**Abstract**
Develop a module that uses various techniques to detect errors produced from forced alignment in order to improve its accuracy.

**What is Forced Alignment**
Forced alignment is the process of matching phonetic segments in an audio sample to its corresponding transcription.

Results become less accurate under adverse alignment conditions caused by transcription errors.

**Motivation**
The use of multimedia is growing exponentially. Forced alignment is a vital part of audio indexing and speech recognition.

**Applications**
- Searching within speech audio
- Linguistic research tools
- Video captions for deaf audiences

---

**Example**
The following is an example of detecting an error in the alignment of a speech sample with its corresponding erroneous transcript.

Example Audio: “I have a dream...”
Example Transcript: I have a south dream

**Architecture**

```
Step 1: A speech audio sample is analyzed in order to extract the relevant acoustic information.

Step 2: The audio is aligned with its transcript using the Penn Phonetic Forced Aligner.

Step 3: The alignment is analyzed for temporal anomalies and probabilistic outliers in order to detect errors.

Step 4: Data collected from error detection is input into a support vector machine (SVM) in order to train the error detection module.
```

---

**Error Detection Heuristics**

```
<table>
<thead>
<tr>
<th>Phones:</th>
<th>Duration Analysis:</th>
<th>Confidence Analysis:</th>
<th>Combined Heuristic:</th>
</tr>
</thead>
<tbody>
<tr>
<td>AY</td>
<td>HH AE V</td>
<td>AH</td>
<td>S AW TH</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

Result: I have a south dream...