**Developing an Accessible Pipeline for the Forced Alignment and Phonetic Analysis of Spanish Dialectal Data**

**Introduction**

The quantitative analysis of sociophonetic data often relies on a cumbersome pipeline of tools for transcription, segmentation, forced alignment, and acoustic measurement. While tools like the Montreal Forced Aligner (MFA) are powerful, they can present a significant barrier to entry for researchers without computational linguistics training. This project proposes the development of an open-source, user-friendly pipeline designed specifically to streamline this process, in this case for Spanish, with particular attention to handling dialectal variation. The intellectual merit of this project is twofold: it addresses a methodological gap by creating a reproducible, accessible tool for the field, and it contributes to linguistic knowledge by enabling more efficient analysis of less-documented Spanish varieties, such as those from transitional zones like La Mancha.

**Literature**

The field of sociophonetics has been revolutionized by forced alignment technology, which automatically segments audio files at the phone level (Rosenberg & Hirschberg, 2009; McAuliffe et al., 2017). The current gold standard, the MFA, uses acoustic models trained on large corpora to achieve high accuracy. However, its performance can degrade on dialects not well-represented in its training data (Wong, 2022). Furthermore, its command-line interface and complex installation process can be prohibitive. This project builds on the principles of forced alignment but aims to create a more accessible and dialectally-aware workflow. It seeks to integrate automatic speech recognition (ASR) for initial transcription and introduce a novel "Phoneme-to-Phone" tier to explicitly model allophonic variation, a feature not standard in existing aligners.

***Research Question and Hypothesis***

**Research Question:**

Can an integrated, user-friendly tool be developed to accurately perform forced alignment and extract acoustic measures from diverse Spanish dialects, and how does its performance compare to existing solutions like the MFA?

**Hypothesis:**

It is hypothesized that the "Shojus" pipeline can be successfully implemented to produce TextGrids with high temporal accuracy. While its overall phone-level alignment accuracy may be slightly lower than the MFA's on standard varieties, it will demonstrate superior handling of dialect-specific phenomena (e.g., /s/-aspiration) through its explicit phoneme-to-phone modeling, and will be rated as significantly more usable by novice linguists.

**Methods**

***Project Overview***

This project involves the development and evaluation of a software pipeline. The core activities are **Tool Development** and **Validation**.

***Data for Development and Testing***

The primary data for training and testing the pipeline will be the existing corpus of sociolinguistic interviews from Ciudad Real, Madrid, and Andalusia (IRB #27734). This data is ideal because it includes naturalistic speech from the target dialectal varieties.

***Pipeline Architecture and Development***

The "Shoejus" pipeline will be built using a combination of Praat scripting (for acoustic analysis and user interface) and Python (for ASR and alignment logic). The development will proceed in stages to create the following core modules:

* **Figure 1: Proposed "Shoejus" Pipeline Architecture**

[Input: Audio File (.wav)]

|

v

[Module 1: Segmentation & ASR]

- Step 1: Voice Activity Detection to isolate speech.

- Step 2: Automatic Speech Recognition (using a pre-trained Whisper model) to generate an initial transcript.

- Output: A TextGrid with a "Phrases" tier containing the ASR transcript.

|

v

[Module 2: Forced Alignment & Phoneme-to-Phone Mapping]

- Step 1: Tokenize the transcript into words.

- Step 2: Use a pronunciation dictionary for Castilian Spanish to generate a "Phoneme" tier (canonical form).

- Step 3: Apply dialect-specific phonological rules (e.g., /s/ -> [h] in coda position) to generate a "Phone" tier (predicted surface form).

- Step 4: Perform forced alignment of the Phone tier to the audio signal.

- Output: A TextGrid with tiers for Phrases, Words, Phonemes, and Phones, all time-aligned.

|

v

[Module 3: Acoustic Measurement]

- A Praat script that takes the aligned TextGrid and extracts user-specified acoustic measures (e.g., F1/F2, duration, HNR) for selected segments.

- Output: A structured .csv file with the extracted measurements.

***Evaluation Plan***

The success of the project will be measured by two primary outcomes:

1. **Accuracy Evaluation:** The alignment accuracy of "Shoejus" will be quantitatively compared against a manually-corrected "gold standard" baseline for a subset (approx. 5-10 minutes) of the data. Key metrics will include:
   * **Boundary Error:** The average millisecond difference between automated and manual phone boundaries.
   * **Phone Recognition Accuracy:** The proportion of correctly identified phones, with special attention to dialect-sensitive variables like [s] vs. [h].
   * A similar evaluation will be run using the MFA on the same data to provide a benchmark.
2. **Usability Evaluation:** A small group (n=3-5) of linguistics students or researchers with limited computational experience will be asked to use both "Shoejus" and the MFA to process a short audio file. They will then complete a standardized System Usability Scale (SUS) questionnaire and provide qualitative feedback on their experience.

**Analysis**

1. **Accuracy Data:** Boundary error will be analyzed using descriptive statistics (mean, standard deviation). Phone recognition accuracy will be compared between "Shoejus" and MFA using a chi-square test.
2. **Usability Data:** SUS scores for "Shoejus" will be calculated and compared to established benchmarks for usability. Qualitative feedback will be analyzed thematically to identify strengths and weaknesses.

**Discussion**

A successfully implemented "Shoejus" pipeline will demonstrate that it is feasible to create a more accessible and dialectally-sensitive tool for sociophonetic analysis. If the hypothesis is confirmed, it would mean that the tool offers a valuable trade-off: potentially slightly lower raw accuracy than the MFA on some measures, but greater transparency (via the Phoneme/Phone tiers) and significantly lower usability barriers. This would make sophisticated phonetic analysis more accessible to a wider range of researchers. If the tool's accuracy is unacceptably low, the discussion will focus on the specific points of failure (e.g., ASR transcription errors, inadequate phonological rules) and outline a plan for improvement, such as incorporating a pre-trained Spanish MFA model for the initial alignment and then applying a post-processing rule layer.

**Overall Plans and Feasibility**

* **Week 1-2:** Develop and test Module 1 (Segmentation & ASR). Create the core Praat script structure.
* **Week 3-4:** Develop and test Module 2 (Forced Alignment & Phoneme-to-Phone Mapping). This is the most complex stage.
* **Week 5:** Develop Module 3 (Acoustic Measurement extraction). Manually create the "gold standard" dataset for evaluation.
* **Week 6:** Run the accuracy evaluation against the baseline and the MFA. Conduct the usability evaluation with participants.
* **Week 7:** Analyze evaluation data and write the final paper, focusing on the methodological contribution and evaluation results.

**Feasibility:**

This project is ambitious but feasible for a semester. The core components (ASR, Praat scripting) are based on established technologies. The scope is appropriately limited to a proof-of-concept for Spanish. The main risk is the complexity of the forced alignment algorithm, but this can be mitigated by leveraging existing open-source libraries or using a hybrid approach with the MFA as a backend.

***Works Cited***