Chapter 2
Experimental methodology

2.1 Introduction
The aim of this chapter is to present the details of the experimental methodology, used to investigate the impact of vowel space and pharyngealized consonants, on V-to-V coarticulation in three Yemeni Arabic dialects: AYA, HYA, and TYA. In Section 2.2, the details of the speakers of each dialect are presented. In Section 2.3, a comprehensive description of the questionnaire, used to elicit information from Yemeni Arabic speakers on their dialectal background, is discussed. In Section 2.4, the instructions that were given orally to the Yemeni native speakers about the recording procedure are presented. It is followed by the details of the speech materials used for the study in Section 2.5. In Section 2.6, a detailed description of the tools used for recording the speech data is given. In Section 2.7, the methods of acoustic analysis are described. The purpose of vocal tract normalization is discussed in Section 2.8. In Section 2.9, various statistical analyses that were used to test the hypotheses are discussed.

2.2 Speakers
This section details the speakers who participated in the experiments. Eight native speakers (4 female and 4 male) of Abyani, Hadhrami and Ta'izzi, participated in the experimental study; the age of the speakers ranged from 20 to 40, 25 to 33, and 26 to 35 years, respectively. Most of the speakers were Masters and Doctoral students at the English and Foreign Languages University and Osmania University, Hyderabad. The
rest of the speakers were undergraduate students who came to Hyderabad to attend the certificate of proficiency course in English, or for different purposes.

The speakers of all the three Yemeni Arabic dialects have an equal number of males and females. The purpose behind taking an equal number of male and female speakers was to ensure an equal representation of the two genders. All speakers reported no speech or hearing impairment and participated in the experimental study voluntarily. The acoustic data of all the speakers did not include their names. All speakers were informed that the recordings of their voices and their identities would not be made public and that they have the right to withdraw from the study at any time.

2.3 Questionnaire

The speakers were requested to provide information on their dialectal background (see Appendix 2.1 for AYA speakers, Appendix 2.2 for HYA speakers and Appendix 2.3 for TYA speakers). The questionnaire contained questions to elicit information on age, gender, qualifications and the mother tongue (i.e., the dialect that the speaker spoke). It is also included questions on the number of dialects that they were proficient in addition to their mother tongue. It also sought to inquire whether their parents spoke the same dialect or not. The questionnaire was written in Arabic script from right to left. For those who spoke English, the questionnaire was given in English.

2.4 Instructions for recording

The speakers were instructed on the procedure of recording, which entailed the number of repetitions of sentences, the speech rate for reading the data, and so on (for
more details of the instructions, see Appendix 2.4). The instructions for the recording procedure were given verbally.

2.5 Stimuli

This section gives a comprehensive description of the stimuli for all experiments conducted in the present study. The stimuli consisted of meaningful and nonce words for the first and third experiment, whereas the stimuli of the second experiment consisted of only meaningful words (see Appendix 2.5 for Experiment 1, Appendix 2.6 for Experiment 2 and Appendix 2.7 for Experiment 3).

The target words of the stimuli were written in Modern Standard Arabic script without diacritical marks (i.e., without vocalization). The target words of the stimuli were presented on power point slide on a 14 inch Sony laptop. Each slide consisted of the target word embedded in a carrier sentence. The same carrier sentence was used for all target words of each experiment. Four repetitions for each carrier sentence were recorded. For all the three experiments, the first repetition was not analyzed (i.e., it was used only as a trial). The stimuli of the experiments are discussed in the following subsections.

2.5.1 Stimuli for Experiment 1

Speech material for this experiment was elicited from 8 speakers for each dialect, to examine the size of the phonemic inventory, the acoustic vowel space and the F1 and F2 z-scores of each vowel identity in the three Yemeni Arabic dialects. Stimuli consisted of monosyllabic, real and nonce words of the form /CVn/ or /CV:n/. The syllable initial consonant was one of the three stop consonants, /b/, /d/ or /ɡ/, and the
syllable medial was one of the six vowels, /i: i u: u a/. The syllable final was controlled in all the target words with the consonant /n/ (see Appendix 2.5).

The target words were presented to the speakers in Arabic orthography (right to the left), without diacritics, in a fixed carrier sentence: “katab <target word> ɵala:ø marra:t” (he wrote <target word> three times). The total number of the randomized target words for all the three Yemeni Arabic dialects is 1296 (18 target words x 3 repetitions of each target word in the carrier sentence x 8 speakers per dialect x 3 dialects).

### 2.5.2 Stimuli for Experiment 2

For the present study disyllabic words were obtained from 8 speakers, for each dialect, to test the impact of vowel space on V-to-V coarticulation, in the three Yemeni Arabic dialects: AYA, HYA and TYA. The stimuli composed of 19 disyllabic words that occurred in multiple types of syllable structure: /CV: CVC/, /CVC CV:C/, /CV CV:C/, /CVC CVC/ and /CV CVC/. These disyllabic target words were embedded in a carrier sentence: “katab <target word> sabaʕ mara:t” (he wrote <target word> three times). All of the target words were real Arabic words (see Appendix 2.6). The target words were presented to the speakers in Arabic script. The total number of the target words for all the speakers per dialect is 456 (19 test words x 3 times x 8 speakers), excluding the discarded words. Five words were discarded from this experiment because Yemeni speakers uttered them with different pronunciations such as the target word /øyu:mah/.

Vowels in pharyngealized contexts are excluded from the set of stimuli of this experiment since previous studies found that pharyngealized consonants affect the F1
and F2 values of the vowels preceding or following them (Card, 1983; Zawaydeh, 1999; Khattab, Al-Tamimi, and Heselwood, 2006; Mohdar, 2016). Taking this into consideration, a separate experiment (Experiment 3), including only pharyngealized consonants and their nonpharyngealized counterparts, was conducted to investigate the impact of the pharyngealized consonants on the degree of V-to-V coarticulation in the three Yemeni Arabic dialects.

2.5.3 Stimuli for Experiment 3

Speech material for this experiment was acquired from 8 speakers for each dialect. Stimuli consisted of disyllabic words with either one of the six vowels, /i: i u: u a: a/. Each target word consisted of either one of the four pharyngealized consonants, /dˁ/, /sˁ/, /tˁ/ or /ðˁ/ or one of their nonpharyngealized counterparts, /d/, /s/, /t/ or /ð/. Pharyngealized and nonpharyngealized consonants occur either syllable-initially or syllable-finally. Target words were presented to the speakers in a written form, in a fixed carrier sentence: “katab <target word> ɵala:ɵ marra:t” (he wrote <target word> three times).

Each speaker read a set of 28 randomized target words embedded in a fixed carrier sentence. This gives a total of 672 target words per dialect (28 target words x 3 times x 8 speakers). The total number of the target words for all the three Yemeni Arabic dialects is 2016. All target words in the pharyngealized contexts are real words, whereas the target words in the nonpharyngealized contexts are nonce words. The nonce words were used to match the exact structure of real words. (The list of the target words that were embedded in the carrier sentence can be found in Appendix 2.7).
2.6 Recording procedure and tools

The tools used in the recording procedure are described. A portion of the recordings was conducted in a sound-treated room in the phonetics laboratory in the English and Foreign Languages University, using a CSL4200 with a Shure condenser microphone SM 48. The Shure condenser microphone was placed one inch and half distance from the speaker’s lips. Speakers who were unable to come to the lab, were recorded in a quiet room at the researcher’s reside in Hyderabad (India), using a Zoom H4n linear PCM portable recorder.

The speakers of the three dialects read a list of randomized sentences that were displayed as a power point slide on a 14 inch Sony laptop screen. Twenty minutes was allotted for each session. The speakers were allowed to take short breaks. Recordings were saved in a CD as mono channel waveform audio files, with a sampling rate of 44.1 kHz. The recordings were analyzed using Praat software version 5.4.04 (Boersma & Weenink 2014).

2.7 Acoustic analysis

In this section, the acoustic analysis of the data was done using Praat software version 5.4.04 (Boersma & Weenink 2014). Praat allows users to view, measure, and manipulate the spectrograms of sound files. It generates measurable points that follow the formant frequencies of a spectrogram. The acoustic analysis of the data involved segmentation and annotation. Firstly, the segmentation procedure, involved identifying the beginning and the end of each vowel of interest, was manually specified. The locations of the vowels were determined by listening to the audio file
in Praat. Following the segmentation of the data, segmented vowels of the acoustic signal were annotated to identify the target vowels.

After the onset and offset labels of each vowel of interest were determined and annotated manually, the next step was to decide the point (i.e., onset, midpoint, offset) at which the F1 and F2 values of vowels would be measured. For Experiment 1 and 2, vowel formant values were measured at the vowel midpoint, whereas vowel formant values were measured at the vowel onset, midpoint and offset for Experiment 3. The reason behind measuring vowel formants at the three positions was to investigate the effect of pharyngealized consonants on V-to-V coarticulation compared to their nonpharyngealized counterparts.

The F1 and F2 values of all vowels of the target words were measured using Praat running scripts for all the experiments. The script gathered measurements for the F1 and F2 values at the midpoint, onset and offset of each labeled interval. Then the results were exported to a spreadsheet. The values in the spreadsheet were later used for vowel plots and for statistical analyses.

Since the data differs in the three experiments of the study, a sample of waveform, spectrogram and textgrid of one target word embedded in a carrier sentence for each experiment are shown for each dialect. For Experiment 1, a sample waveform, spectrogram and textgrid are displayed for one female native speaker for each dialect. The monosyllabic word is embedded in a carrier sentence: ‘/katab <da:n> ɵala:ɵ marra:t/’. Figure 2.1 shows the F1 and F2 values of the vowel /a:/ in the target word /da:n/. The waveform, textgrid, and spectrogram are shown on the left side for AYA speaker, in the middle for HYA speaker, and on the right side for TYA speaker.
Figure 2.1 shows the F1 and F2 values at the vowel midpoint. As can be seen in Figure 2.1, the F1 value for /a:/ in AYA is 885 Hz and the F2 is 1900 Hz. For HYA, the F1 value for /a:/ is 925 Hz and the F2 is 1936 Hz. The F1 value is 845 Hz and the F2 value is 1850 Hz in TYA. TYA has the lowest F1 value for /a:/ compared with HYA and TYA, whereas HYA has the highest F2 value for /a:/ compared with AYA and TYA.

*Figure 2.1: Waveform, spectrogram and textgrid of the word /da:n/ uttered by a female speaker of AYA, HYA and TYA*
For Experiment 2, a sample waveform, spectrogram and textgrid are displayed for one female native speaker for each dialect. The disyllabic word is embedded in a carrier sentence: ‘/katab <ta:mir> sabaʕ mara:t/’. Figure 2.2 shows the F1 and F2 values of the vowels of each syllable V1=/a:/ and V2=/i/, in the target word /ta:mir/. The waveform and spectrogram of AYA speaker are shown on the left side, the HYA speaker is shown in the middle, and the TYA speaker is shown on the right side.

**Figure 2.1**: Waveform, spectrogram and textgrid of the word /ta:mir/ uttered by a female speaker of AYA, HYA and TYA
As can be observed in Figure 2.2, the three dialects differ in the F1 and F2 values at the vowel midpoint, the following results show the F1 and F2 values in Hz.

<table>
<thead>
<tr>
<th>Dialect</th>
<th>V1 F1</th>
<th>V1 F2</th>
<th>V2 F1</th>
<th>V2 F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AYA</td>
<td>803</td>
<td>1920</td>
<td>761</td>
<td>1724</td>
</tr>
<tr>
<td>HYA</td>
<td>767</td>
<td>2101</td>
<td>700</td>
<td>1911</td>
</tr>
<tr>
<td>TYA</td>
<td>812</td>
<td>1703</td>
<td>625</td>
<td>2004</td>
</tr>
</tbody>
</table>

For AYA, the F1 value of /i/ is raised and the F2 value is lowered. The difference in the F1 value of V1 and the F1 value of V2 is 42 Hz and in the F2 value is 196 Hz. The variance in the F1 value is less than the F2 value. The difference in the F1 value of V1 and the F1 value of V2 is 67 Hz and in the F2 value is 190 Hz in HYA. The variance in the F1 value is less than the F1 value. For TYA, The difference between V1 and V2 is higher in the F1 and F2 values. Based on the above results, TYA has the lowest degree of V-to-V between the two vowel identities /a:/ and /i/.

For Experiment 3, sample waveforms, spectrograms and textgrids are displayed for one male native speaker of AYA. The disyllabic words /tˁariːq/ and /tariːq/ are embedded in a carrier sentence: ‘katab <target word> ḥalaː marraːt’. Figure 2.3 shows the F1 and F2 values of the vowels in the pharyngealized context (leftside), and vowels in the nonpharyngealized context (rightside) at the vowel midpoint in AYA.

As can be seen in in Figure 2.3, the F1 value for V1 /a/ following the pharyngealized consonant /tˁ/ is 732 Hz and the F2 value is 1199 Hz. The F1 value for V2 /iː/ is 660 Hz and the F2 value is 1270 Hz. The F1 value for V1 /a/ following the
nonpharyngealized consonant /t/ is 667 Hz and the F2 value is 1554 Hz. The F1 value for V2 /i:/ is 641 Hz and the F2 value is 1340 Hz. The results show that higher F1 value and lower F2 value (F1=732 Hz and F2=1199 Hz) of the vowel following the pharyngealized consonants compared to vowels in the nonpharyngealized contexts (F1=667 Hz and F2=1554 Hz).

**Figure 2.2:** Waveform, spectrogram and textgrid of the words /tˁariːq/ and /tariːq/ produced by a male speaker of AYA

Figure 2.4 shows the F1 and F2 values of the vowels in the pharyngealized context (leftside), and vowels in the nonpharyngealized context (rightside) at the vowel midpoint in HYA. Figure 2.4 shows that the F1 value for /a/ following the pharyngealized consonant /tˁ/ at the vowel midpoint in HYA is 627 Hz and the F2 value is 1250 Hz. The F1 value for V2 /i:/ is 434 Hz and the F2 value is 1814 Hz. The F1 value for /a/ following the nonpharyngealized consonant /t/ is 597 Hz and the F2
value is 1388 Hz. The F1 value for V2 /iː/ is 410 Hz and the F2 value is 1815 Hz. The effect of pharyngealized consonant /tˁ/ is clear in V1. The pharyngealized consonant /tˁ/ raises F1 and lowers F2.

Figure 2.3: Waveform, spectrogram and textgrid of the words /tˁariːq/ and /tariːq/ produced by a male speaker of HYA

Figure 2.5 shows the F1 and F2 values of the vowels in the pharyngealized context (left side), and vowels in the nonpharyngealized context (right side) at the vowel midpoint in TYA. Figure 2.5 shows that the F1 value for /a/ following the pharyngealized consonant /tˁ/ at the vowel midpoint in TYA is 701 Hz and the F2 value is 1224 Hz. The F1 value for V2 /iː/ is 558 Hz and the F2 value is 1682 Hz. The F1 value for /a/ following the nonpharyngealized consonant /t/ is 693 Hz and the F2 value is 1577 Hz. The F1 value for V2 /iː/ is 541 Hz and the F2 value is 1703 Hz.
Figure 2.4: Waveform, spectrogram and textgrid of the words /tʰariːq/ and /tariːq/ produced by a male speaker of TYA

2.8 Data normalization

The F1 and F2 data were normalized using R software Version 3.2.5 (2016). The vowel formant data were normalized, to reduce the differences in formant data between speakers due to physiological differences. Speakers have vocal tracts of different sizes, causing different formant values for the same vowels (Ladefoged 2005). The F1 and F2 data were normalized using z-score transformation method (Lobanov, 1971). The continuous variables in this study are transformed to z-score transformation (also known as Lobanov’s transformation).
The purpose of normalization is to decrease the differences in the physiological and anatomical variations of the speakers’ vocal tract size that affect the F1 and F2 values. Adank et al. (2004) found that Lobanov’s (1971) z-score transformation reduces variation owing to anatomical variations between speakers and retains phonetic information. Lobanov’s transformation formula used in this study is one of the earlier vowel-extrinsic formulas for vowel normalization, but it stays among the best. The vowel-extrinsic procedure requires the F1 and F2 values of all vowels for each speaker (Adank et al., 2004). This procedure calculates the mean and the standard deviation across all vowels for an individual speaker. Lobanov transformation is not a formant value but a z-score. A z-score of -1 and 1 implies that the vowel is one standard deviation below and above the mean, respectively.

2.9 Statistical analyses

This section provides the details of the statistical tests used to test the hypotheses addressed in chapter 1. The F1 and F2 data were analyzed using R software Version 3.2.5 (2016). The F1 and F2 formant values of the vowels were plotted using ggplot2 package (Halley & Chang, 2015). The results of the z-score normalized vowels were plotted using PhonR packages (McCloy, 2015), to determine the positions of the vowels and their acoustical vowel space in the three Yemeni Arabic dialects.

The linear mixed effects model shows the difference between vowel formants in monosyllabic and disyllabic words. The linear mixed effects model contains fixed and random effects. Fixed effects are associated with the average effect of predictors on the response; the random effects are associated with the covariance structure of the random effects, and of the error term. In many practical applications, estimates of the
random effects are also of interest. Both the fixed and the random effects contribute linearly to the response function.

The linear mixed effects regression was used to measure the first two formants (F1 and F2) values of the vowels of the monosyllabic words at the vowel midpoint for experiment 1. The same statistical test was also utilized to measure the first two formant values of the vowels of the disyllabic words, at the vowel midpoint, to test the degree of V-to-V coarticulation in the three dialects for experiment 2. In addition, the test was used to measure the F1 and F2 formant values of the vowels in the pharyngealized contexts and their non-pharyngealized counterparts for experiment 3. The formant values of the vowels of the disyllabic words were measured at the vowel onset, midpoint and offset, to test the effect of the pharyngealized consonants on the vowels following them and, to measure the degree of V-to-V coarticulation in Yemeni Arabic dialects. The random and fixed variables of all experiments of the study are shown in Table 2.1.
Table 2.1: Random and fixed variables of the experiments

<table>
<thead>
<tr>
<th>Random variables</th>
<th>Fixed variables</th>
<th>Continuous variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speaker</strong></td>
<td>Vowel identity</td>
<td>Formant values</td>
</tr>
<tr>
<td>(eight speakers per dialect: 4 females and 4 males)</td>
<td>HLBV /u:/</td>
<td>F1 and F2</td>
</tr>
<tr>
<td></td>
<td>HSBV /u/</td>
<td></td>
</tr>
<tr>
<td>Repetition</td>
<td>HFLV /i:/</td>
<td>Vowel space area</td>
</tr>
<tr>
<td>(four repetitions of all target words in the carrier sentences)</td>
<td>HFSV /i/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LCLV /a:/</td>
<td>Duration</td>
</tr>
<tr>
<td>Dialect</td>
<td>LCSV /a/</td>
<td></td>
</tr>
<tr>
<td>AYA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The lmerTest was conducted to measure the random and fixed effects. The random effects report speakers and repetitions in AYA, HYA and TYA. In addition, the lmerTest measures the fixed effect of the study, such as vowel identity and dialect. In order to assess the statistical significance of the F1 and F2 mean formant values of the fixed effects of monosyllabic and disyllabic words, the lmerTest-package (Brockhoff, 2016) was used. The lmerTest-package yields Anova function, that produces a data frame akin to the lme4 package with p-values computed from F statistics of types I - III hypotheses. The computation of ANOVA with Kenward-Roger’s approximation is based on the function from pbkrtest package. The results of the test are used, to accept or reject the null hypothesis (Kuznetsova, 2016).
A post hoc Tukey test was utilized to verify the categorical variables that have more than two levels. One-way ANOVA test was also done to measure the significant difference in vowel space and vowel duration in the three dialects.