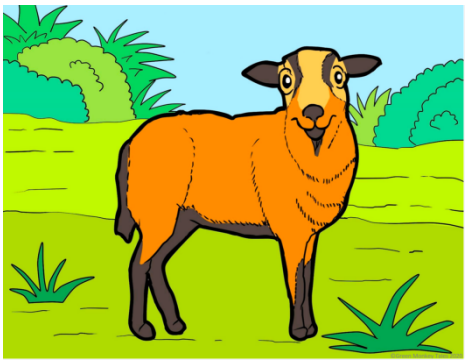


Figure 5.1 and 5.2: Original and revised picture for the word *ǹtómbá* "sheep"

Another challenge was finding copyright-free images for certain words. When suitable images were not available, I created a few of the pictures with *Procreate* (Savage Interactive Pty Ltd, 2023) myself. For instance, *Figure 6* displays a stimulus picture I drew for the word *gʷǒ* "yams".

Figure 6: stimulus picture for the word gʷǒ “yams”

2.2.4 Stage 2: literature review

The second stage of my involvement in this project was to conduct a literature review on specific phonemes and phonological features in various languages worldwide, with the aim of providing insights for the current research. I looked into acoustic and articulatory measures, as well as the acquisition of these phonemes in previous studies across different languages.

I began the task by examining labialized and non-labialized phonemes. However, the vast amount of existing literature turned out to be overwhelming, making it difficult to synthesize a coherent summary out of them. Unfortunately, as my internship ended, the literature review task had to be concluded without significant progress.

2.2.5 What remains to be done?

My progress on the literature review task has been minimal, and it may be taken over by other future interns. The last update I received from Dr. Mah indicated that he was going to the field for data collection. I’m still awaiting his next update and look forward to seeing the progress of the study.

2.2.6 What did I learn

Through the task of collecting visual stimuli, I gained firsthand experience with international collaborations across institutions and cultures, which made me aware of

the importance of effective communication. I also acquired valuable experience in experimental design, particularly in selecting appropriate stimuli for experiments involving children. Additionally, I had the opportunity to apply my interest in drawing by creating some visual stimuli, allowing me to explore how to integrate this skill into academic work.

Through the literature review task, I gained valuable insights into my own shortcomings, particularly my lack of communication with my supervisor and my misjudgment of my time and abilities. Recognizing these areas for improvement will help immensely in helping me enhance my academic skills in the future. Moreover, although my progress was minimal, I gained valuable experience in navigating and extracting research methods and linguistic data from a vast body of literature, while also improving my proficiency in reading academic papers efficiently.

3. Other Internship Components

3.1 Supervision

The supervision during the internship took two forms: weekly progress meetings with Marloes, my supervisor for lab works, and bi-weekly catch-up sessions with my supervisor Dr. Benders along with my fellow intern.

The weekly progress meetings with Marloes primarily centered on Project 1. These meetings mainly involved discussing and refining annotation schemes, helping me monitor and adjust my workflow, and exploring methods for data collection, processing, and analysis. During these sessions, Marloes also generously shared her experience in both her studies and conducting linguistics-related lab work.

The bi-weekly catch-up sessions with Dr. Benders typically involved sharing recent progress and feelings and discussing any questions that had arisen since the last meeting. Additionally, we explored a range of linguistics and academic topics that we found interesting. One particularity of these catch-up sessions was that the agenda was set by my peer intern and me, allowing us to control the content and flow of the meetings.

From both sides of the supervision, I received comprehensive guidance and support.

My supervisors were consistently supportive and encouraging, generously sharing their expertise in linguistics and academic research, and providing practical and mental support. Through supervision, I gained not only knowledge in linguistics and academic research but also valuable skills in managing my time and emotions more reasonably. Most importantly, I acquired crucial insights into teamwork, such as the importance of accountability and the need for providing timely updates.

3.2 Meetings, lectures and seminars

During the internship, I also got the opportunity to attend various lab meetings and linguistics-related lectures and seminars.

One of the lab meetings I have attended was the BiPhon meetings, which stands for "bidirectional phonology and phonetics", named after a theoretical model proposed by Boersma (2011) that reflects the theoretical concept of integrating phonology with phonetics and examining production alongside comprehension. In line with its name, this meeting was designed to offer a platform for both phonologists and phoneticians, as well as researchers focused on production and comprehension to discuss their current and upcoming stages of research — including design, execution, analysis, and publication—with a diverse group possessing extensive expertise in various aspects. Although I didn't attend frequently, I was honored to engage with highly esteemed scholars and gain insights from their expertise. Beyond the sophisticated academic perspectives, I also gained a valuable glimpse into academic life.

Another lab meeting I attended was the Lab Linguistics meeting, a monthly gathering of UvA Speech Lab members. This meeting also served as a platform for lab members to share their research progress and engage in discussions to exchange insights. Unlike the BiPhon meetings, which focus specifically on phonology and phonetics, the Lab Linguistics meetings encompass researchers from all areas of linguistics. This broader scope offers a wider range of insights from diverse fields of study. In addition to the academic knowledge I gained, I have also learned from these meetings about the lab's history, funding, and budgetary matters.

The seminars I attended were primarily hosted by the Amsterdam Center for Language and Communication (ACLC). The ACLC is an organization that aims to foster an inclusive research environment where linguists with diverse theoretical perspectives collaborate to integrate multiple sources of evidence. The seminars covered a broad range of linguistics-related topics, from field research on isolated languages in Rondônia, Brazil, to statistical learning in autism. The ACLC seminars offered a valuable opportunity to explore diverse fields of linguistic research, greatly assisting in the development of my own interests and guiding my choice of future academic pursuits. In addition to the ACLC seminars, I also attended several webinars hosted by LangVIEW and CLaS, following my supervisor’ s recommendations. These webinars, which primarily focused on phonetics and children's language development, were invaluable in deepening my understanding of the projects I was working on.

3.3 “Looking over each other's shoulders”

Another valuable aspect of the internship was the opportunity to engage with my peer intern’s project and share my own progress with him. This mingling, initially suggested by my supervisor as “looking over each other’s shoulders”, involved reviewing each other’s work and offering feedback. My peer intern was very generous in sharing his progress, questions, and insights, as well as providing both assistance and emotional support to my questions and feelings. This experience has not only expanded my knowledge in linguistics and presented intellectual challenges, but also deepened my understanding of the vital role of communication and collaboration in academia.

3.4 Lab experiments

While working at the lab, I also gained the opportunity to observe other researchers’ experiments conducted there. Specifically, due to my interest in neurolinguistics, I inspected two EEG studies, one led by a master’s student investigating the priming effect of grammatical aspects in English, and the other by a PhD candidate exploring color perception among Dutch speakers. Although brief, the experience offered me

invaluable firsthand knowledge of conducting neurolinguistic experiments, which was not available in the classroom. Additionally, as they introduced their projects, I learned that both researchers faced the common challenge of insufficient participants. Therefore, I assisted in recruiting participants to help address this issue. During a supervision session, I raised the concern about participant recruitment and suggested that the relatively low compensation compared to other local institutes might be a factor. At my supervisor's request, I conducted a brief survey on participant compensation for EEG and other linguistics-related experiments at local institutions.

4. Reflections

4.1 What went well

In addition to the invaluable experiences, insights and assistance previously mentioned, I particularly appreciated how well my internship aligned with my coursework. The course on First Language Acquisition, which began right at the start of my internship, provided a strong theoretical foundation for understanding children's language development. My data processing and analysis followed shortly after completing my course in Research Methods and Statistics, allowing me to immediately apply the R programming skills and statistics knowledge I had just acquired. Also, while studying neurolinguistics, I observed EEG experiments in the lab. Similarly, after finishing my introductory course in Python, I promptly applied my knowledge by conducting data processing and analysis for assessing reliability.

Furthermore, the internship has also helped me to practice my social skills and improve my time management. I have also learned how to handle feelings of inadequacy during periods of slow progress and navigate the challenges of multitasking.

4.2 What could have gone better, and what I've learned from that

The internship highlighted several areas for improvement in my abilities. Many tasks could have been performed more effectively and efficiently, but my limited knowledge and skills resulted in less-than-optimal results. For example, my inadequate

proficiency with *Praat* scripts led to significant challenges and delays in data extraction and processing for the reliability analysis. This problem might have been avoided had I completed the course Speech Processing in the first semester. I also recognized that my general knowledge of linguistics is lacking. If I encounter similar issues in the future, I will know to address these gaps by engaging in more self-directed learning.

Significant issues were addressed in my time management skills. As I struggled to balance coursework and internship responsibilities in May, June, and July, I made limited progress on internship tasks during those months. Consequently, my internship extended longer than anticipated. Although this did not lead to significant issues, it deviated considerably from my original plan. Additionally, I discovered that I struggle with utilizing fragmented time effectively and switching swiftly between tasks. This was compounded by my tendency to overestimate my abilities and energy levels. For instance, I underestimated the time needed for the literature review task, resulting in my completing the internship without making any progress on it. This will serve as a reminder for the future to better manage my time and energy, and to create more realistic plans.

I also realized that I was not very proactive in working on projects and tended to complete tasks passively and without deep insight. It wasn't until I observed Tyler working on his project that I saw how much more engaged and proactive he was compared to me. This should be improved as academic research involves more than just completing tasks but requires active thinking and deep insights.

My social anxiety also hindered my performance during the internship. As a result, I struggled with communication, often feeling too intimidated to ask questions or seek help from my supervisors and peer intern. Similarly, I avoided attending the BiPhon and Lab Linguistics meetings in the second semester for the same reason, which led me to miss numerous opportunities to learn directly from esteemed scholars and gain valuable insights. After completing the annotation task, fewer lab-related responsibilities led to a reduced frequency of visits and less active communication with my lab supervisor. Additionally, I tended to avoid sending weekly reports when I felt I

was underperforming, which was detrimental to teamwork and reflected poorly on my accountability. Although I have already made noticeable improvements in my social skills within academic settings, I plan to continue developing in this area by actively engaging in more interactions with others.

4.3 How can the internship help with my future study

During the internship, I discovered my keen interest in phonetics and phonology, particularly in speech prosody and speech perception. Furthermore, I have established a strong interest in academic research and am eager to pursue further academic endeavors. This exploration of personal interests and understanding of academia will serve as a solid foundation for developing my bachelor's thesis topic and may also guide my future academic pursuits.

My skills in statistics and data processing have notably improved. More importantly, as previously mentioned, I've significantly enhanced my ability to conduct speech analysis using spectrograms and waveforms, leading to a deeper understanding of speech characteristics. Additionally, I have become proficient in collecting stimuli and have gained a better grasp of experimental design. These improvements will greatly benefit the research I plan to undertake for my bachelor's thesis.

5. Final Remarks

I would like to express my sincere gratitude to my supervisors for their unwavering guidance and support throughout the internship. I am also grateful to my peer intern for sharing valuable insights and being an inspiring and enjoyable presence. My thanks to Lab Manager Ray Pelupessy and Technician Dirk J. Vet for their support and insights. Lastly, I would like to thank my parents for their invaluable support as programming advisors and for their ongoing encouragement and mental support.

I am deeply grateful for the opportunity to work as a research intern at the UvA Speech Lab. This experience not only allowed me to expand and deepen my knowledge

of linguistics but also gave me my first real glimpse into the world of academia, to be among scholars and witnessing how their brilliance collide. I believe this experience has laid a solid foundation for my future academic pursuits.

Bibliography

Blache, S. E. (1978). *The Acquisition of Distinctive Features*. University Park Press.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. *Linguistik Aktuell*, 33–72. https://doi.org/10.1075/la.180.02boe

Boersma, P., & Weenink, D. (2024). *Praat: doing phonetics by computer*. http://www.praat.org

Cai, R., Boersma, P., Yuen, I., Demuth, K., & Benders, T. (2024, July 2). Prosodic Clitics in English-speaking Children's Speech Production – An Acoustic Study. *Speech Prosody 2024*, Leiden, The Netherlands.

Carter, A., & Gerken, L. (2002). Do children's omissions leave traces?. *Journal of child language*, *31*(3), 561–586. https://doi.org/10.1017/s030500090400621x

Cohen, J. (1960). *A coefficient of agreement for nominal scales*. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Ferguson, C. A., & Macken, M. A. (1980). Phonological development in children: Play and cognition. *Papers and Reports in Child Language Development, 10*, 138-177.

Jakobson, R. (1968). *Child Language, Aphasia and Phonological Universals*. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783111353562

Lane, S. P., Aslinger, E. N., & Shrout, P. E. (2023). Reliability. In H. Cooper, M. N. McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica, 22*(3), 276–282.

Makasso, E.-M., & Lee, S. J. (2015). Basaá. J*ournal of the International Phonetic Association, 45*(1), 71–79. https://www.jstor.org/stable/26352217

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. (2017) Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Proc. Interspeech 2017*, 498-502. https://doi.org/10.21437/Interspeech.2017-1386

Menn, L. (1975). Counter example to "fronting" as a universal of child psychology. *Journal of Child Language, 2*(2), 293–296. https://doi.org/10.1017/S0305000900001124

Nwokah, E. E. (1986). Consonantal substitution patterns in Igbo phonological acquisition. *Language and Speech, 29*(2), 159–176.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Savage Interactive Pty Ltd. (2023). *Procreate* (Version 5.3.5) [Mobile app]. App Store. https://apps.apple.com/us/app/procreate/id425073498

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam

Wheeldon, L., & Lahiri, A. (1997). Prosodic Units in Speech Production. *Journal of Memory and Language, 37*(3), 356–381. https://doi.org/10.1006/jmla.1997.2517

Wynne, H. S. Z., Wheeldon, L., & Lahiri, A. (2018). Compounds, phrases and clitics in connected speech. *Journal of Memory and Language, 98*, 45–58. https://doi.org/10.1016/j.jml.2017.08.001

Yuen, I., Demuth, K., & Shattuck-Hufnagel, S. (2022). Planning of prosodic clitics in Australian English. *Language, Cognition and Neuroscience, 37*(10), 1271–1276. https://doi.org/10.1080/23273798.2022.2060517

Appendix

The *Reliability Report* and relevant files can be found here:
Reliability Report-Yuying Zhu-240911.pdf
Reliability Files_Yuying

The Logbook of the internship can be found here:
Lab Logbook_Yuying Zhu.xlsx