The Phonetics and Phonology of the Three-Way Laryngeal Contrast in Madurese

Misnadin

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Linguistics

The University of Edinburgh

2016

Declaration

I hereby declare that this thesis has been composed by myself, that the work is all my own work except where I indicate otherwise by proper use of quotes and references, and that the work has not been submitted for any other degree or professional qualification except as specified.

Misnadin

List of Tables

Table of Figures

Acknowledgements

I have been blessed during my PhD study at Edinburgh for having the opportunity and privilege to work with James Kirby and Bert Remijsen. James as my primary supervisor has been very inspiring, motivating, helpful, kind, patient and always generous in his comments, feedback and time to me. Thank you for supporting and guiding me to face and go through the challenges and difficult parts of my PhD journey during these past four years. I am also grateful to Bert as my second supervisor, who has also been very helpful, supportive and generous in his comments and feedback. Thank you, James and Bert, for your attention to details and for showing me how to improve the clarity and conciseness of my writing.

I would also like to thank the people in the Department of Linguistics and English Language especially members of P-Workshop for their support and feedback on my work. Thank you to Alice Turk, who has taught me phonetics during the early years of my PhD study and also for her willingness to be my first year reviewer and also my internal examiner. Thank you to Justin Watkins for his willingness to come to Edinburgh to be my external examiner. I highly appreciate your challenging and critical questions as well as great suggestions during the viva that both of you gave to me. My gratitude also goes to Bob Ladd, Pavel Iosad, Patrick Honeybone and my fellow PhD friends Laura Arnold, Ziming Lu, Daniel Lawrence, Soundes Azzabou-Kacem, Ernisa Marzuki, Maki Kubota, George Starling, Qian Luo and all my other office mates at Room 1.15, Dugald Stewart Building (DSB).

I would also like to thank the staff in the Postgraduate Office of the School of Philosophy, Psychology and Language Sciences (PPLS), Katie Keltie, Toni Noble and Lynsey Buchanan. My gratitude also goes to PPLS Computing Support staff, who always helped me anytime I had problems with my computer and software during the course of my PhD study. Thank you to Stephen Boyd, James Bathgate, and Alisdair Tullo for teaching me computer programming.

I would also like to express my gratitude to the Indonesian Directorate General for Higher Education for providing scholarships for both my Masters and PhD studies. Without the scholarships, studying abroad at a world-class university such as the

University of Edinburgh may have only become a dream for me. Thank you for helping me pursue and realise my dream. My gratitude also goes to Universitas Trunojoyo Madura, the university where I belong, for motivating and giving me the opportunity to pursue my PhD study.

I would like to express my solemn and utmost gratitude to my late mother, who passed away one year after I was back in Edinburgh from my phonetic fieldwork in Madura. I am so sorry for not having been able to see you and say goodbye to you for the last time in your life. Mother, you are always in my heart and will always be missed. My gratitude also goes to my father, who always prays for me, and my brothers and sisters, who always support me.

My grateful thanks go to my wife Wiwik Ratnasari, my son Krisna Adi Pratama and my daughter Tanti Amarta Putri for accompanying me during my PhD study in Edinburgh. I hope what I have done so far will become an inspiration and motivation for you now and in the future. I cannot imagine doing PhD without your accompanying me. Once again, thank you for always listening to my same old stories about the ups and downs of doing a PhD.

Many thanks also go to all my research participants who made the study possible. Thank you to all of my Indonesian friends in Edinburgh, Pak Alex, Bu Alex, Pak Beni, Pak Hari, Pak Totok, Bu Ofita, Bu Maria, and Bu Ari. Big thanks also go to all my friends and colleagues at Universitas Trunojoyo Madura for supporting and motivating me to finish my study.

Abstract

Madurese, a Western Malayo-Polynesian language spoken on the Indonesian island of Madura, exhibits a three-way laryngeal contrast distinguishing between voiced, voiceless unaspirated and voiceless aspirated stops and an unusual consonant-vowel (CV) co-occurrence restriction. The CV co-occurrence restriction is of phonological interest given the patterning of voiceless aspirated stops with voiced stops rather than with voiceless unaspirated stops, raising the question of what phonological feature they may share. Two features have been linked with the CV co-occurrence restriction: Advanced Tongue Root [ATR] and Lowered Larynx [LL]. However, as no evidence of voicing during closure for aspirated stops is observed and no other acoustic measures except voice onset time (VOT), fundamental frequency (F0), frequencies of the first (F1) and the second (F2) formants and closure duration relating to the proposed features have been conducted, it remains an open question which acoustic properties are shared by voiced and aspirated stops.

Three main questions are addressed in the thesis. The first question is what acoustic properties voiced and voiceless aspirated stops share to the exclusion of voiceless unaspirated stops. The second question is whether [ATR] or [LL] accounts for the patterning together of voiceless aspirated stops with voiced stops. The third question is what the implications of the results are for a transparent phonetics-phonology mapping that expects phonological features to have phonetic correlates associated with them. In order to answer the questions, we looked into VOT, closure duration, F0, F1, F2 and a number of spectral measures, i.e. H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. We recorded fifteen speakers of Madurese (8 females, 7 males) reading 188 disyllabic Madurese words embedded in a sentence frame.

The results show that the three-way voicing categories in Madurese have different VOT values. The difference in VOT is robust between voiced stops on the one hand and voiceless unaspirated and voiceless aspirated stops on the other. Albeit statistically significant, the difference in VOT values between voiceless unaspirated and voiceless aspirated stops is relatively small. With regard to closure duration, we found that there is a difference between voiced stops on the one hand and voiceless unaspirated and aspirated stops on the other. We also found that female speakers distinguish F0 for the three categories while male speakers distinguish between F0 for voiced stops on the one hand and voiceless unaspirated and voiceless aspirated stops on the other. The results for spectral measures show that there are no significant differences in H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP between vowels adjacent to voiced and voiceless aspirated stops. In contrast, there are significant differences in these measures between vowels adjacent to voiced and voiceless unaspirated stops and between vowels adjacent to voiceless aspirated and voiceless unaspirated stops.

Regarding the question whether voiced and voiceless aspirated stops share certain acoustic properties, our findings show that they do. The acoustic properties they share are H1*-A1* for both genders, H1*-H2* for females, H1*-A3* and H2*-H4* for males, and CPP for females at vowel onset and for males at vowel midpoint.

However, they do not share such acoustic properties as VOT, closure duration and F0. Voiceless unaspirated and voiceless aspirated stops can be distinguished by VOT, F0 and spectral measures, i.e. H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. However, these two voiceless stop categories have similar closure durations.

As regards the question if [+ATR] or [+LL] might be responsible for the patterning together of voiceless aspirated stops with voiced stops, our findings suggest that either feature appears to be plausible. Acoustic evidence that lends support to the feature [+ATR] includes lower F1 and greater spectral tilt measures, i.e. H1*-A1*, H1*-A3*, H1*-H2* and H2*-H4*, and lower CPP values. Acoustic evidence that supports the feature [+LL] includes lower F1 and greater spectral tilt measures, i.e. H1*-A1*, H1*-A3*, H1*-H2* and H2*-H4*, and lower CPP values. However, the fact that voiceless aspirated stops are voiceless during closure raises a problem for the feature [+ATR] and the fact that F0 for voiceless aspirated stops is higher than for voiced stops also presents a problem for the feature [+LL].

The fact that not all acoustic measures fit in well with either feature is problematic to the idea that the relationship between phonetics and phonology is transparent in the sense that phonological features can be directly transformed into their phonetic correlates. Following the view that not all phonological features may not be expected to be phonetically grounded, for example, when they are related to historical sound change, we hold the idea of a phonetics-phonology mapping which allows for other non-phonetic factors to account for a phonological phenomenon. We also provide historical and loanword evidence which could support that voiceless aspirated stops in Madurese may have derived from earlier voiced stops, which probably retain their historical laryngeal contrast through phonologisation.

Lay Summary

Madurese, a Western Malayo-Polynesian language spoken on the Indonesian island of Madura, shows a three-way laryngeal contrast differentiating between voiced, voiceless unaspirated and voiceless aspirated stops, and an unusual consonant-vowel (CV) co-occurrence restriction. The CV co-occurrence restriction is phonologically interesting because voiceless aspirated stops pattern together with voiced stops rather than with voiceless unaspirated ones in terms of vowels they co-occur with. Two features have been proposed: Advanced Tongue Root [ATR] and Lowered Larynx [LL]. As no evidence of voicing during closure for aspirated stops is observed and no other acoustic measures except voice onset time (VOT), fundamental frequency (F0), frequencies of the first (F1) and second (F2) formants and closure duration relating to the features have been conducted, it remains an open question which acoustic properties voiced and aspirated stops share.

We address three main questions in the thesis: (1) what acoustic properties voiced and voiceless aspirated stops share to the exclusion of voiceless unaspirated stops, (2) whether [ATR] or [LL] accounts for the patterning together of voiceless aspirated stops with voiced stops and (3) what the results mean for a transparent phonetics-phonology mapping, which expects phonological features to have phonetic correlates associated with them. To answer the questions, we looked at VOT, closure duration, F0, F1, F2 and spectral measures, i.e. H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. We recorded fifteen speakers of Madurese (8 females, 7 males) reading 188 disyllabic Madurese words within a sentence frame.

The results show that the three-way voicing categories in Madurese have different VOT values. The difference in VOT is robust between voiced stops on the one hand and voiceless unaspirated and voiceless aspirated stops on the other. Although statistically significant, the VOT difference between voiceless unaspirated and voiceless aspirated stops is not large. There is also a difference in closure duration between voiced stops on the one hand and voiceless unaspirated and aspirated stops on the other. We also found that female speakers distinguish F0 for the three voicing categories while male speakers distinguish between F0 for voiced stops on the one hand and voiceless unaspirated and voiceless aspirated stops on the other. The results for spectral measures show that there are no significant differences in H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP between vowels following voiced and voiceless aspirated stops. However, there are significant differences in those measures between vowels following voiced and voiceless unaspirated stops and between vowels following voiceless aspirated and voiceless unaspirated stops.

Our findings also show that voiced and voiceless aspirated stops share some acoustic properties: H1*-A1* for both genders, H1*-H2* for females, H1*-A3* and H2*-H4* for males and CPP for females at vowel onset and for males at vowel midpoint. However, they do not share VOT, closure duration and F0. While voiceless unaspirated and voiceless aspirated stops have similar closure durations, they can be distinguished by VOT, F0 and spectral measures, i.e. H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. Our findings suggest that either [+ATR] or [+LL] might be

responsible for the patterning together of voiceless aspirated stops with voiced stops. Acoustic evidence that supports the feature [+ATR] is lower F1 and greater spectral tilt measures, i.e. H1*-A1*, H1*-A3*, H1*-H2* and H2*-H4*, and lower CPP values. Acoustic evidence that supports the feature [+LL] is lower F1 and greater spectral tilt measures, i.e. H1*-A1*, H1*-A3*, H1*-H2* and H2*-H4*, and lower CPP values. Since voiceless aspirated stops are voiceless during closure, this raises a problem for the feature [+ATR], and since F0 for voiceless aspirated stops is higher than for voiced stops, this also presents a problem for the feature [+LL].

Due to the fact that not all acoustic measures correspond well to either feature, it is problematic to the idea that the relationship between phonetics and phonology is transparent in the sense that phonological features can have phonetic correlates directly associated with them. Following the view that not all phonological features may not be expected to have phonetic motivation particularly when they are related to historical sound change, we subscribe to the idea of a phonetics-phonology mapping that allows for other non-phonetic factors to account for a phonological phenomenon. We also provide historical and loanword evidence supporting that voiceless aspirated stops in Madurese may have derived from earlier voiced stops, which probably retain their historical laryngeal contrast through phonologisation.

Contents

1 Introduction

1.1 Background of the Study

Madurese is a Western Malayo-Polynesian language spoken primarily on the island of Madura and a number of regions in East Java, Indonesia. The language can be divided into three mutually intelligible dialect regions, Western, Central, and Eastern (Stevens, 1968). Of these three, East Madurese is considered as the standard dialect and is taught from elementary schools to senior high schools across Madura and a number of the regencies along the northern coast of East Java. With the number of its speakers reaching around 14 million, Madurese becomes the fourth largest language spoken in Indonesia following Indonesian, Javanese and Sundanese.

Table 1. Words showing the CV co-occurrence restriction in Madurese

A		B	
[binɛʔ]	‘female’	[patɛʔ]	‘dog’
[pʰikʰɤl]	‘robber’	[tɔdiʔ]	‘knife’
[dɤpɔr]	‘kitchen’	[cɛlɔʔ]	‘sour’
[tʰusah]	‘sin’	[kəʈʰ:ɤŋ]	‘banana’
[ɟikar]	‘horse cart’	[nasɛʔ]	‘rice’
[cʰɤcʰɤl]	‘try’	[sɔʈɔk]	‘push’
[gɤɟi]	‘salary’	[ɲɛlɔ]	‘painful’
[kʰubɤŋ]	‘break in’	[ŋɛtək]	‘hide’

Madurese is often described as having eight surface vowels [a, ɛ, ə, ɔ, ɤ, i, ɨ, u] and as contrasting voiced, voiceless unaspirated and voiceless aspirated stops at five places of articulation (Cohn, 1993a; Davies, 2010; Stevens, 1968, 1980). Two particularly interesting aspects of Madurese are the fact that the language shows a surface three-way voicing contrast among its stop series (voiced, voiceless unaspirated and voiceless aspirated) and a robust consonant-vowel (CV) co-occurrence restriction (Stevens, 1968, 1980; Cohn, 1993a). The three-way contrast is areally unusual given that its related languages such Javanese and Sundanese exhibit two-way contrasts in their stop consonants and do not show CV co-occurrence restrictions.

The expression of laryngeal contrast in Madurese is typologically unusual given that its voiced stops /b, d, ɖ, ɟ, ɡ/ and voiceless aspirated stops /pʰ, tʰ, ʈʰ, cʰ, kʰ/ are always followed by the high vowels [ɤ, i, ɨ, u] as in Table 1 (A) while its voiceless unaspirated stops /p, t, ʈ, c, k/ and other consonants such as /m, n, s, ɲ, ŋ/ are always followed by the non-high vowels [a, ɛ, ə, ɔ] as in Table 1 (B).

The CV co-occurrence restriction is of phonological interest given the patterning of voiced stops with voiceless aspirated stops, raising the question of what phonological feature they might share. That is to say, it is natural for voiceless aspirated stops to pattern with voiceless unaspirated stops because they are voiceless during closure. However, the fact that voiced stops pattern together with voiceless aspirated stops suggests that they belong to the same natural class and therefore share a feature.

Previous studies (Cohn, 1993a, 1993b; Cohn & Lockwood, 1994; Trigo, 1991) have associated the CV co-occurrence restriction with two phonological features of the preceding consonants: Advanced Tongue Root ([ATR]) and Lowered Larynx ([LL]). The proposals tacitly assume that there is an articulatory gesture (advancement of the tongue root and/or lowering of the larynx) shared by the voiced and voiceless aspirated stops, and this is why these segments pattern together phonologically. This prediction is consistent with theories where a distinctive feature characterises a restricted set of possible phonetic realisations (e.g. Chomsky and Halle, 1968), and so segments which share a feature are predicted to share at least some aspects of the phonetic implementation of that feature (Keating, 1990; Pierrehumbert, 1990). As they are usually defined, the features [ATR] and [LL] seem to assume a close, relatively *transparent* association between the phonological feature and its articulatory implementation. If this could be the case, we would expect that segments characterised by such features might also share certain acoustic correlates associated with the physical gesture of either tongue root advancement or larynx lowering.

However, it has also been questioned whether phonetic implementation can always be predictable from phonological features. For one thing, features like [±continuant] do not have a clear articulatory basis: for example, there is no single gesture shared by [f], [s] and [x], yet they often pattern together phonologically (Clements & Hallé, 2010). There are also well-known cases of ‘unnatural’ rules such as the well-known Indo-European ‘ruki rule’, where [s] became [š] before the segments /r u k i/. It is not clear that a phonetically transparent account of such rules is possible. This subsequently leads us to consider a second possible option which is that phonological features can have a more *abstract* relationship to their phonetic realisation (Anderson, 1981; Bach & Harms, 1972; Mielke, 2004). This means that there are some cases in which phonological features do not have any phonetic motivation, which may be attributed to historical sound change or other non-phonetic factors instead (Blevins, 2004; Hyman, 2001; Ohala, 2005).

Madurese is an interesting language to consider in this regard, because the features that have been proposed to account for its CV co-occurrence restriction are fundamentally articulatory in nature. However, this phonological pattern does not seem to be very common cross-linguistically, raising the possibility that these segments pattern together for other reasons, i.e. the pattern is not transparently phonetic. The goal of this dissertation is therefore to study the voiced and voiceless aspirated stops of Madurese, to see if they share any acoustic properties that would suggest a shared articulatory correlate. This will permit us to evaluate the proposed phonological features ([ATR] and [LL]) that have been associated with the CV co-occurrence restriction in Madurese based on new acoustic data. The results of the phonetic analysis are also expected to contribute to the discussion about the relationship between phonological features and phonetic implementation.

The remainder of this chapter addresses two major issues. First, we establish the phonological nature of the Madurese laryngeal contrast, arguing that it is best analysed as a three-way system (Section 1.2). Next, we review previous proposals that have been made regarding the nature of the phonological feature shared by voiced and voiceless aspirated stops in Madurese, and lay out their articulatory and acoustic correlates (Section 1.3). This serves as the basis for the specific research questions posed in this thesis (Section 1.4).

1.2 Establishing the Laryngeal Contrast in Madurese Stops

This section establishes the phonological status of the laryngeal contrast in Madurese stops. We provide phonological evidence that Madurese can be best described as a language with a three-way phonological contrast in stop consonants. The phonological evidence includes consonant-vowel interactions, vowel harmony processes and morphophonemic processes. All this evidence will also be used as phonological evidence in support of the proposal that consonants trigger vowel height alternations rather than vowels trigger consonant allophony in Madurese.

1.2.1 The description of the laryngeal contrast in Madurese

Two related questions need to be addressed in relation to the laryngeal contrast in Madurese: (1) how should the contrast be better described, a two-way or a three-way phonological contrast? (2) what are the phonological consequences for favouring one type of contrast over another? In the following, we argue that despite previous studies suggesting that the surface phonetic distribution differs from that of a 'classic' three-way laryngeal contrast language like Thai, the preferred phonological analysis for Madurese is one with three stop phonemes and four vowel phonemes.

1.2.1.1 Is there a phonological three-way contrast in Madurese?

Three types of VOT were observed by Lisker and Abramson (1964): voicing begins before the release of the stop, voicing begins after the release and voicing lags behind the release of the stops, corresponding respectively to voiced, voiceless unaspirated and voiceless aspirated stops. As has been suggested in a number of studies (Cohn, 1993a; Cohn & Ham, 1998; Cohn & Lockwood, 1994; Stevens, 1968, 1980, 1991), Madurese also has three stop categories, namely voiced, voiceless unaspirated and voiceless aspirated. This being so, the contrast in Madurese typologically appears to bear some resemblance to the voicing contrast in languages such as Thai and East Armenian (Lisker & Abramson 1964).

However, there are three reasons why it is tempting to think that Madurese may instead have only a two-way laryngeal contrast distinguishing between voiced and voiceless stops. First, the VOT values between the two voiceless categories do not exhibit the typical distribution characterising a language with a three-way laryngeal contrast (Cohn & Ham, 1998; Cohn & Lockwood, 1994; Misnadin, Kirby, & Remijsen, 2015). In particular, although statistically significant, the VOT difference between voiceless unaspirated and voiceless aspirated stops reported by Cohn and Lockwood (1994) is not so large, i.e. on average 11 ms and 25 ms respectively (Cohn & Lockwood, 1994, p. 76).

The second reason why it is tempting to consider Madurese as having a two-way contrast is related to the fact that voiceless unaspirated stops only occur before non-high vowels, while voiceless aspirated stops only occur before high vowels. That is, one could analyse the occurrence of each stop type as conditioned by different vocalic environments and in this way they should be considered allophonic. It is also possible that the difference in their VOT values may simply reflect variations due to the different vowel types which follow them. In fact, there is some evidence that VOT also depends on vowel quality: VOT is longer before tense vowels and shorter before lax vowels (Port & Rotunno, 1979). There is also evidence that VOT is longer before high vowels than before low vowels in other languages with prevoiced stops such as Hungarian (Gósy, 2001) and Canadian French (Nearey & Rochet, 1994).

The third reason is concerned with the fact that there is no minimal triplet of stops exemplifying the three-way contrast in Madurese. The true distinction is only between voiced and voiceless aspirated stops because this is the only contrast where true minimal sets can be found, for example [bɤrɤ] ‘swell’ vs. [pʰɤrɤ] ‘lung’, [bɤlɤ] ‘tell’ vs. [pʰɤlɤ] ‘family’ and [dɤlɨm] ‘deep’ vs. [tʰɤlɨm] ‘residence’. In contrast, we cannot find minimal pairs which show the distinction either between voiced and voiceless unaspirated stops or between voiceless unaspirated and voiceless aspirated stops due to the CV co-occurrence restriction. Recall that voiced and voiceless aspirated stops only co-occur with high vowels while voiceless unaspirated stops only co-occur with non-high vowels, for example, voiced vs. voiceless unaspirated stops, [bɤrɤʔ] ‘west’ vs. [paraʔ] ‘almost’ and [gɤgɤn] ‘dumb’ vs. [kakan] ‘eat’ and voiceless aspirated vs. voiceless unaspirated stops, [kʰɤlɤ] ‘pole’ vs. [kala] ‘lose’ and [cʰɤlɤ] ‘net’ vs. [cala] ‘defective’. As we can see, they are not minimal pairs because the difference not only resides in the stops but also in the following vowels.

1.2.1.2 Assessing different proposals regarding the Madurese laryngeal contrast

As we can only find voiceless aspirated stops before high vowels and voiceless unaspirated stops before non-high vowels, we might argue that the two voiceless stop categories are allophonic. That is, they do not belong to phonologically different voicing categories since they may be conditioned by, or depend on, the following vowel. Thus, with this case in mind, we could argue that the stop consonants that we observe in Madurese are actually not stops with a three-way laryngeal contrast but ones with a two-way distinction, distinguishing between voiced and voiceless stops.

If this could be the case, voiceless unaspirated and voiceless aspirated stops in Madurese seem similar to voiceless aspirated and unaspirated stops in English, which are also allophonic in certain environments. The difference probably lies in the fact that in Madurese voiceless aspirated and unaspirated stops occur in any position in word as long as they co-occur with the ‘right’ vowels. In contrast, English voiceless unaspirated and voiceless aspirated stops can be followed by any vowel type but their occurrences are not as free as those in Madurese particularly in terms of position in word. In English, voiceless aspirated stops only occur in the stressed syllable onset while voiceless unaspirated stops occur elsewhere (see e.g. Iverson & Ahn, 2007; Iverson & Salmons, 1995).

In the following we will consider three scenarios with respect to whether Madurese has a two- or three-way laryngeal contrast in its stops and decide which scenario is more parsimonious phonologically and can best describe the laryngeal system of Madurese. The scenarios are that Madurese may have (1) a two-way contrast distinguishing between voiced and voiceless stops, (2) a two-way maximum contrast distinguishing between voiced and voiceless aspirated stops and (3) a three-way contrast distinguishing between voiced, voiceless unaspirated and voiceless aspirated stops. The first two scenarios assume that there are two underlying stop consonants (i.e. voiceless and voiced stops for the first scenario, and voiced and voiceless aspirated in the second scenario) and eight underlying vowels (a, ɛ, ə, ɔ, ɤ, i, ɨ, u). The third scenario assumes that there are three underlying consonants (voiced, voiceless unaspirated and voiceless aspirated) and four underlying vowels (a, ɛ, ə, ɔ). The third scenario has been suggested by other authors in previous studies (e.g. Stevens, 1968, 1980; Trigo, 1991; Cohn, 1993b) and is considered the default here. Therefore, in the following we only focus on assessing the first two scenarios.

Scenario 1: 8 vowels, voiced and voiceless stops. Suppose Madurese has a two-way voicing contrast as in the first scenario, the contrast that may describe the system is that the language may have underlying voiced and voiceless stops. By this account, voiceless stops are assumed to have two allophones, i.e. voiceless unaspirated and voiceless aspirated stops, occurring in complementary distribution. That is, voiceless unaspirated stops only occur before non-high vowels while voiceless aspirated stops only occur before high vowels. This can be schematised as in (1) below.

(1) C [-voice] → [+asp] /__ (+high vowels), where C = stop consonants

Considering voiceless stops having two allophones such as these bears a resemblance to some extent to allophonic voiceless unaspirated and voiceless aspirated stops in English. By this analysis, we have to consider that the vowels affect the consonants as opposed to the other way around. Consequently, we may not need to think about what phonological feature voiced and voiceless aspirated stops share because they just happen to be two different voicing categories with no effects on vowels. Hence, it may be simply due to some sort of phonological coincidence that voiced and voiceless aspirated stops pattern together in this manner.

Considering Madurese as having a two-way voicing contrast such as the first scenario will imply that Madurese has eight underlying vowels. That is, we would have to view that the eight vowels (a, ɛ, ə, ɔ, ɤ, i, ɨ, u) are all phonemic. They are not allophones of the four 'underlying' non-high vowels as has been suggested, for example, in Stevens 1968, Cohn 1993a and Cohn 1993b. If this could be the case, we do not need to find out what phonological feature is shared by voiced and voiceless aspirated stops for triggering vowel raising as there is no vowel raising in the first place. Therefore, the issue of feature spreading becomes no longer relevant here.

In addition, if we hold the assumption that there is only a two-way phonological contrast in stops and hence eight vowel phonemes in Madurese, we could argue that what we have observed with respect to voicing and aspiration and their relationships to vowel height is not really unusual in the language, either areally or typologically. In this case, the laryngeal contrast in Madurese would be similar to its related languages such as Javanese and Sundanese, both of which show a two-way contrast, i.e. tense versus lax stops for the former and voiced versus voiceless stops for the latter. The question is whether this assumption is in line with the results of acoustic measures and more importantly whether it is also consistent with the phonological facts of Madurese, one of which is that non-high vowels only occur in word-initial position while high vowels never occur in this position.

Scenario 2: 8 vowels, voiced and aspirated stops. A second possible scenario is that there may be a two-way maximum contrast in Madurese, distinguishing between underlying voiced and voiceless aspirated stops (Brett Baker, personal communication). As it stands, the contrast in the second scenario is different from the account in the first scenario, which proposes that the two-way contrast in Madurese is between underlying voiced and voiceless stops, where voiceless stops can be realised as voiceless unaspirated and voiceless aspirated stops. They are, however, similar in their assumptions that Madurese has eight underlying vowels. Specifically, the two-way maximum contrast proposes that voiced stops and voiceless unaspirated stops are allophonic; voiced stops are underlying and the voiceless unaspirated stops are the surface variant that occurs before non-high vowels. This can be represented as in the following rule in (2) below.

(2) C [+voice] → [-voice] /__ (-high vowels), where C = stop consonants

Like the first scenario, this proposal assumes no feature spreading or consonant-vowel interactions whatsoever. However, we would need to explain why voiced stops become voiceless before non-high vowels, which is not trivial either phonetically or phonologically.

The assumption that there may be only a two-way contrast in Madurese stops would make sense if we consider that the occurrences of voiceless unaspirated and voiceless aspirated stops as in the first scenario are considered environment-dependent. The question is whether the vowels with a height difference following the consonants can be considered as a phonological environment here. Furthermore, considering the vowels as the environment which predicts consonant allophony, i.e. high vowels predict voiced and voiceless aspirated stops while non-high vowels predict voiceless unaspirated stops, is also phonologically problematic. This is because it cannot explain a number of phonological phenomena in Madurese such as the distribution of only low vowels word-initially; vowel height harmony; transparent consonants; the behaviour of /s/; and non-high vowel suffixes, as discussed in the following section.

1.2.2 Phonological evidence against a two-way contrast in Madurese stops

Distribution of vowels word-initially. High vowels [i u ɤ ɨ] never occur in absolute word-initial position. This restriction is mysterious on an account that posits 8 underlying vowels, but if high vowels are surface allophones of non-high vowels and are triggered by the presence of a voiced or aspirated consonant, this distributional restriction makes more sense.

Vowel height harmony: transparent consonants. The consonants /l/, /r/ and /ʔ/, when occurring in word-medial position, are transparent in the sense that the height of the vowels following them depends on the height of the vowels preceding them (Stevens, 1968; Trigo, 1991). That is, if the vowels preceding them are high, the vowels that follow them will also be high. Some examples are shown in (3) below.

(3)	[bɤrɤ]	‘swell’
	[bɤʔɤ]	‘flood’
	[bulu]	‘feather’
	[kʰɤru]	‘scratch’
	[kʰulɤ]	‘sugar’
	[tʰɤʔɤr]	‘eat’
	[tʰuʔum]	‘distribute’

On the other hand, if the vowels before l, r and ʔ are non-high, the vowels following them will also be realised as non-high. Some examples are shown in (4) below.

(4)	[lɛʔɛr]	‘neck’
	[paʔaʔ]	‘chisel’
	[pɛlak]	‘kind’
	[pɔla]	‘probably’
	[pɔrak]	‘cleave’
	[raʔa]	‘water germ’
	[tɔrɔk]	‘deficit’

Vowel height harmony: /s/. Another aspect which needs to be mentioned here is the behaviour of /s/. In word-initial position, /s/ behaves in the same manner as the other voiceless stops, nasal consonants and liquids. However, it behaves differently when it occurs in intervocalic position. In this position, the height of the vowels following

/s/ depends on whether /s/ occurs morpheme-internally or at a morpheme boundary (Cohn, 1993b; Stevens, 1968). If it occurs morpheme-internally, it co-occurs with non-high vowels, for example [kasar] 'rude', [tʰisa] 'village', [sɛsɛt] 'dragonfly', and [pɛsːɛ] 'money'. However, if /s/ occurs at a morpheme boundary, the vowel following /s/ is determined by the vowel height preceding it, as shown in (5) below.

(5)	[bɤlɨs]+an	→	[bɤlɨsɤn]	'reply'
	[kʰɤrus]+an	→	[kʰɤrusɤn]	'selling faster'
	[pʰuŋkɔs]+an	→	[pʰuŋkɔsan]	'package'
	[tɔlɛs]+an	→	[tɔlɛsan]	'writing'

The vowel height harmony shown in (3), (4) and (5) above would only make more sense if we hold the idea that it is the consonants that determine the phonological environment conditioning vowel height, i.e. vowel allophony instead of the other way around, i.e. consonant allophony.

Morphophonemic processes: Nasal Substitution, vowel deletion, and aspiration. Other evidence that supports the idea that consonant type triggers vowel alternations, rather than vice versa, comes from vowel height alternation as a result of affixation. This can be seen in morphophonemic alternation involving a nasal prefix 'N' indicating the 'actor voice' form of verbs (Cohn, 1993, p. 110; Davies, 2010, p. 32; Stevens, 1991, p. 363), a process known as Nasal Substitution. In this case, when the prefix 'N' replaces an underlying voiced or voiceless aspirated stop with its homorganic nasal equivalent, the following vowel subsequently becomes non-high, as exemplified in (6) below.

(6)	N+[bɤca]	→	[maca]	'read'
	N+[bɤlɨs]	→	[maləs]	'reply'
	N+[bɤgi]	→	[magi]	'divide up'
	N+[bɨlːi]	→	[məlːɛ]	'buy'
	N+[pʰɤlik]	→	[malɛʔ]	'turn over'
	N+[pʰuruk]	→	[mɔrɔk]	'teach'
	N+[tʰutʰ:uʔ]	→	[nɔtʰːuʔ]	'finger-point'
	N+[cʰucʰːu]	→	[ɲɔcʰːu]	'push'

Other phonological evidence in support of the idea that it is the consonants that trigger vowel harmony comes from a process called vowel deletion. Vowel deletion, which is optional and appears to be dialect-specific in Madurese, can occur in an open first syllable of a word consisting of at least three syllables. That is, the vowel of the word in the first syllable can undergo an optional deletion if it is preceded by a consonant and followed by an approximant, a liquid, or a glide (Davies, 2010; Stevens, 1968). As we can see in (7), even after the vowel in the first syllable is deleted and therefore in the absence of the preceding vowel, the vowel following the transparent consonants /l, r/ does not change. This indicates that the harmony trigger is the consonant preceding the transparent consonants, rather than the vowel itself.

(7) bɤlɤntʰɤ → [blɤntʰɤ] ‘the Dutch’
parabɤn → [prabɤn] ‘virgin’
paraɟɤ → [praɟɤ] ‘make bigger’
paraɔ → [praɔ] ‘boat’
salamət → [slamət] ‘safe’
sakalaŋkɔŋ → [skalaŋkɔŋ] ‘thank you’
saratɔs → [sratɔs] ‘a hundred’

A third process that supports the idea of phonologically condition vowel height alternations is aspiration as a result of a morphophonemic process. This type of aspiration occurs when a root-final stop, which is always voiceless unaspirated in Madurese, meets with a vowel-initial suffix, which will necessarily begin with a non-high vowel. In this position, the voiceless unaspirated root-final stop will be realised as voiceless aspirated stops and the non-high vowel suffix will subsequently be realised as a high vowel. Examples of this morphophonemic aspiration are shown in (8) below. The suffix *-ɛ* is attached to a noun to form an imperative verb while the suffix *-an* is attached to a verb to form a noun.

(8) [ɔbat] + ɛ → [ɔbatʰi] ‘treat’
[karɛt] + ɛ → [karɛtʰi] ‘tie’
[pɛkɔt] + ɛ → [pɛkɔtʰi] ‘entangle’
[tɔtɔp] + ɛ → [tɔtɔpʰi] ‘cover’
[ɟɤwɤp] + an → [ɟɤwɤpʰɤn] ‘answer’
[kərap] + an → [kərapʰɤn] ‘(bull) race’
[səmprɔt] + an → [səmprɔtʰɤn] ‘spray’
[sɛkɔt] + an → [sɛkɔtʰɤn] ‘tailoring’

The examples in (8) above also provide further evidence that it is the consonants which trigger the vowel height alternation, as opposed to vice versa. This is because the suffixes that underlyingly begin with non-high vowels become high vowels as the root-final stops become aspirated. In this case, it appears that final stops in (8) are in fact underlyingly aspirated and that aspiration becomes neutralised word-finally.

Thus, of the three possible scenarios, subscribing to the idea that Madurese has three stop phonemes (voiced, voiceless unaspirated and voiceless aspirated) and four vowel phonemes (a, ɛ, ə, ɔ), i.e. the third scenario, can best account for the laryngeal system in the language. Proposing that Madurese has only a two-way phonological contrast fails to explain the robust consonant-vowel interaction as well as feature spreading associated with the prevocalic consonants. Put differently, the two-way contrast proposal seems to simplify the description of the consonants, but it complicates the analysis of the vowels, the vowel harmony process and the morphophonemic alternation. In addition, it does not need to account for the phonological patterning of voiced and voiceless aspirated stops since it reduces the CV co-occurrence restriction in Madurese to a trivial phonological phenomenon that does not require a further phonological analysis.

1.3 A Review of Features Proposed to Explain the CV Co-occurrence Restriction in Madurese

The previous section established that the CV co-occurrence restriction in Madurese is best analysed as being triggered by some property of phonologically voiced and voiceless aspirated stops. We must then ask the question of what phonological feature(s) they might share. For phonetic reasons, we might think that it is more natural for voiceless aspirated stops to pattern with voiceless unaspirated stops instead of with voiced stops, since they are phonetically voiceless during closure. However, the fact that voiced stops and voiceless aspirated stops pattern together in that they are only followed by high vowels suggests that they belong to the same natural class and therefore can share a phonological feature that distinguishes them from the other consonants.

In relation to this, there have been a number of proposals that attempt to account for the consonant-vowel interactions or vowel-height alternations in Madurese. In the following, we will discuss the proposals, which include: a tense-lax distinction (Section 1.3.1), a register system (Section 1.3.2), a feature [advanced tongue root] (ATR, Section 1.3.3) and a feature [lowered larynx] (LL, Section 1.3.4).

1.3.1 Tense-lax

Stevens (1980, pp. 136-137) argues that Madurese only has four underlying vowels /e, a, ə, ɔ/. These vowels become tense or high when they occur following voiced and voiceless aspirated stops. However, they remain lax when they occur in word-initial position and after the other consonants. Specifically, he suggests that the four vowels surface as high [i, ɤ, ɨ, u] and non-high [ɛ, a, ə, ɔ] and that they can be characterised primarily in terms of the tense-lax distinction. That is, the high vowels have the feature [tense] while the non-high vowels have the feature [lax].

However, Cohn (1993a) rules out the tense-lax account proposed by Stevens (1980) given that this account appears to be contradictory with the observed phonetic patterns. She argues that the feature realised in the vowels is not consistent with the consonantal feature, assuming the consonants trigger the vowel height alternation. This is because considering voiced stops as having the feature [lax] should expect that vowels following voiced stops are also realised as lax vowels, i.e. non-high vowels. However, this is not what we observe in Madurese since voiced stops only co-occur with tense vowels, i.e. high vowels. Similarly, considering voiceless unaspirated stops as having the feature [tense] should also expect that vowels that co-occur with them are tense vowels. The fact is that voiceless unaspirated stops only co-occur with non-high vowels, which have the feature [lax].

In this case, the only consistency in feature spreading we may observe in Madurese by the tense versus lax proposal is that voiceless aspirated stops assumed to have the feature [tense] are followed by high vowels which also have the feature [tense]. However, as Cohn (1993a) also points out, this is not really the case since voiceless aspirated stops has the feature [Heightened Subglottal Pressure] rather than [tense], assuming that such a distinction is based on *The Sound Pattern of English* (SPE) tradition of Chomsky and Halle (1968). In conclusion, the tense-lax distinction cannot account for the patterning together of voiced stops and voiceless aspirated stops with [tense] vowels and voiceless unaspirated stops with [lax] vowels.

1.3.2 Register

Another proposal which attempts to account for the vowel height alternation in Madurese is a register system (Trigo, 1991; Cohn, 1993a; Cohn & Lockwood, 1994). In the literature on the phonology of Southeast Asian languages, the term *register* is primarily used to refer to two sets of vowels that can be distinguished by differences in their voice quality, fundamental frequency, vowel quality, intensity and vowel duration (Abramson & Luangthongkum, 2009; Brunelle & Kirby, 2016; Gregerson, 1976; Henderson, 1952; Kirby & Brunelle, in press; Wayland & Jongman, 2003). A register system in Southeast Asian languages perspective is defined as one where a historical voicing contrast, which is now neutralised, is synchronically manifested on the vowels by a constellation of such phonetic properties as voice quality, fundamental frequency, vowel quality, intensity and vowel duration.

A number of terms have been used to label the two vowel types such as first register versus second register, head register versus chest register, and tense register versus breathy register. For the sake of convenience, in this dissertation we refer to the vowel set which has phonetic features such as clear or creaky voice quality, higher fundamental frequency, higher F1, and tendency to diphthongise as 'upper register' and the other set which has phonetic features such as breathy voice, lower fundamental frequency, lower F1 and tendency to centralise as 'lower register'.

1.3.2.1 Register as a phonological system

As a phonological concept, the term 'register' was first introduced and used by Henderson (1952) in the description on the vowel system of standard Khmer, the national language of Cambodia. She classified the vowels of the language into upper register and lower register. She characterised the upper register as having a number of phonetic properties such as modal voice quality and higher pitch while the lower register as breathy voice quality and lower pitch. The lower register was also described as being produced with a lowered larynx and sometimes accompanied by a widening of the nostrils.

Henderson (1952) also observed that the two registers had different vowel quality, i.e. the upper register appeared to be more open than the lower register. In essence, Henderson suggests that the primary feature of a register system is contrastive voice quality, namely 'normal' or 'head' versus 'breathy' or 'sepulchral' while pitch and vowel quality can be regarded as secondary or tertiary properties. However, it is important to note that Henderson later reported that modern standard Khmer does not really have registers in the sense that there is a synchronic dichotomy of phonation type in its vowel system as she described in her earlier study on the language. She agreed with Huffman (1978) and other linguists working on the language that Khmer is not a register language. However, it is clear that it was a historical distinction.

Table 2. Huffman's (1976) classification on fifteen Mon-Khmer languages

No.	Type	Consonant Contrast	Vowel	Example
1.	*Conservative*	/p t c k/ vs. /b d ɟ g/	Little or no differentiation in vowels	Loven, Lawa, Stieng, Brao
2.	*Transitional*	/p' t' c' k'/ vs. /p t c k/	Sub-phonemic register distinction in vowels	Alak, Souei, Ngeʔ, Mal
3.	*Register*	No contrast in initial consonants	Phonemic vowel register; retention of sub-phonemic differentiation in stops vis-à-vis register	Kuy, Chaobon, Chong, Bru, Mon
4.	*Restructured*	No contrast in initial consonants	Loss of register through vowel system restructuring; complete merger of consonants	Cambodian

Huffman (1976) classifies the fifteen Mon-Khmer languages into four main groups. As summarised in Table 2 above, the first group is called 'conservative', which is a group of languages in which the voiced and voiceless stop distinction is maintained with little or no effect on the following vowel. The second group is called 'transitional', which covers a group of languages in which the tense and lax contrast in initial stops /p', t', c', k'/ versus /p, t, c, k/ is also maintained and the stops phonetically affect the following vowel. The third group is a group of 'pure register' languages with a merger of initial stops with some retention of sub-phonemic differentiation and a complete register in the vowels. The fourth group is called 'restructured' languages in which initial stops have been in a complete merger and the vowel register has also lost through changes in articulation of the vowels or diphthongisation in vowels.

It is clear that Huffman (1976) characterises register as a system where the voicing distinction in onsets has been neutralised, and instead the distinction is now realised on the following vowel. Thus, the term 'register' is used to refer to specifically languages that have transphonologised laryngeal contrasts onto the following vowels as bundles of correlated acoustic properties.

1.3.2.2 Is Madurese a register language?

The question of whether Madurese is a register language becomes relevant because Madurese vowels also show some features commonly associated with phonetic features observed in register languages. Indeed, Cohn (1993a) discusses a register account but raises three important issues, two of which are particularly worth mentioning here. The first issue is related to the fact that the vowel alternations in Madurese primarily differ in vowel height. However, they do not seem to differ in F0 and voice quality[1] in her speaker as commonly observed in register languages. The second issue is that Madurese and canonical register languages have very different phonological systems. This is because on the one hand all Mon-Khmer register languages have undergone a loss of voicing contrast and therefore they automatically do not have the CV co-occurrence restriction. On the other hand, Madurese maintains the voicing contrast and shows the CV co-occurrence restriction.

In a later study, however, Cohn and Lockwood (1994) interpret the high and non-high vowel sets in Madurese as a register difference which shows similar patterns to the Mon-Khmer register languages. Specifically, they suggest that the two vowel sets can be systematically distinguished by their F1 (vowel height) and F0 (pitch) values, which they claim that both F1 and F0 are lower following voiced and voiceless aspirated stops. However, if we look at Huffman' (1976) definition on register system summarised in Table 2, it is clear that Madurese cannot be considered as a register language. This is primarily due to the fact that the voicing contrast in Madurese is still preserved.

---

[1] However, it is important to note that Cohn did not do voice quality measurement herself. This claim is only based on her impression.

[2] See Husson (1997), who looks at the socio-political and economic aspects of the Madurese migration

In conclusion, we rule out the tense-lax distinction because there is a mismatch between the proposed phonological feature in the preceding consonants and the realisation of the feature in the following vowels. We also discount the proposal for Madurese as a register language due to the fact that the voicing contrast in canonical register languages is lost and thus it cannot explain the CV co-occurrence restriction.

1.3.3 Advanced tongue root (ATR)

In the following sections, we will consider two other possible phonological features that have also been proposed to account for the consonant-vowel interactions in Madurese. They are advanced tongue root ([ATR]) and lowered larynx ([LL]) (Trigo, 1991; Cohn, 1993b; Cohn & Lockwood, 1994). First, we will discuss what the predictions of these features would be under a transparent phonetics-phonology mapping. Specifically, we will discuss how and why advancing the tongue root and lowering the larynx could affect the acoustics. Second, we will discuss previous studies which propose ATR and LL as possible consonantal features responsible for the CV interactions in Madurese.

The feature [ATR] is a phonological feature which is commonly used to distinguish different types of vowels and has also been associated with vowel distinctions and vowel harmony in a number of African languages such as Akan, Maasai, Kinande, Yoruba, and Zulu. In these languages, it has been widely known that in addition to vowel quality there is another important articulatory dimension which seems to contrast a pair of vowels and the dimension is related to whether the vowel sets are produced with an advanced or retracted tongue root (e.g. Ladefoged & Maddieson, 2001; Lindau, 1979; Stewart, 1967; Trigo, 1991).

Studies on ATR vowel harmony are particularly relevant with the present study because the consonant-vowel co-occurrence restrictions in Madurese have also been associated with ATR vowel harmony in the sense of feature spreading from consonants to vowels (Trigo, 1991). With regard to ATR harmony in African languages, however, it is important to note that vowels act as both the harmony trigger and the harmony target. In those languages, vowel harmony can spread either rightward or leftward (Casali, 2008). In the case of Madurese, however, it appears that prevocalic consonants act as harmony trigger while vowels become the target of harmony, i.e. harmony is always rightward-spreading. The question is whether in fact there are some phonetic differences between ATR vowels and ATR consonants. This question will become relevant when we later consider that the feature ATR may also play a role in the consonant-vowel interactions in Madurese.

Some scholars (e.g. Casali, 2008) classify ATR languages based on the number of vowels in their phoneme inventories. What is also interesting about languages with an ATR system is the fact that some of them demonstrate an ATR harmony as well. For example, Akan and Maasai, which belong to nine-vowel ATR harmony languages, exhibit 'affix harmony and root-internal [ATR] agreement' (Casali, 2008). The feature [ATR], which in these languages belongs to vowels, can spread either rightward or leftward from the triggering vowels. The direction of the spreading depends on which morphemes (roots or affixes) are dominant as the possible harmony trigger in the languages. The spreading of the feature continues as long as it is not intervened by consonants that can act as harmony blockers. Casali (2008) also provides a detailed account of types of ATR vowel harmony.

1.3.3.1 Articulatory correlates of [ATR]

In languages where the feature [ATR] is considered to distinguish two sets of vowels, the difference primarily resides in the fact that [+ATR] vowels are articulated with a relatively higher tongue body position and also more fronted than their [-ATR] counterparts. In general, each member of the [+ATR] vowels impressionistically sound higher than each member of the [-ATR] vowels (Casali, 2008). Moreover, a number of following impressionistic studies (Pike, 1967; Stewart, 1967) and instrumental ones (Jacobson, 1978; Ladefoged, 1968; Lindau, 1979) provide compelling evidence that the [+ATR] vowels in Akan and a number of West and East African languages were also produced with the tongue root position more advanced than their [-ATR] counterparts.

As pointed out by Casali (2008), even though describing the harmonic feature [ATR] as entailing either advancement of the tongue root for [+ATR] vowels or retraction of the tongue root for [-ATR] vowels provides a good preliminary estimation, the description requires further qualifications. Firstly, Lindau (1979) observed that advancing or retracting the tongue root is not the only articulatory mechanisms for increasing (in the case of [+ATR] vowels) or decreasing (in the case of [-ATR] vowels) the overall size of the pharyngeal cavity. In fact, other articulatory mechanisms such as vertical movement of the larynx and other gestures also contribute to either an increase or a decrease in the overall volume of the pharyngeal cavity. On the basis of this, Lindau (1979) proposed an alternative feature, i.e. the feature 'Expanded' to replace the feature [ATR].

More recently, using a magnetic resonance imaging (MRI) technique, Tiede (1996) found that his Akan speaker produced the [+ATR] vowels by a combination of tongue root advancement, larynx lowering and tension maintenance in the pharyngeal walls. In contrast, the [-ATR] vowels were produced by retracting the tongue root, constricting the pharyngeal passage and raising the larynx. Although most phonologists have a general agreement that [ATR] involves an expansion of the pharyngeal cavity as a whole rather than merely an advancement of the tongue root, the term [ATR] has been retained instead of the alternative feature 'Expanded' as proposed by Lindau (1979).

Secondly, there are cases in which some languages which exhibit ATR harmony do not base their harmonic feature on tongue root advancement or other pharyngeal cavity expansion whatsoever. One example of this comes from Ateso, a language whose ATR harmony primarily involves changes in the height of the tongue body instead (Lindau & Ladefoged, 1986). There is also other evidence showing that different speakers of even a single language may use different mechanisms for implementing an [ATR] contrast. As reported by Lindau and Ladefoged (1986), speakers of the Nilotic language Dho-Luo implement ATR contrasts in the language by either adjusting the tongue height or moving the tongue root with or without vertical displacement of the larynx. Edmondson and Esling (2006) also demonstrate another interesting possibility. Using a laryngoscopic technique, they reveal that the production of [-ATR] vowels in two West African languages, Akan and Kabiyè, involves a constriction made by the epiglottis and aryepiglottic folds.

With regard to possible ATR contrast mechanisms, Casali (2008) notices that the number of languages for which we have access for direct articulatory observation is still relatively very few in comparison with the large number of ATR-harmony languages available. Therefore, it is possible that other articulatory mechanisms in the expression of ATR contrasts can be further revealed as more instrumental studies on languages of this type are conducted in the future.

Voice quality has also been associated with ATR distinctions in many ATR languages. That is, vowels which are produced with an advanced tongue root are usually attributed to a lax or breathy voice quality while those which are produced with a retracted tongue root are usually associated with a tense or creaky voice quality. For example, the [+ATR] vowels in some dialects of Akan have been reported to sound relatively breathier, fuller or deeper than their [-ATR] counterparts (Stewart, 1967) and the [+ATR] vowels in Maasai has also been described to have a ‘somewhat breathy voice quality’ (Tucker & Mpaayei, 1955). Pike (1967, p. 130) describes that vowels produced with an expanded pharyngeal cavity sound ‘fuller’ or ‘deeper’ while those produced with a constricted pharyngeal cavity sound ‘choked up’. Stewart (1967, p. 199) even speculates that breathy voice is the most important auditory correlate of tongue root advancement. This is because tongue root advancement would result in an expanded pharynx that may account for the breathy voice. This is also the case for Shilluk, a Western Nilotic language, in which [+ATR] vowels sound impressionistically breathier compared to their [-ATR] counterparts (Remijsen, Ayoker, & Mills, 2011).

However, there seem to be no clear explanations why advancing the tongue root would result in breathy voice quality. To the best of my knowledge, a number of studies which associate [+ATR] vowels produced with a rather breathy voice quality barely account explicitly for how and why this mechanism occurs articulatorily. For example, Stewart (1967, p. 199) suggests that breathy voice may result from a wide pharynx due to tongue root advancement, but he does not explain explicitly how a wide pharynx might affect particularly the vocal fold settings which may lead to breathy voice. However, Kingston *et al.* (1997, p. 1697) provides a rather explicit mechanism of how voice quality and tongue root position may be interrelated. They suggest that voice quality can be physiologically dependent on tongue root position ‘if the aryepiglottic ligament and membrane, which connect the tongue root to the arytenoid cartilages via the epiglottis, cause the arytenoids to slide forward slightly and/or rock slightly apart, slackening or separating the vocal folds enough to lax the voice, when the tongue’s root is advanced or its body raised’. The vocal folds may not completely close and as the glottis is partially open, it could generate turbulence noise, which leads to breathy voice quality.

The feature ATR has also been associated with the tense-lax distinction in some Germanic languages such as English and German. In this case, [+ATR] vowels have been considered as similar to tense vowels while [-ATR] vowels have been associated with lax vowels. However, different from what is observed in many African languages, there seems to be no solid evidence that tongue height and tongue root advancement in Germanic languages are two independently controlled gestures. That is, in Germanic languages such as English the tense-lax, vowel pairs can be primarily distinguished with reference to only two variables (see Lindau 1978 and Ladefoged & Maddieson 2001).

Ladefoged and Maddieson (2001) suggest that the tongue root advancement in African languages such as Akan and Igbo constitutes an independent tongue gesture but in Germanic languages such as English and German it appears to be rather an epiphenomenon of vowel height. In general, [+ATR] vowels in an ATR system appear to be raised and advanced. Front vowels in languages which distinguish [+ATR] and [-ATR] vowels have formant frequency properties similar to tense and lax vowel pairs (Ladefoged & Maddieson, 2001). That is, both [-ATR] and lax front vowels are lowered and centralised in the vowel space. However, this does not seem to be the case for back vowel pairs. That is, lax back vowels are normally more centralised while [-ATR] back vowels do not always exhibit this characteristic. In fact, in languages such as Akan, Ateso, Igbo and Ijo, [-ATR] vowels are always further back whereas [+ATR] vowels always appear to be further forward (Ladefoged & Maddieson, 2001).

1.3.3.2 Acoustic correlates of ATR

It has been established that both retracting and advancing the tongue root will have certain acoustic effects on vowel height. Specifically, the manipulation of the overall size of the pharyngeal cavity in ATR contrasts is predicted to produce a number of acoustic consequences. First, the larger size of the pharyngeal cavity will result in a lower frequency of the first formant (F1) (Halle & Stevens, 1969). This is because expanding the pharynx essentially enlarges the cross-section of the back cavity, and this will lower F1 since the first resonant frequencies of the front and back cavities are predicted to be close together, creating coupling effects between the two tubes (Stevens, 1989). F1 lowering constitutes the most reliable acoustic correlate of the ATR contrast in a variety of Nilotic languages (Jacobson, 1978; Lindau, 1978), Degema (Fulop, Kari, & Ladefoged, 1998), Maa (Guion, Post, & Payne, 2004), and Shilluk (Remijsen *et al.*, 2011).

In a number of cases, there is also evidence that the frequency of the second formant (F2) varies systematically across ATR vowel sets. However, this does not seem to be consistent across languages and also it is not clear why pharyngeal expansion should also change F2. Some languages show that their [-ATR] front vowels have greater F2 values (more front) and their [-ATR] back vowels have smaller F2 values (more back) than the corresponding [+ATR] vowels. This shows that [-ATR] vowels are more peripheral than their [+ATR] counterparts. For example, Jacobson (1980) observed that some vowel sets in the Nilotic languages Dho-Luo and Shilluk demonstrate such F2 effects, but he did not find the same effects in all vowel pairs for Dinka, another Nilotic language.

The finding is, however, contrary to what Fulop *et al.* (1998) reported for the Niger-Congo language Degema. They found that the F2 values of some [+ATR] vowels in Degema were consistently more peripheral, i.e. higher F2 values for front vowels and lower F2 values for back vowels, compared to their [-ATR] counterparts. Unlike both studies, Guion *et al.* (2004, p. 536) reported that there was no significant difference in F2 values between the [+ATR] and [-ATR] vowels in Maa, a Nilo-Saharan language. Hence, it appears that F2 values may not be a reliable correlate for the ATR contrast given that even languages from the same family appear to vary with regard to this parameter. It is also possible that F2 can be important correlates for ATR distinction in one language, but not in another, suggesting a language-specific nature of the phonetic realisation of an ATR system.

Vowels with [+ATR] and [-ATR] features can also be differentiated by their spectral slopes, another acoustic correlate of the [ATR] contrast. In this case, [+ATR] vowels often sound as ‘deeper’, ‘hollow’, or ‘breathy’ and laxer, while [-ATR] vowels are frequently described as ‘brighter’, ‘brassy’ or ‘creaky’. Auditory impressions such as these have been associated with the overall slope of the spectrum. The impressions result from the fact that [+ATR] vowels tend to have energy concentration in the lower frequency region while [-ATR] vowels tend to have energy concentration in the higher frequency region. For example, Denning observed that breathy vowels which are associated with [+ATR] in Dinka have lower F1 than [-ATR] vowels which are associated with [-ATR]. They can also be consistently distinguished by two measures of spectral tilt H1-H2 and H1-A1 whereby breathy vowels have higher H1-H2 and H1-A1 values than non-breathy vowels. In addition, there is also a general tendency that high vowels which are often produced with a large pharynx sound breathy whereas non-high vowels produced with a small pharynx sound tense (see e.g. Gregerson 1976, Laver 1980 and Maddieson & Ladefoged 1985).

The question is what mechanisms may contribute to the spectral correlates of the ATR contrast. Related to this, Guion *et al.* (2004) summarise three possible origins of the spectral slope differences. First, they can be as a consequence of differences in voice quality. Compared with modal phonation, the source spectrum for breathy phonation shows less harmonic energy in the upper frequency range. This is because the vocal folds never close completely and as a result the vocal fold vibration during breathy voicing is more sinusoidal. This sinusoidal nature of the vocal fold vibration results in less energy at high frequencies and more energy at low frequencies around the first and second harmonics (Stevens, 1977). Spectra of vowels produced with breathy voicing may have a reduction in amplitude of about 15 dB in the higher frequencies than those produced with modal voicing while the first harmonic for these two phonation types is similar (Stevens, 1998, p. 89). This may explain why the spectral slope for breathy vowels is much steeper than that for modal vowels.

Second, differences in the spectral slope may result from different tensions in the pharyngeal walls. The pharyngeal walls with a tenser or stiffer configuration are more likely to produce less loss of acoustic energy than those with a comparatively laxer configuration. Third, it is also possible that the spectral slope differences are due to the effect of the pharyngeal constriction on the first formant amplitude whereby pharyngeal constriction itself can influence formant damping. [-ATR] vowels that are produced with a firmly constricted pharynx can damp friction in the F1 frequency region due to the air viscosity (Fulop *et al.*, 1998, p. 84).

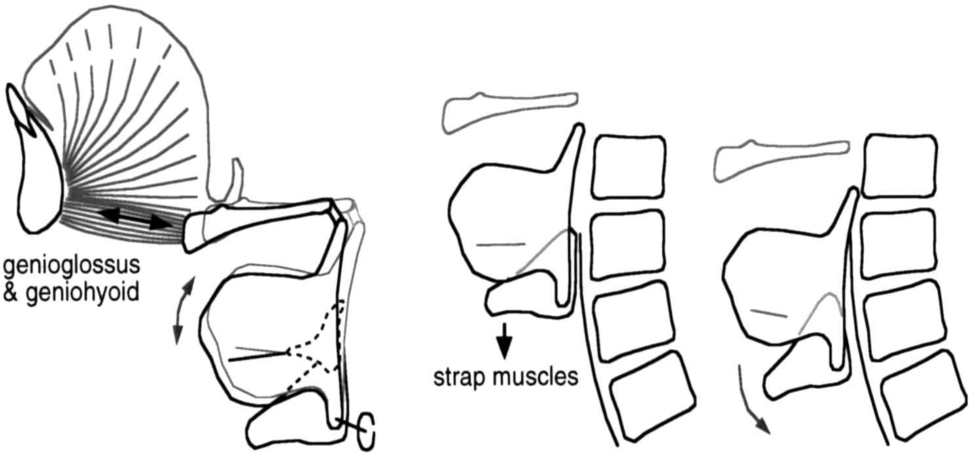
Another important correlate of ATR contrast is spectral emphasis (Traunmüller & Eriksson, 2000). This spectral measure compares energy distribution in the spectrum between the fundamental frequency and the rest of the harmonics. The measure relates the amount of energy in the high frequency region, which is defined as energy upward from 1.5 times the fundamental frequency, to the overall energy. For example, [-ATR] vowels in Shilluk were found to have significantly higher values for spectral emphasis than their [+ATR] counterparts (Remijsen *et al.*, 2011). This indicates that the energy is concentrated in the rest of the harmonics for [-ATR] vowels while it is concentrated in the fundamental for [+ATR] vowels.

1.3.4 Lowered larynx (LL)

The feature [LL] has also been proposed as a phonological feature associated with the consonant-vowel interactions in Madurese (Cohn, 1993b; Trigo, 1991). This feature has also been suggested to be found active in Buchan Scots and proposed as a phonological feature which accounts for the consonant-vowel interaction in that language as well (Paster, 2004; Youssef, 2010). Using a quite different term, Avery and Idsardi (2001) also categorise larynx height as a laryngeal feature. They propose that the laryngeal dimension has two values, namely [raised], which corresponds to a raised larynx and [lowered], which corresponds to a lowered larynx.

1.3.4.1 Acoustic correlates of LL

In general, lowering the larynx will make the vocal tract volume above the glottis longer and the lengthening of the vocal tract will result in lower formant frequencies. It is for this reason that vowels following consonants produced with a lowered larynx tend to sound more closed in comparison with consonants articulated with a raised larynx. In this regard, the effect of larynx lowering is particularly obvious for the first formant frequency (F1) as it primarily depends on the cavity size between the glottis and the place of maximum constriction between the tongue and the palate.



(a) horizontal component (b) vertical component

Figure 1–1. Horizontal and vertical components of extralaryngeal F0 control mechanism (Honda, Hirai, Masaki, & Shimada, 1999)

Larynx lowering is also known to have a lowering effect on the fundamental frequency (F0). Evidence of this comes from a magnetic resonance images (MRI) study by Honda *et al*. (1999) who investigated the F0 control mechanism and how it was related to the vertical laryngeal displacement. They observed that the larynx height remains high and constant in the high F0 range in order that the horizontal movement of the hyoid bone facilitates the rotation of the thyroid cartilage, subsequently leading to the stretching of the vocal folds and raising F0. In the low F0 range, they found that the jaw, hyoid bone and the larynx move downward in order that the cricoid cartilage rotates along the cervical spine, leading to vocal fold shortening and relaxation and lowering F0.

Figure 1–1(a) above shows that there is a consistent horizontal movement of the hyoid bone in the high F0 range. The horizontal movement is made possible by the suprahyoid muscles such as genioglossus and geniohyoid muscles (Honda, 1983), which facilitate the thyroid cartilage rotation resulting in stretching the vocal folds and subsequently raising F0. There is only a minimum vertical movement in this high F0 range. This is different from what is shown in Figure 1–1(b), which shows the low F0 range, in which there is a big vertical movement involving the hyoid bone and larynx along the cervical spine. This vertical movement is produced by the action of the infrahyoid muscles, which rotate the cricoid cartilage along the cervical curvature. As a consequence, this action of larynx lowering gives rise to the shortening and relaxation of the vocal folds.

In short, the vocal folds will become shorter and relaxed when the larynx lowers and as a result the shorter vocal folds will vibrate more slowly and consequently produce lower F0 (see Honda, Hirai, Masaki, and Shimada, 1999 for more discussion on this mechanism and Brunelle 2010). Put differently, the reason why lowering the larynx may lower F0 resides ‘in the relaxing effect on the pitch mechanism of the larynx by the mechanical downwards pull of the infrahyoids’ (Laver, 1980, p. 30).Larynx lowering has also been suggested to yield breathy voice quality (Laver, 1980, p. 31) and longer VOT following lax stops (Brunelle, 2010). Brunelle (2010) argues that although it is probably difficult to account for such effects as a consequence of larynx position, two mechanisms can explain why this may happen. This is due to the fact that larynx lowering not only triggers higher subglottal pressure but also exert direct effects on the configuration of the vocal folds as a whole.

The vocal folds tend to be quite close together when the larynx is in its default position. However, when the larynx is lowered, the degree of the contact between them may decrease, which subsequently could facilitate a leakage of the air from the subglottal space through the glottis. Given that the glottis is open in the course of the production of voiceless stops (considering that lax stops are voiceless), the high airflow that runs through the glottis does not influence lax stops themselves. Instead, it delays the voicing onset and carries on onto the beginning of the following vowel, resulting in breathiness in the vowel (Brunelle, 2010). In short, as lowering of the larynx slackens the vocal folds, it results in an increase in glottal opening and this glottal aperture may consequently contribute to breathiness.

1.3.5 ATR and LL: are they different features?

Both tongue root advancement and larynx lowering have also been associated with consonant voicing maintenance mechanisms. Specifically, tongue root advancement and larynx lowering are two articulatory mechanisms that can be used to help maintain voicing during the production of voiced stop consonants (Westbury, 1983). In relation to this, Perkell (1969) also provides evidence that voiced stops are produced with larger pharyngeal width while voiceless stops are produced with smaller pharyngeal width. He argues that this pharyngeal expansion which corresponds to tongue root advancement is also used as a mechanism to sustain voicing during the production of voiced stops.

Similarly, Westbury (1983) found that the tongue root was more advanced during the production of voiced stops than during the production of voiceless stops in the majority of his American English data. He also observes that consonantal voicing appears to determine the larynx position in which voiced stops tend to be produced with a lowered larynx position. Thus, all else being equal, the larynx lowering during the closure of a stop would generally facilitate voicing and that is why we would expect that voiced stops are produced with a downward movement of the larynx.

The articulatory mechanisms of tongue root advancement and larynx lowering are also known to produce similar acoustic consequences, one of which is that they both robustly lower the frequency of F1. The lowering of F1 arises from the fact that both advancing the tongue root and lowering the larynx will result in a wider and longer vocal tract and a wider and longer vocal tract will make formant frequencies go down even further. Furthermore, Lindau (1978, p. 552) suggests that the tongue root and the larynx cooperate to achieve pharyngeal expansion and because of this, she labels the feature as ‘Expanded’. Therefore, it does not come as a surprise that due to the close association between these two articulatory gestures, distinguishing their acoustic consequences is also difficult, if not impossible. In addition, both advancing the tongue root and lowering the larynx can result in lower fundamental frequency (F0). They have been associated with demonstrating similar consequences in voice quality as well. That is, vowels which are articulated with either an advanced tongue root or a lowered larynx are generally expected to sound breathy.

However, it is important to note that we do not suggest that [ATR] and [LL] are the same phonological features for two reasons. First, they are produced by different articulatory gestures and second, they are suggested to be found independently active in different languages. What we want to show here is the fact that to some extent these two gestures appear to exhibit similar acoustic manifestations. In relation to phonological features associated with these articulatory gestures, as discussed earlier, the feature [ATR] is particularly used in the description of two vowel sets observed in many African languages. On the contrary, the feature [LL] has been associated with two vowel sets found in a number of Southeast Asian languages although it is not explicitly suggested as a phonological feature, except with respect to the CV co-occurrence restriction in Madurese and voice register in Javanese (for Javanese, see e.g. Fagan, 1988; Brunelle, 2010).

Furthermore, there is one important distinction between the phenomenon involving vowel sets in these two language areas. In African languages with ATR harmony it is the vowels that act as the harmony trigger and the harmony target and most studies concerning this phenomenon particularly deal with the vowels per se. Although a number of studies also mention that [ATR] vowels are produced with a lowered larynx, the feature [LL] does not appear to be considered dominant in those languages. In contrast, in languages such as Javanese and Madurese, there has been an association between consonant voicing and its observed effects on vowels. This has been suggested to occur as a result of a feature spreading from consonants to following vowels. Unlike ATR harmony languages in African languages, the features held responsible for this type of assimilation or harmony have been suggested for Madurese as ATR and LL (Trigo, 1991) or only LL (Cohn, 1993b).

The feature [LL] has also been proposed as a phonological feature that may account for the tense and lax stop distinction in Javanese. In his analysis on the tense and lax distinction of Javanese stops, Brunelle (2010) argues that there is some evidence that larynx height plays a role in their distinction. He observes that the larynx is consistently lower during the production of lax stops and higher during the production of tense stops. He mentions a number of acoustic characteristics of stops produced with a lowered larynx; these include longer VOT, lower F0 in the following vowel, lower F1 in the following vowel, and breathy vowels. In contrast, stops produced with a raised larynx tend to have shorter VOT, higher F0 in the following vowel, higher F1 in the following vowel, and modal vowels (Huffman, 1976; Kirby & Brunelle, in press; Thurgood, 2007).

1.3.6 ATR and LL as possible consonantal features in Madurese

Trigo (1991) proposes two possible features that may account for the consonant-vowel interaction in Madurese, i.e. why certain consonants are only followed by certain vowels. The features are [Lowered Larynx]/[Raised Larynx] ([LL/RL]) and [Advanced Tongue Root]/[Retracted Tongue Root] ([ATR/RTR]). Trigo (1991) claims that both voiced stops and voiceless aspirated stops in Madurese share the feature [LL] that spreads to the following vowel. She also claims that the voiceless aspirated stops of Madurese have some similarities to the Javanese lax stops, which are produced with a lowered larynx, lower pitch and breathy phonation (see Brunelle, 2010; Fagan, 1988; Hayward, 1995 on Javanese laryngeal contrast). The claim was based on her personal communication with Kenneth Stevens, who suggests that the voiceless aspirated stops of Madurese also have a lowering effect on the fundamental frequency of a following vowel, similar to the lax stops of Javanese. However, she does not provide any phonetic evidence in support of her claim on Madurese.

Trigo (1991) also argues that both voiced and voiceless aspirated stops may share the feature [ATR]. Although the feature [ATR] is relevant for voiced stops since tongue root advancement is a common strategy used for maintaining voicing during stop closure (Chomsky & Halle, 1968; Perkell, 1969; Ohala & Riordan, 1979; Westbury, 1983), as she also admits, it may not be compatible with voiceless aspirated stops. This is because according to Perkell (1969) voiceless aspirated stops are not produced with either a pharyngeal expansion or tongue root advancement. However, following Stevens (1966), who proposes historical evidence that voiceless aspirated stops in Madurese may have derived from earlier voiced stops, Trigo (1991) maintains that voiceless aspirated stops can be phonologically [ATR] as well. She also claims that they also have a lowering effect on the fundamental frequency of the following vowel, but she did not provide any articulatory mechanism that may explain why ATR should lower F0, let alone acoustic evidence. This idea, however, probably comes from the assumption that Madurese voiceless aspirated stops have been claimed to bear similarity to the lax stops of Javanese (Catford, 1977, p. 106).

In order to confirm whether the ATR account is a possible explanation for the consonant vowel interaction, Cohn (1993a) conducted a phonetic investigation into the realisation of the vowel height alternations in Madurese. Following an ATR account, she predicted lower F1, higher F2 and longer duration for [+ATR] vowels and higher F1, lower F2 and shorter duration for [-ATR] vowels. She looked into F1, F2 and duration from eight tokens of each vowel ([ɛ] ~ [i], [ɔ] ~ [u], and [a] ~ [ɤ]). She also considered the three stop categories (i.e. voiced, voiceless unaspirated and voiceless aspirated) and measured the duration and VOT of the stops. The tokens were produced by one male speaker from Western Madura.

She found that the high and non-high vowels are systematically distinguished by F1 while the differences in F2 are not systematic. Except for [a] ~ [ɤ], which show a marked difference in duration, the other vowel pairs have very small differences in duration. She also found that the three stop categories have the same durations but on average the closure duration of the voiceless aspirated stops appears to be slightly longer. In terms of VOT, they are different whereby voiced stops show very little voicing lag while the voiceless aspirated stops indicate a slightly longer VOT than the voiceless unaspirated stops. Based on these acoustic findings, Cohn (1993a) rejected the ATR explanation because only the F1 alternations follow the expected direction of an ATR system while the F2 alternations and vowel duration do not.

In another study, Cohn (1993b) also provides a phonological analysis in an attempt to account for what possible phonological feature which may be shared by voiced and voiceless aspirated stops of Madurese in triggering the vowel height alternations. She argues that the spreading of the feature [LL] from consonants to vowels accounts for the consonant-vowel interactions and vowel harmony in Madurese. She refers to the rule in which the preceding consonant conditions the height of the following vowel as ‘Vowel Raising’. Cohn (1993b) asserts that the vowel-raising rule has long-distance and categorical effects.

In this case, it is important to note that unlike Trigo (1991), who proposes that the features [ATR] and [LL] are responsible for the vowel height alternations and considers them as pharyngeal and privative features, Cohn (1993b) argues that the only feature responsible for the vowel alternations is the feature [LL] and considers it as a binary laryngeal feature instead. It is a binary feature since the consonants associated with the feature can either lower or raise the following vowels, depending on whether the feature is [-LL] or [+LL] respectively. As discussed earlier, she excludes the likelihood of the feature [ATR] involvement here because doing so would require to also consider the voiceless aspirated stops voiced or breathy voiced. More importantly, such a ATR-based contrast is not supported by the phonetic findings in her other study (see Cohn 1993a). However, recall that she has not provided any acoustic data on voice quality measures considering the fact that, as discussed earlier, either advancing the tongue root or lowering the larynx has also been associated with breathy voice quality.

In an attempt to further unravel what phonetic properties voiced and voiceless aspirated stops may share, Cohn and Lockwood (1994) conducted another acoustic study looking at voicing during closure, stop duration, aspiration, formant structure (F1 and F2), vowel duration and fundamental frequency. This study involved two speakers (one male, one female) of Eastern Madurese. The results confirm that aspirated stops do not show any phonetic voicing indicated by the fact that there is no vocal fold vibration during occlusion. Therefore, the question which remains unanswered is what phonetic properties, if any, the voiced and voiceless aspirated stops share synchronically by which they pattern together in triggering vowel raising (Cohn & Lockwood, 1994). It is important to bear in mind that they assume that non-high vowels become high (raised) following voiced and voiceless aspirated stops and the feature responsible for this has been suggested to be a consonantal feature associated with the preceding consonants.

In addition, they also found that there were small differences in closure duration and aspiration between voiceless unaspirated and voiceless aspirated stops. That is, on average the duration for voiceless aspirated stops was approximately 10 ms longer than for voiceless unaspirated stops while the VOT for voiceless aspirated stops was

14 ms longer than for voiceless unaspirated stops (Cohn & Lockwood, 1994, p. 76). They suggest that these two voiceless stops appear relatively similar synchronically and can mostly be distinguished by the following vowel quality.

In the case of vowels, they found that there were systematic differences in F1 and F0 between the high and non-high vowels. Specifically, F1 values following voiced and voiceless aspirated stops are systematically lower than those following voiceless unaspirated stops and nasals. With regard to F0, they found that F0 values following voiced and voiceless aspirated stops were lower than those following voiceless unaspirated stops and nasals. However, this is not the pattern we would expect to see for F0 of high vs. non-high vowels, as it is high vowels which show lower F0 rather than vice versa in Cohn and Lockwood's (1994) study.

Thus, previous studies discussed above (i.e. Cohn 1993a and Cohn & Lockwood 1994) provide some acoustic measures mainly on voice onset time, closure duration, vowel quality and fundamental frequency. However, they did not consider voice quality, which can be another important acoustic correlate of tongue root advancement and larynx lowering (Brunelle, 2010; Denning, 1989; Fulop *et al.*, 1998; Guion *et al.*, 2004; Laver, 1980; Remijsen *et al.*, 2011). Moreover, the data in previous studies were collected from at most two speakers of Madurese. In this study, we present new data on the phonetic realisation of Madurese stops from a larger sample size of 15 native speakers. In addition to VOT, closure duration, F0 and vowel quality, we also examined several acoustic correlates of voice quality which have been mentioned but not examined in previous studies of Madurese (Cohn, 1993a; Cohn & Lockwood, 1994). The results will particularly help us assess the hypotheses of Trigo and Cohn that Madurese voiced and voiceless aspirated stops may share a phonetically transparent phonological feature such as [ATR] or [LL].

1.4 Research Questions

1. What acoustic property or properties, if any, do voiced and voiceless aspirated stops (and the vowels which follow them) share in comparison with voiceless unaspirated stops?

2. Are the acoustic properties of these consonants consistent with what we would expect if they share an articulatory feature, specifically [ATR] or [LL]? That is, do we find acoustic evidence for a phonetically grounded phonological feature that could explain the patterning together of voiced and voiceless aspirated stops to the exclusion of voiceless unaspirated stops in the CV co-occurrence restriction? If we do, which feature do the results support? If we do not, how can we account for the co-occurrence pattern?
3. What are the implications of the results of the study for our understanding of the phonetics-phonology mapping? Are the findings consistent with a concrete, transparent phonetics-phonology mapping, or do they suggest a more flexible, abstract phonetics-phonology relationship instead?

1.5 Structure of the Dissertation

The rest of the dissertation is organised in the following way. In Chapter 2, we provide some background about Madurese. We introduce the inventory of consonants and vowels, phonological and morphophonemic processes and also discuss some non-phonological aspects. We also provide some background information about the Madurese people and linguistic situation in Madura, orthography and speech levels in Madurese, i.e. *kasar* ‘coarse’, *biasa* ‘ordinary’, *tengnga* ‘middle’ and *alos* ‘refined’. In Chapter 3, we address and discuss a number of theoretical frameworks which function as the foundation for understanding voicing and laryngeal contrasts. We will describe how voicing and laryngeal contrasts are manifested through, for example, voice onset time, fundamental frequency and formant frequencies, and how they are also related to voice quality as well as what acoustic measures are commonly used to examine voice quality. This description provides a foundation for later analyses. We will also review and discuss a number of studies which report relevant empirical findings on these issues.

In Chapter 4, we present the methodology used in the present study. The methodology section provides information regarding the study’s participants, the process of data collection, and the process of data segmentation, measurement and acoustic analyses. We also introduce the statistical analyses used in the study. In

Chapter 5, we present the results of statistical analyses of the acoustic measurements and the implications of the findings with respect to research questions. In Chapter 6, we address and discuss the results of the study presented in Chapter 5 by contextualising them with literature and earlier findings. In Chapter 7, we conclude the study and show how it has bearings on wider issues in phonetics and phonology. We also provide suggestions for future studies.

2 Language Background

2.1 Social and Language Situations

Madura is a small island located north of Java, Indonesia (see Figure 2–1 below showing the map of Indonesia and the position of Madura in the archipelago, as indicated by a red circle). The island is a main producer of salt, which is why it is also widely known as the island of salt. The geographical condition in particularly the western part of the island is not as fertile as other islands in Indonesia. The island itself constitutes part of East Java province. In addition to the main island, there are a number of other small islands around its eastern part. Madura is administratively divided into four regencies, namely, Sumenep, Pamekasan, Sampang and Bangkalan (ordered from east to west). Pamekasan is the administrative capital city of Madura.

Figure 2–1. Map of the Indonesian archipelago. Madura is circled (accessed from http://www.lib.utexas.edu/maps).

The condition of most land in the eastern regencies of Sumenep and Pamekasan is more fertile than the condition of land in the western regencies of Sampang and Bangkalan. Such a natural condition may partly have become a push factor for some Madurese people living in these areas to emigrate to other Indonesian islands that they consider would provide better livelihoods for them. In fact, the migration already took place during the Dutch colonial period, which was in part triggered by the need for labourers to work in the Dutch plantations in East Java in particular (Husson, 1997). The same thing also occurred during the relatively short Japanese occupation in Indonesia (1942-1945), during which many Madurese people were forced to work as labourers in Java. Moreover, the Indonesian government under the Soeharto administration from the late 1960s to the late 1990s also organised a planned migration and spreading of Madurese people across Indonesia[2]. With all this in mind, it is therefore not surprising that at present Madurese people can also be found living in different parts of Java, Kalimantan and other islands across the Indonesian archipelago. Figure 2–2 below shows the map of Madura, which is separated from Java by a small strait known as the strait of Madura.

Figure 2–2. Map of the Island of Madura (accessed from http://peta-kota.blogspot.co.uk/2011/07/peta-pulau-madura.html)

In terms of employment, the majority of Madurese people work either as farmers especially in the areas where water availability does not rely on rainfall, or as fishermen in the areas that are close to the sea. Some others also work in formal sectors and informal sectors other than farming and fishing. However, it is also

---

[2] See Husson (1997), who looks at the socio-political and economic aspects of the Madurese migration dating back from the 14th century.

common to find Madurese people who take different jobs simultaneously in order to better support their lives. For example, some farmers may also work as merchants or as fishermen during certain periods of the year.

Madurese is the main language spoken on the island of Madura. The number of Madurese speakers who speak the language for daily communication at home is approximately 7.8 million (Ananta *et al.,* 2015, p. 278) and this number includes the speakers living on the island of Madura itself and other islands across the Indonesian archipelago[3]. Madurese is formally taught at school from grade 1 to grade 12. Although Indonesian or Bahasa Indonesia is generally used as the language of instruction at school, Madurese is also used instead of Indonesian in some parts of Madura particularly for the first four grades of elementary school. The goal of using Madurese in tandem with Indonesian at grades 1-4 is to facilitate the learning of Indonesian because children mostly speak Madurese at home particularly those who live in villages and other remote areas.

Due to its important role as the uniting national language in the country where different ethnic groups live and hundreds of local languages are actively spoken, Indonesian is obligatorily taught up to university level. That is why it is not surprising if the majority of Madurese people also understand Indonesian today and in fact, Ananta *et al.* (2015, p. 290) mentions around 114,482 Madurese people use Indonesian for daily communication at home. However, there is no information about the definitive number of monolingual speakers of Madurese, but we believe they can still be found in remote villages and especially among older Madurese speakers who may have had no access to formal education due to the lack of facility or poverty.

---

[3] Madurese is also spoken in a number of small adjacent islands such as Bawean, Sapudi, and Kangean and some regencies spread along the northern coast of the eastern part of East Java province such as Pasuruan, Probolinggo, Kraksaan, Besuki, Situbondo, Bondowoso, and Jember. Some Madurese speaking people in those regencies were former Madurese migrants but they still maintain close contact with their relatives who live in Madura by making regular visits. In fact, the tradition of visiting relatives among Madurese people, which is usually made during the annual celebrations of important Islamic festivals, has been maintained from generation to generation.

There has also been a growing interest for Madurese younger generation in learning foreign languages such as English and Arabic. In this regard, English has been introduced as a subject at elementary school and indeed it is formally taught from secondary school to university level. Another foreign language which is also commonly taught in Madura is Arabic. This language has become an obligatory subject in the majority of religious schools called *madrasa* and particularly at Islamic boarding schools, known as *pesantren,* across Madura.

In addition, it is quite common to find Madurese people who speak other local languages as well. This is usually made possible when they migrate to other Indonesian islands where different local languages are spoken. For example, some Madurese people in Kalimantan may not only speak Madurese and Indonesian but also speak Banjar Malay. Similarly, depending on which part of Java they live in, Madurese people may also speak Javanese or Sundanese (see Ananta *et al.* 2015 for a review of languages spoken by different ethnic groups in Indonesia).

As a language, Madurese also have dialects in the sense of regional variations. However, there have been limited studies which describe Madurese dialects. Stevens (1968) only mentions in passing that Madurese can be divided into three major dialects. They are West Madurese, which covers Bawean and Bangkalan, Central Madurese, which includes Pamekasan and Sampang, and East Madurese, which comprises Sumenep and Sapudi. In relation to this, two studies which particularly deal with describing dialects in Madurese are worth mentioning: *Pemetaan Bahasa Madura di Pulau Madura* ‘The Mapping of Madurese on the Island of Madura’ (Soegianto *et al*., 1986) and *Geografi Dialek Bahasa Madura* ‘A Geography of Madurese Dialects’ (Soetoko *et al*., 1998). Unlike Stevens’ (1968) proposal, these studies focus on describing Madurese dialects on the main island. In this dissertation, however, we did not look at specific dialects although, as we will see in Chapter 4 later, the participants came from different dialect areas. More important is the fact that all the dialects have the same CV co-occurrence restriction.

Specifically, Soegianto *et al.* (1986) describe Madurese dialects according to the distribution of vocabulary across Madura. Based on this parameter, they classify Madurese into three major dialects[4]. They are Eastern Madurese, which is primarily spoken in Sumenep, Central Madurese, which is mainly spoken in Pamekasan, and Western dialect, which is primarily spoken in Bangkalan. They argue that the Madurese dialect spoken in Sampang can be categorised as a mixture of Western and Central dialects and could not be considered as a different dialect. That is to say, the people of Sampang who live in close boarders with Pamekasan will have the tendency to speak Central Madurese while those who live close to Bangkalan will tend to speak Western Madurese. It is important to note, however, that none of these studies provide any instrumental phonetic data to substantiate their dialect analyses. In this case, they rely more on vocabulary mappings and their impressions of how the words would be pronounced in different dialect areas.

Eastern Madurese spoken in Sumenep is considered as the standard dialect. This may be associated with the fact that Sumenep used to become the centre of some former Madurese kingdoms particularly in its connection with a number of former Javanese kingdoms such as Singosari, Majapahit and Mataram. Another reason may be related to the fact that Sumenep is located at the easternmost part of Madura. In this way, Eastern Madurese spoken in that area is considered relatively free from influences of other local languages compared with, for example, Western Madurese spoken in Bangkalan, which is close to Java where Javanese is mainly spoken.

Native speakers are aware of certain differences among the Madurese dialects, but such differences do not hinder successful communication between people who come from different dialect areas. This should also be the case for Madurese dialects that are spoken outside Madura. It is true that the same words may have different meanings depending on which Madurese dialects they are used in. In this regard,

---

[4] Madurese spoken outside Madura may form different dialects as well and the dialects may partly depend on which part of Madura the speakers originally come from. They may also be influenced by other local languages spoken in their respective areas.

Soegianto *et al.* (1986) suggest that some words are more common to be used in certain dialects and they are sometimes pronounced quite differently as well. For example, they observe that people from Pamekasan who speak Central Madurese tend to pronounce words such as *barampa* ‘how many’ and *jareya* ‘that’ as [bɤrɤmpa] and [ɟɤrijɤ] respectively without vowel reduction. In contrast, people from Sumenep who speak Eastern Madurese have the tendency to lengthen word-final vowels such as *baramma* ‘how’ and *paneka* ‘this’ as [bɤrɤmma:] and [panɛka:] respectively. In fact, it can also be observed that Sumenep people speak with a different intonation from Bangkalan people who speak Western Madurese. To my knowledge, no phonological and instrumental phonetic studies that particularly examine these prosodic aspects of Madurese dialects have been done to date.

2.2 Genetic Affiliation of Madurese

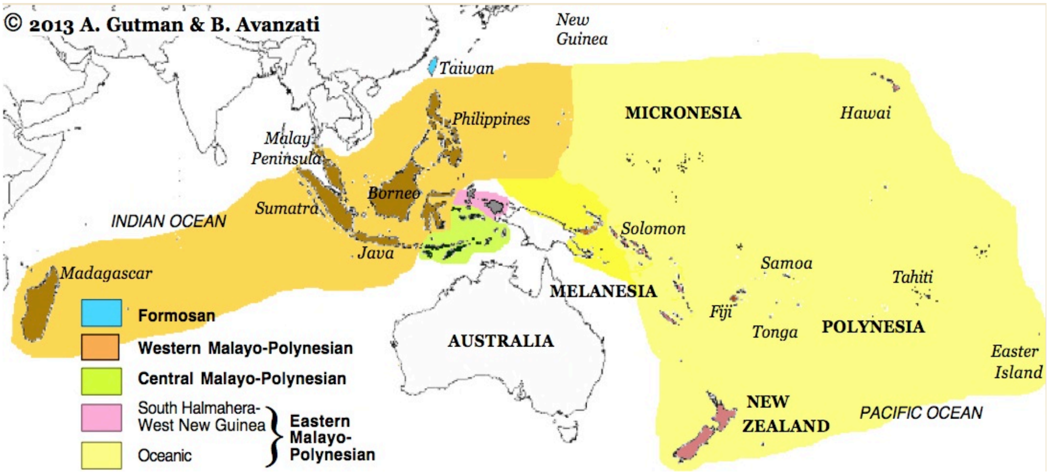


Figure 2–3. Map of the Austronesian language family (accessed from http://www.languagesgulper.com/eng/Austronesian.html)

Lynch *et al.* (2002) divide the Austronesian language family into two main groups, namely the Formosan languages of Taiwan and the Malayo-Polynesian languages. Malayo-Polynesian languages, which constitute the majority of the family, are subdivided into Western Malayo-Polynesian and Central/Eastern Malayo-Polynesian languages. Eastern Malayo-Polynesian languages are further divided into South Halmahera/West New Guinea and Oceanic languages. According to this classification, Madurese belongs to other Western Malayo-Polynesian group together with other languages of Sumatra, Borneo, Sulawesi, Java, Bali, Lombok, West Sumbawa, the Philippines and Madagascar.

In relation to this, it is worth noting that Adelaar (2005a) suggests a difference classification particularly with regard to the so-called Malayo-Javanic subgroup, which has been suggested to include for example Javanese, Malay, Madurese and Sundanese under the same subgroup (Dyen, 1965). Based on phonological and lexical evidence, Adelaar proposes that the Malayo-Javanic subgroup should be replaced by a ‘Malayo-Sumbawan’ subgroup. This subgroup puts Malayic, Chamic, and the Balinese-Sasak-Sumbawa group into one branch while Madurese and Sundanese in two other branches. He excludes Javanese from the subgroup. Thus, by this classification Madurese is not considered closely related to Javanese.

2.3 The sound System of Madurese

2.3.1 Madurese consonants

Table 3. Madurese consonant inventory

		Bilabial	Dental/ Alveolar	Retroflex	Palatal	Velar	Glottal
Stops	Unaspirated	p	t	ʈ	c	k	ʔ
	Aspirated	pʰ	tʰ	ʈʰ	cʰ	kʰ	
	Voiced	b	d	ɖ	ɟ	g	
Nasals		m	n		ɲ	ŋ	
Fricative		(f)	s				(h)
Liquids			l r				
Glides		(w)			j		

As shown in Table 3 above, Madurese has 27 consonants, most of which belong to the class of stops. Of the 27 consonants, fifteen are oral stops consisting of three labials, three dentals/alveolars, three retroflexes, three palatals, and three velars; and four belong to nasal stops comprising one labial, one dental/alveolar, one palatal, and one velar. Other consonants existing in Madurese are one labio-dental fricative, one alveolar fricative, one glottal fricative, two dental/alveolar liquids, one labial glide, and one palatal glide. Three consonants in parentheses shown in Table 3 are not considered native to Madurese: /f/, /h/ and /w/. Words beginning with /f/ and /h/ may have been borrowed from Arabic, Malay or Indonesian while those beginning with /w/ may have been borrowed from Javanese and Arabic (Stevens, 1968).

Madurese has a relatively larger consonant inventory than its related languages such as Javanese, Sundanese and Indonesian. This larger consonant inventory derives from the fact that Madurese has a series of stop consonants with a three-way laryngeal contrast: voiced, voiceless unaspirated, and voiceless aspirated. Other related languages such as Javanese, Sundanese and Indonesian only have stops with a two-way contrast. However, despite such a relatively rich inventory, not all Madurese consonants can occur in word-final position. That is, only /p, t, k, m, n, l, r, s, j, ʔ/ occur word-finally. The glottal stop /ʔ/ is the only native consonant which does not occur word-initially and in word-medial or intervocalic position it can only function as a syllable coda (see e.g. Davies, 2010; Stevens, 1968, 1991).

Previous researchers (Cohn & Ham, 1998; Davies, 2010; Stevens, 1968, 1991) agree that there are five places of articulation in Madurese. However, they differ in the way they label them. In this case, Stevens (1968, 1991) names them as labial, dental, alveolar, palatal and velar. Cohn and Ham (1998) add retroflex but they do not distinguish retroflex from alveolar by labelling them labial, dental, retroflex/alveolar, palatal and velar. In contrast, Davies (2010, p. 12) does not distinguish between dental and alveolar, but he distinguishes dental/alveolar from retroflex instead. In this dissertation, we follow Davies (2010), who does not distinguish between dental and alveolar but distinguishes them from retroflex. However, it is important to note that even though dental/alveolar and retroflex are contrastive, the functional load of their contrast is not very high. This is evidenced by there being very few minimal pairs that show their distinctions particularly in word-initial position. Because word-initial retroflex stops are very rare in Madurese, we decided to exclude them from our phonetic analysis. Thus, we focus on four places of articulation, namely bilabial, dental/alveolar, palatal and velar.

Stevens (1968, 1991) and Davies (2010) note that the glides /j/ and /w/ have somewhat special phonological status in Madurese. The glide /j/ in native Madurese words only occurs in word-final position, for example, in words such as [kərpʰuj] ‘water buffalo’, [apɔj] ‘fire’, [laŋ:ɔj] ‘swim’ and [sɔroj] ‘comb’. However, the glide /j/ which occurs in intervocalic environment is not phonemic as it is there as a result of epenthesis, a process which will be further discussed in Section 2.4.1. Moreover,

/j/ in word-initial position can only be found in loanwords, for example, [jəkɛn] ‘convinced’ from Arabic. Unlike /j/, the glide /w/ does not occur in word-final position. However, it has a similarity to the glide /j/ in a way that it is not phonemic in intervocalic position and that in word-initial position it can only be found in loanwords, for example [wəɟip] ‘obligatory’, also borrowed from Arabic.

2.3.2 Madurese vowels

Most previous work agrees that Madurese has eight surface vowel qualities, but researchers differ as to the number of vowel phonemes it has. Such differences may partly arise because some researchers base their distinction of Madurese vowels purely on sounds as found in lexical items while some others base the vowel distinction on a particular phonological analysis of the language. The disagreements also result from the fact that some researchers do not distinguish between native vowels of Madurese and non-native ones that are found in some loanwords.

2.3.2.1 Monophthongs

Table 4. Madurese surface vowels (Stevens, 1980; Cohn and Lockwood 1994)

	Front	Central	Back
High	i	ɨ	u
Mid	ɛ	ɤ ə	ɔ
Low		a	

As shown in Table 4 above, Madurese vowels can basically be grouped into two sets: high vowels [i, ɨ, ɤ, u] and non-high vowels [ɛ, ə, a, ɔ] (Stevens, 1968, 1980, 1991). Stevens (1968, p. 18) suggests that about 95% of the Madurese lexical items in his corpus use these eight surface vowels. He considers the non-high vowels as the underlying vowels because they occur in word-initial position, a position which is considered neutral to the conditioning phonological context and hence a position which is not occupied by the high vowels (Stevens, 1980; 1991, pp. 359–360).

A quite different view with regard to vowel phonemes and their alternations in Madurese is postulated by Anderson (1991). She claims that the ‘default’ vowels in the language consist of three non-high vowels /ɛ, a, ɔ/ which surface as [ɛ, a, ɔ] and [i, ʌ, u] and that there is no distinction between ə and ɨ. Following Kiliaan (1897), Anderson argues that the vowel /ə/ does not alternate and hence it can occur after voiced and voiceless aspirated stops. It is also important to note that Anderson uses the IPA symbol [ʌ] instead of [ɤ]. In contrast, Davies (2010, pp. 36-37) argues that Madurese has six phonemic vowels, namely /ɛ/, /ɔ/, /a/, /ə/, /i/, /u/. Unlike Stevens, Davies includes /i/ and /u/ in the Madurese vowel inventory arguing that they are also found in word-initial position. He shows that these two vowels are particularly found in Madurese loanwords such as [imigrasi] 'immigration' and [uɟiɤn] 'exam'.

Researchers use different symbols in particular for the vowel [ɤ]. The IPA symbol [ɤ] was first used by Stevens (1985) and this is then followed by other researchers such as Trigo (1991), Cohn (1993a, 1993b), and recently Davies (2010). However, Davies (2010, pp. 19-20) notices that the Madurese vowel symbolised with [ɤ] is in fact a mid-close central unrounded vowel, which is normally transcribed using the IPA symbol [ɘ], whereas [ɤ] is the IPA symbol for a mid-close back unrounded vowel instead. Davies suggests deciding to use the symbol in order to conform to the tradition of previous researchers including Stevens (1985), Cohn (1993a, 1993b), and Cohn and Lockwood (1994). In addition, the latest Madurese dictionary written by Pawitra (2009) uses a low central vowel [ɐ] for his phonetic transcription. These differences suggest that the phonetic status of Madurese vowels requires further research. Since previous instrumental studies on Madurese only involved one or two speakers of Madurese, we address this problem here by involving more speakers.

2.3.2.2 CV co-occurrence restrictions

One interesting aspect of Madurese in CV sequences is the fact that high vowels only occur after voiced and voiceless aspirated stops while non-high vowels only occur in word-initial position, after word-initial liquids, and the other consonants. This was first described by Kiliaan (1897) and discussed extensively by Stevens (1968, 1980, 1992), Trigo (1991), Anderson (1991), Cohn (1993a, 1993b), Cohn & Lockwood (1994), and Cohn & Ham (1998). Table 5 on the following page provides examples illustrating the CV co-occurrence restriction involving the alternations of non-high and high vowels. It also provides examples in which only non-high vowels occur. Note that the vowel pair [ə - ɨ] never occurs before retroflex stops and we cannot find Madurese words where the syllable onset is a retroflex followed by the pair.

Table 5. Words illustrating Madurese non-high and high vowel alternations

Place	Vowel	Non-high		High	
Bilabial	[a - ɤ]	paʈɛ	‘coconut milk’	pʰɤʈɛ bɤtaʔ	‘profit’ ‘lift up’
	[ɔ - u]	pɔʈɔ	‘cake’	pʰuʈʰu budu	‘stupid’ ‘stale fish’
	[ɛ - i]	pɛrak	‘happy’	pʰiʈak bisa	‘bird’ ‘able’
	[ə - ɨ]	pəsːɛ	‘money’	pʰɨlːis bɨrːɤʔ	‘furious’ ‘heavy’
Dental/ Alveolar	[a - ɤ]	talam	‘pan’	tʰɤlɨm tʰɤntʰɤn dɤlɨm	‘house’ ‘dress up’ ‘deep’
	[ɔ - u]	tɔrɔt	‘let’	tʰuka duri	‘angry’ ‘thorn’
	[ɛ - i]	tɛla	‘sweet potato’	tʰika dipan	‘you’ ‘bed’
	[ə - ɨ]	təpːaʔ	‘correct’	tʰɨrːɨs dɨpːa	‘heavy’ ‘fathom’
Retroflex	[a - ɤ]	ʈarat	‘scream’	ʈʰɤrɤt ʈʰɤnʈʰɤn	‘land’ ‘draw’
	[ɔ - u]	paʈɔk	‘marker’	pʰɤʈʰuk	‘eat’
	[ɛ - i]	pɔʈɛk	‘cut’	ʈʰimɛn	‘ago’
Palatal	[a - ɤ]	cala	‘defective’	cʰɤlɤ ɟɤgɤ	‘net’ ‘guard’
	[ɔ - u]	cɔcɔk	‘suitable’	cʰukɔʔ ɟuɟuʔ	‘fish’ ‘grandparent’
	[ɛ - i]	cɛrɛt	‘kettle’	cʰilɤ ɟikar	‘tongue’ ‘cart’
	[ə - ɨ]	cəlːɔt	‘clay’	cʰɨcʰːɨl ɟɤdːiŋ	‘insert’ ‘bathroom’
Velar	[a - ɤ]	kanca	‘friend’	kʰɤncʰɤ gɤgɤn	‘pubertal’ ‘unaware’
	[ɔ - u]	kɔraŋ	‘lack’	kʰuriŋ lagu	‘fry’ ‘song’
	[ɛ - i]	kɛkɛr	‘sharpener’	kʰikʰir gibɤs	‘scold’ ‘sheep’
	[ə - ɨ]	kəlːar	‘able’	kʰɨlːɨm	‘willing’
Vowel-initial	[a - ɤ]	anaʔ	‘child’	n.a.	
	[ɔ - u]	ɔrɛŋ	‘man’	n.a.	
	[ɛ - i]	ɛntar	‘go’	n.a.	
	[ə - ɨ]	əlːa	‘don’t’	n.a.	
l and r-initial	[a - ɤ]	lapar raɟɤh	‘hungry’ ‘big’	n.a.	
	[ɔ - u]	lɔrɔŋ rɔsak	‘road’ ‘damaged’	n.a.	
	[ɛ - i]	lɛcaʔ rɛbɤh	‘soft’ ‘lap’	n.a.	
	[ə - ɨ]	ləkːas rəmpɛk	‘quick’ ‘broken’	n.a.	

2.3.2.3 Exceptions

Stevens (1968, pp. 41–45) points out that there are a number of words containing vowels that do not alternate according to the general rule on vowel alternations. However, they are quite rare and appear to be only limited to loanwords given that the vast majority of the Madurese lexicons follow the CV interaction rule. Stevens (1968, p. 18) suggests that more than 95% of his corpus can be accounted for by the rule. Examples of words showing such exceptions are shown in (9) below.

(9)	[bal][5]	‘ball’
	[ban]	‘tyre’
	[baŋ]	‘bank’
	[baŋku]	‘bench’
	[bɛcaʔ]	‘trishaw’
	[bijasa]	‘usual’
	[buku]	‘book’
	[dasi]	‘tie’
	[dɔktər]	‘doctor’
	[dɔmpɛt]	‘wallet’
	[əmba]	‘grandparent’
	[gaŋ]	‘alley’
	[gas]	‘gasoline’
	[kiblat]	‘facing Mecca’
	[kɔpi]	‘coffee’
	[mɔgɔʔ]	‘strike’
	[ɔbat][6]	‘medicine’
	[pɛnsiun]	‘retired’
	[piŋpɔŋ]	‘Ping-Pong’
	[pɔlisi]	‘police’
	[ranɟaŋ]	‘bed’
	[rɔmbɛŋ]	‘old clothes’
	[rɔmbɔŋan]	‘group’
	[sandal]	‘sandal’
	[satrika]	‘iron’
	[susu]	‘milk’
	[tabraʔ]	‘hit’
	[taksi]	‘taxi’
	[tɔpi]	‘hat’
	[udur]	‘hindrance’

---

[5] Some speakers pronounce monosyllabic content words, which are mostly borrowed, by adding [ə] in front of them, for example /bal/ → [əbːal], /gaŋ/ → [əgːaŋ].

[6] Some speakers pronounce this word as [ɔpʰɤt], which follows the general rule.

However, all of these words appear to be borrowed words. For example, the words [baŋ], [dɔktər], [gas], [pɛnsiun], [pɔlisi] and [taksi] are borrowed from Dutch or English while the words [kiblat] and [udur] are borrowed from Arabic. The words such as [dɔmpɛt], [əmba], [rɔmbɔŋan], [tabraʔ] and [tɔpi] are probably borrowed from Malay or Indonesian. Interestingly, a very small number of those words have seemingly native counterparts, which in fact follow the rule, for example [bukɔ] 'joint', [bɤn] 'and' and [sɔsɔ] 'breast'.

2.3.2.4 The status of ɨ

As mentioned earlier, until now the number of Madurese vowels has been debated and scholars have some disagreements with regard to this. Once source of these disagreements probably arises because some scholars do not distinguish between phonemic and phonetic vowels. Among the Madurese vowels mentioned in the literature, it is the status of [ɨ] as a surface vowel which is the most debated. As noted by Davies (2010, p. 37) none of the Indonesian authors recognise the existence of this vowel. Davies points out that other than Stevens (1980, 1992) and Cohn and Lockwood (1994), no scholars postulates the distinction between [ə] and [ɨ]. This is also reflected in the fact that none of the orthographies ever devised for Madurese so far have made a distinction between them (see Table 7 in Section 2.5.1).

In relation to this, two studies have attempted to provide phonetic evidence that [ə] and [ɨ] are distinguishable in the vowel space. Cohn and Lockwood (1994) provide phonetic evidence that these two vowels are only different in terms of their F1 values but they have a very small difference in their F2 values. As mentioned in Davies (2010), Bortscheller (2007) also reports that the vowel space occupied by the vowels [ə] and [ɨ] is relatively distinct for his single speaker. Since they looked at them with only one or two speakers, we will be looking again at the ambiguity surrounding their phonetic realisations based on new phonetic data in the present study.

2.3.2.5 Diphthongs

Madurese has also been suggested to have diphthongs in its vowel system. Unlike monophthongs, diphthongs never occur in word-initial position, but they can occupy other positions in a word. Interestingly, the same as monophthongs, the vowel quality for diphthongs also depends on the consonant preceding them. In connection with this, there are two diphthongs in Madurese, namely /aj/ and /ɔj/, which, depending on the preceding consonant, are realised as [aj - ɤj] and [ɔj - uj] respectively. This is shown in the following examples.

(10)	[aŋkʰɤj]	‘a kind of insect’
	[aŋkʰuj]	‘use’
	[apɔj]	‘fire’
	[bɤrɤkaj]	‘lizard’
	[kərpʰuj]	‘water buffalo’
	[laŋːɔj]	‘swim’
	[pəlːaj]	‘pale’
	[lambɤj]	‘to wave’
	[sɔŋaj]	‘river’
	[sɔrɔj]	‘comb’

The diphthongs can also occur in word-medial position, but this seems to be quite rare. In fact, the diphthongs which occur in word-medial position appear to be limited to reduplication-type words such as [pajpaj] ‘plain’, [bɤjbɤj] ‘too soft’.

However, the phonological status of the Madurese diphthongs mentioned above is questionable. This is because they never occur in word-initial position and their occurrences in word-medial position can also be arguable. The only obvious position for them to occur is in word-final position, which also turns out to be debatable. The reason for this is that in word-final position, the so-called diphthongs appear to be a combination of a vowel and a glide /j/. If we argue against the existence of the phoneme /j/ in word-final position, the sequence of a vowel and a glide can be regarded as a diphthong. In this case, evidence against the existence of Madurese diphthongs may come from the behaviour of the word-final /j/ when it is followed by a vowel-initial suffix. As shown in (11) below, the glide /j/ becomes geminated and becomes the onset of the following syllable.

(11)	[aŋkʰuj] + a	→	[aŋkʰujːɤ]	‘the clothes’
	[kərpʰuj] + a	→	[kərpʰujːɤ]	‘the water buffalo’
	[sɔŋaj] + a	→	[sɔŋajːa]	‘the river’
	[sɔrɔj] + a	→	[sɔrɔjːa]	‘the comb’
	[salɔj] + a	→	[salɔjːa]	‘the mixture’

Also relevant to mention here is the existence of a number of vowel clusters in Madurese. However, vowel clusters are different from diphthongs given that the clusters are pronounced fully like the way each member of the clusters is pronounced in their single forms. Interestingly, vowel clusters of this type do not trigger any segmental epenthesis, which in other cases we find that when two vowels occur in a sequence, they usually result in either glide-insertion or glottal insertion. In this regard, Davies (2010, p. 28) mentions four surface vowel clusters: [aɛ], [aɔ], [ɤi] and [ɤu] and they appear in words shown in (12) below.

(12)	[bɤiʔ]	'seed'[7]
	[bɤu]	'smell'
	[cʰɤi]	'ginger'
	[cʰɤu]	'far'
	[cʰɤlɤuʔ]	'in the south'
	[kaɛʔ]	'hook'
	[laɔʔ]	'south'
	[paɛʔ]	'bitter'
	[paɔ]	'mango'
	[saɛ]	'well'
	[pʰɤi]	'instead'
	[pʰɤu]	'shoulder'

As shown in (12) above, the vowel clusters do not appear to trigger any type of glide insertion. This is because they do not differ in their front-back dimension. The clusters do not trigger glottal epenthesis either given that they are not identical vowels, as will be discussed further in Section 2.4.1 later.

2.3.3 Phonotactics and syllable structure in Madurese

As also noted by Davies (2010, p. 25) and Stevens (1968, pp. 51-52), most roots in Madurese consist of two syllables with CV and CVC being the most common structures. In terms of word categories, the majority of content words are disyllabic. Monosyllabic words are mostly limited to function words and are also associated with borrowings. Some possible syllable structures for monosyllabic and disyllabic words are shown in Table 6 on the following page.

---

[7] Impressionistically the first vowel in the cluster is stressed and longer.

Table 6. Word shapes in Madurese

Word shape	Example
CV	ka ‘to’, la ‘already’, sɛ ‘that/who’
CV.V	paɔ ‘mango’, taɔ ‘know’, cʰɤu ‘far’
CVC	taŋ ‘my’, taʔ ‘not’, kʰiʔ ‘yet’
CV.CV	caca ‘talk’, padɤ ‘same’, sala ‘wrong’, sabɤ ‘rice field’
CV.CVC	bulɤn ‘moon’, kɔlɔr ‘a type of jackfruit’, sabɤn ‘possessed’, bɤrɤs ‘well’, bɤtʰɤl ‘origin’
CVC.CV	kanɟi ‘a kind of flour’, kʰɤntaʔ ‘cricket’, tɛŋka ‘conduct’, pəŋkɔ ‘stubborn’, pəllɔ ‘sweat’, kʰɤncʰɤ ‘pubertal’
CVC.CVC	kəmpʰɤŋ ‘flower’, kampat ‘a kind of crab’, lancʰɤŋ ‘long’, lɔncaʔ ‘jump’, mancʰɨŋ ‘stand up’, kantʰɨl ‘thick’, kʰɤrriŋ ‘sick’
V	ɛ ‘at’
V.VC	aɛŋ ‘water’, ɔɛŋ ‘nod head’, aɛp ‘shame’
V.CV	ɔbu ‘raise’, ɔbɤ ‘change’, ɔpa ‘wage’
VC.CV	əmba ‘grandparent’, anca ‘provoke’, ɔŋkʰɤ ‘rise’
V.CVC	aɟɤm ‘chicken’, anɔm ‘uncle’, ɔtək ‘brain’
VC.CVC	ɛntar ‘visit’, andiʔ ‘have’, aŋkaʔ ‘lift’, əmpaʔ ‘four’

It is worth noting that when pronouncing borrowed monosyllabic words, Madurese people tend to add [ə] to the words. For example, the words *bis* ‘bus’, *ban* ‘tyre’, *kol* ‘pick-up truck’ and *truk* ‘truck’ are pronounced as [əbːis], [əbːan], [əkːɔl], and [ətːruk] respectively. As noted by Davies (2010), the glottal stop cannot occur in word-initial position, but occur in syllable-final position. However, vowels that occur in word-initial position are often glottalised.

Consonant clusters never occur in syllable-initial and syllable-final position in Madurese. The occurrence of initial consonant clusters in native Madurese words is primarily due to vowel reduction and borrowing (Davies, 2010; Stevens, 1968). However, vowel reduction only applies to words that have more than two syllables and the vowel can only undergo vowel reduction if it is preceded by a consonant and followed by an approximant, a liquid or a glide. For example, the word *parabân* [parabɤn] ‘virgin’, which has an initial syllable structure CV, can be pronounced as [prabɤn]; *biasa* [bijasa] ‘usual’ as [bjasa] and *soara* [sɔwara] ‘voice’ as [swara]. Examples of words with consonant clusters from borrowings are *pramuka* [pramuka] ‘scout’ and *prangko* [praŋkɔ] ‘stamp’.

Like the other non-high vowels, the vowel [ə] can occur in word-initial position, and similar to the other high vowels, the vowel [ɨ] cannot occur in word-initial position. Davies (2010, p. 36) provides evidence that [i] and [u] can also occur in word-initial position. However, other than the word [uwɤʔ][8] ‘steam’, it appears that all the words he uses as examples are all recent borrowings from Indonesian. Moreover, [ə] and [ɨ] never occur in syllable-final position. The other vowels can occur in word-medial position with or without geminate consonants and they can also occur in word-final position. It is clear that consonant gemination following [ə] and [ɨ] are predictable while that following the other vowels are not. In another case, the only way these two vowels can occur in syllable-final position when the syllable has a coda. Thus, with regard to disyllabic words, these two vowels require the syllable pattern CVCCVC, where the vowel in the first syllable is followed by either a geminate or a consonant cluster and the second syllable has a coda.

2.3.4 Word stress in Madurese

To my knowledge, word stress is an area that has not been studied in Madurese. However, its close neighbour, Indonesian, has been described as a language with free word stress (van Zanten & van Heuven, 2004; van Zanten & van Heuven, 1998). The following description is based on my intuition as a native speaker. Word stress is never lexically contrastive in Madurese, but we have the intuition that word stress in both disyllabic and trisyllabic words occurs in the first syllables.

The stressed syllables in the following examples are written in bold. For example, disyllabic words such as **pa**raʔ ‘almost’, **pɛn**tər ‘smart’, **pɛ**lak ‘kind’, **pɔ**rɔ ‘ulcer’, **pʰu**ru ‘afterwards’ and trisyllabic words such as **təp**ːaʔan ‘more accurate’, **pa**təpːaʔ ‘make correct’, **pa**lap:a ‘spices’, **kɛ**nɛʔan ‘smaller’ and **pɔs**ːaʔan ‘more full’ all have stress on the first syllables. Word stress for words with four syllables appears to fall on the antepenultimate syllable, for example, ka**bɤ**dɤʔɤn ‘presence’, ka**lɔp**ːaʔan ‘forgotten’, ka**la**kɔan ‘job’, a**sa**pɔan ‘to sweep’, a**ta**retan ‘make brotherly relations’, a**sɔŋ**kɔʔan ‘wear a hat’ and a**pɔ**rɔan ‘to have ulcer’.

---

[8] In Central Madurese, this word is pronounced as [õwãʔ], which clearly obeys the CV interaction rule. It is interesting that the vowel [õ] is nasalised and it raises the question how it occurs as there is no environment contributing to its occurrence, assuming it an allophone of the vowel /a/. However, the reason why the vowel [a] is nasalised here is possibly because it occurs after the vowel [õ] as /w/ is transparent to nasalisation.

It appears that word stress in Madurese is neither dependent nor affected by vowel height. This is due to the fact that even the first syllable of a disyllabic or trisyllabic word contains a schwa vowel, the stress remains on the first syllable, for example **pəl**ːɔ ‘sweat’, **pəl**kaʔ ‘thirsty’, **pəs**ːɛ ‘money’, **təl**ːɔʔ ‘three’, **təl**ːɔʔan ‘there are three’, **sən**ːəŋːan ‘happier’ and **cər**ːɛʔan ‘more stingy’. If we look at the word stress for words with more than three syllables, we see a fairly consistent and regular pattern that it always occurs in the first syllable of the root. For example, bɤdɤ ‘exist’→ ka-**bɤ**dɤʔ-ɤn ‘presence’, sapɔ ‘broom’→ a-**sa**pɔ-an ‘to sweep’.

2.4 Phonological and Morphophonemic Processes

There are a number of phonological processes in Madurese which are relevant to discuss in this section because they are related to the CV co-occurrence restriction and vowel harmony processes. They include epenthesis, gemination and deletion. Epenthesis includes such phonological processes as glottal insertion, j-epenthesis, and glide insertion. In addition, a number of morphophonemic processes have also been identified and are also relevant to discuss for the same reason as the phonological processes mentioned above. They consist of j-epenthesis, nasal substitution, aspiration, gemination and vowel reduction. See Stevens (1968, 1980, 1991) and Davies (2010) for more complete reviews on these aspects.

2.4.1 Epenthesis

Davies (2010) identifies three types of consonantal epenthesis in Madurese, namely glottal stop insertion, j-insertion, and glide-insertion. A glottal stop is inserted when two identical vowels occur in a sequence at either word-internal position or a morpheme boundary. Some examples of glottal insertion occurring at word-internal position are shown in (13) below.

(13) bɤɤ → [bɤʔɤ] ‘flood’[9]
paaʔ → [paʔaʔ] ‘chisel’
saar → [saʔar] ‘coffee residue’
taal → [taʔal] ‘a kind of palm fruit’

---

[9] Postulating that the glottal stop is derived is quite problematic. This is because in this way Madurese is expected to have a long vowel phoneme, which does not appear to be the case (Bert Remijsen, personal communication).

	taat	→	[taʔat]	‘obedient’
	tɔɔt	→	[tɔʔɔt]	‘kneel’
	kɔɔl	→	[kɔʔɔl]	‘a type of snails’

Examples of glottal insertion occurring at a morpheme boundary are shown in (14).

(14)	ɔk^{h}ɤ + an	→	[ɔk^{h}ɤʔɤn]	‘looser’
	bɤca + an	→	[bɤcaʔan]	‘reading’
	maca + an	→	[macaʔan]	‘love reading’
	patɛ + ɛ	→	[patɛʔɛ]	‘to kill’
	sakɛ + ɛ	→	[sakɛʔɛ]	‘to hurt’

The second type of epenthesis is j-insertion. This epenthesis occurs in principally the same environment as that of the glottal stop insertion, but it has a rather limited distribution. Unlike the glottal insertion, the j-epenthesis can only occur at a morpheme or word boundary. At a morpheme boundary, it only involves the prefix /ɛ/, which is a prefix used for indicating passive voice while at a word boundary it only occurs with the proposition /ɛ/. Examples of j-epenthesis occurring at a morpheme boundary are shown in (15) and examples of j-epenthesis occurring at a word boundary are shown in (16) below.

(15)	ɛ + ɛntarɛ	→	[ɛjɛntarɛ]	‘to be visited’
	ɛ + ɛnɔm	→	[ɛjɛnɔm]	‘to be drunk’
	ɛ + ɛncər	→	[ɛjɛncər]	‘to be wanted’

(16)	ɛ + ɛlɔŋ	→	[ɛjɛlɔŋ]	‘at the nose’
	ɛ + ɛpar	→	[ɛjɛpar]	‘at in-laws'
	ɛ + ɛp^{h}u	→	[ɛjɛp^{h}u]	‘at my mother'

As we can see in (15) and (16) above, the two identical vowels do not require the insertion of the glottal stop; they require j-epenthesis instead. Thus, this process is different from the one that we see in (13) and (14) shown earlier, where identical vowels require glottal insertion.

The third type is glide-insertion involving the insertion of either [w] or [j] at word-internal position or at a morpheme boundary. This epenthesis occurs when two vowels which differ in backness occurs next to each other either word-internally or before a suffix. The first vowel determines which glide to be inserted. That is, [j] is inserted after a front vowel while [w] is inserted after a back vowel. Examples of word-internal glide insertion are shown in (17) and examples of glide insertion involving a suffix are shown in (18).

(17) kεaε → [kεjaε] ‘religious teacher’
kʰuɤ → [kʰuwɤ] ‘cave’
kɔat → [kɔwat] ‘strong’
lεaʔ → [lεjaʔ] ‘hard’
rɔa → [rɔwa] ‘that’

(18) εka + tɔa + ε → [εkatɔwaε] ‘to be led’
məl:ε + akʰi → [məl:εjakʰi] ‘to buy for’
sarε + akʰi → [sarεjakʰi] ‘to search for’
tɔpɔ + akʰi → [tɔpɔwakʰi] ‘to cover up’
ε + parlɔ + akʰi → [εparlɔwakʰi] ‘to be needed’

2.4.2 Gemination

Gemination in Madurese can be contrastive as well as non-contrastive. As the name suggests, contrastive gemination occurs when it is not predictable while non-contrastive gemination occurs when its occurrence is contextually conditioned and therefore predictable. The only example of predictable gemination which occurs root-internally involves the vowels [ə] and [ɨ]. These vowels, which constitute a pair of the non-high and high vowels, always trigger gemination in the following consonants, as shown in (19) below.

(19) cələŋ → [cəlːəŋ] ‘black’
pəʈεk → [pəʈːεk] ‘break’
pələm → [pəlːəm] ‘fat’
gɨna → [gɨnːa] ‘proper’
pʰɨlis → [pʰɨlːis] ‘angry’
kʰɨta → [kʰɨtːa] ‘sap’

Davies (2010) and Stevens (1968, pp. 126-127) also identifies two suffixes that trigger gemination occurring across morpheme boundary, namely the benefactive or causative *–aghi* [akʰi] and the definite suffix *-na*. Some examples of gemination which is triggered by the suffix *–aghi* are shown in (20) below.

(20) εntar + akʰi → [εntarːakʰi] ‘go for’
pɔlɔŋ + akʰi → [pɔlɔŋːakʰi] ‘pick for’
pɔtər + akʰi → [pɔtərːakʰi] ‘turn on for’
ɲət:εl + akʰi → [ɲət:εlːakʰi] ‘set for’
ŋεncʰɤm + akʰi → [ŋεncʰɤmːakʰi] ‘borrow for’

The consonant /n/ in the suffix *–na* undergoes a change if the noun to which it will attach ends in a consonant. The change depends on the final consonant of the word that it is preceded. Examples of gemination involving this suffix are shown in (21).

(21)	kɔcɛŋ + na	→	[kɔcɛŋːa]	‘the cat’
	sandal + na	→	[sandalːa]	‘the sandal’
	aɟɤm + na	→	[aɟɤmːa]	‘the chicken’
	pakʰɤr + na	→	[pakʰɤrːɤ]	‘the fence’
	kəʈʰ:ɤŋ + na	→	[kəʈʰːɤŋːa]	‘the banana’
	kɔntak + na	→	[kɔntakʰːɤ]	‘the ignition key’

2.4.3 Nasal substitution

Nasal substitution is a morphophonemic process that has a similarity to what is found in Indonesian and most of the Western Malayo-Polynesian languages as well. Nasal substitution occurs when a stem-initial unaspirated stop or in some cases a stem-initial voiced stop is substituted with its homorganic nasal counterpart following an N-prefix,[10] indicating the ‘actor voice’ form of verbs (Stevens, 1968, 1991). Some examples of this process are shown in (22) below.

(22)	N+bɤbɤ	→	[mabɤ]	‘low’
	N+bɨlːi	→	[məlːɛ]	‘buy’
	N+kakan	→	[ŋakan]	‘eat’
	N+patɛ	→	[matɛ]	‘die’
	N+tɔrɔʔ	→	[nɔrɔʔ]	‘follow’
	N+sɔrɔ	→	[ɲɔrɔ]	‘ask’

Although it is relatively rare, nasal substitution can also be found in a stem beginning with a voiceless aspirated stop such as in (23) below.

(23)	N+cʰucʰːu	→	[ɲɔcʰːu]	‘push’
	N+pʰɤkta	→	[makta]	‘bring’
	N+tʰutʰ:uʔ	→	[nɔtʰːuʔ]	‘finger-point’
	N+pʰukpʰuk	→	[mɔkpʰːuk]	‘hit repeatedly’

As we can see in (23), all vowels following ‘N’ become non-high. It is also based on this nasal substitution process that Cohn (1993b) proposes a binary feature for the consonantal feature, which is in this case [-LL] for nasals. That is, while voiced and

---

[10] The N-prefix can be realised as [m], [n], [ɲ] and [ŋ] depending on the stem-initial stop.

voiceless aspirated stops raise vowels that follow them through vowel raising, nasal consonants that pattern with voiceless unaspirated stops lower the vowels following them through vowel lowering. The nasal substitution process also provides evidence that it is the consonants that trigger either vowel lowering or raising. This once again supports the proposal that there are four underlying vowels in Madurese.

2.4.4 Reduplication in Madurese

Reduplication is another interesting element in Madurese morphology. Madurese reduplication has also been considered unique compared with reduplication in its related languages. Unlike reduplication found in Indonesian, for instance, Madurese reduplication mostly shows a pattern where the last syllable of a word is copied and put this copied syllable before the original word as shown in the following examples.

(24)	pʰɤkʰus	‘good’	→	kʰus-pʰɤkʰus	‘all good’
	pɛntər	‘smart’	→	tər-pɛntər	‘all smart’
	kapʰuru	‘hasty’	→	ru-kapʰuru	‘very hasty’
	kɔrɔs	‘thin’	→	rɔs-kɔrɔs	‘all thin’
	ratʰːin	‘pretty’	→	tʰin-ratʰːin	‘all pretty’
	sɔkʰi	‘rich’	→	kʰi-sɔkʰi	‘all rich’

It is important to note that the copied syllable is exactly the same as the original. Stevens (1968, 1991) and recently Davies (2010, pp. 129-148) also provide a detailed discussion of the reduplication patterns and processes in Madurese.

2.5 Non-phonological Aspects

2.5.1 Madurese orthography

Madurese used to be written using a syllabary originating from the Javanese script called *Aksara Jhaban*, which literally means the Javanese letters. This writing system originally derives from the Grantha- or Palava-script of South India, which also has an indirect relation to the Devanagari script of North India (Adelaar, 2005b, pp. 3–4). Although this system is no longer in use, it remains formally taught at school from grade 1 to grade 12 along with the language. Madurese uses the Roman script and its orthography has also undergone some revisions to make it uniform and easy to read and write. The reader is recommended to read Davies (2010, pp. 51–60), who provides an overview of the history of the language’s writing systems. The writing system we use in the dissertation is the 2008 orthography (see Table 7 below), which is based on the result and recommendation of the 2008 congress held in Pamekasan.

Table 7. The sounds and symbols based on the 2008 Madurese orthography

Sound	Symbol	Sound	Symbol
p	p	n	n
p^h	bh	ɲ	ny
b	b	ŋ	ng
t	t	s	s
t^h	dh	r	r
d	d	l	l
ʈ	ṭ	j	y
ʈ^h	ḍh	w	w
ɖ	ḍ	f	f
c	c	v	v
c^h	jh	z	z
ɟ	j	i	i
k	k	ɛ	è
k^h	gh	a	a
ɡ	g	ɤ	â
ʔ	'	ə	e
h	h	u	u
m	m	ɔ	o

Although effort has been made to improve Madurese people's ability to use the revised orthography, it is common to find Madurese people especially the younger generation who cannot write it properly. Some even find it easier to write Madurese words like the way they write in Indonesian. One difficulty may arise from the fact that there are two consonant clusters used to stand for single sounds in Madurese. For example, the voiceless aspirated stops p^h, t^h, and k^h are orthographically written as 'bh', 'dh', and 'gh' respectively. Another difficulty is probably also due to the orthographic forms of certain vowels. For example, the vowels [ɛ] and [ɨ] are orthographically written as 'è' and 'e' while [a] and [ɤ] are 'a' and 'â', respectively. This may also happen because Madurese is rarely used in written communication.

2.5.2 Speech levels in Madurese

Similar to its neighbouring languages such as Javanese and Sundanese, Madurese also has speech registers. Speech registers refer to choices of words whose uses are dependent on the relations between the speaker and the addressee as well as on the status of the referee. As the term 'register' has been used for referring to phonological register in the dissertation, we use ‘speech level’ to avoid confusion. Madurese speech levels involve two types of systems, namely style and reference levels. The style level is related to the status of and the degree of familiarity between the speaker and the addressee while the reference level is concerned with reference to an honoured and high-status person. The style level consists of *kasar* ‘coarse’, *biasa* ‘ordinary’, *tengnga* ‘middle’ and *alos* ‘refined’ while the reference level consists of *alos tèngghi* ‘high refined’ and *alos mandhâp* ‘low refined’. *Alos tèngghi* words are used to refer to the actions and possessions of an honoured and high-status person whereas *alos mandhâp* words are used to refer to the actions toward an honoured or high-status person (Stevens, 1965).

In order to get an understanding of how the system works, let us look at the following examples. The word *aberri'* [abɨr:iʔ] ‘to give’ is classified as a *biasa* word. If we would like to say that an honoured person gives something to someone, we should use the word *marèngè* [marɛŋɛ], the *alos tèngghi* for that word. However, if we would like to say that we give something to an honoured person, we should say *ngatorè* [ngatɔrɛ], the *alos mandhâp* for the word. The word *ngoca'* [ŋɔcaʔ] ‘to speak’ is a *biasa* word. If we would like to say that an honoured person talks, we should use the *alos tèngghi* style for that word, which is *adhâbu* [aʈʰɤbu]. However, if we would like to say that we talk to an honoured person, then we should use *mator* [matɔr], the *alos mandhâp* for the word.

2.6 Concluding Remarks

This chapter has discussed a number of important issues relating to the phonetics and phonology of Madurese. All of this information provides an important phonological foundation and framework for our analyses later. Some examples of outstanding issues that need to be addressed include particularly the phonetic status of certain Madurese vowels, about which researchers have a disagreement. Since the issues of Madurese vowels and consonants are closely intertwined, we will further address them by looking again at the ambiguity surrounding their phonetic realisations based on new acoustic data from a more representative sample.

3 Phonetics and Phonology of Laryngeal Contrasts

3.1 Introduction

There are two major aspects which we will discuss in this chapter and how they can be particularly relevant to the topics addressed in the dissertation. The first aspect is some issues pertaining to the laryngeal contrast itself. For this purpose, we look at a number of studies which are concerned with both phonetic and phonological aspects of laryngeal contrasts in general and how such contrasts are represented phonologically and manifested acoustically. Thus, phonological aspects such as types of laryngeal contrasts found in languages with two-way, three-way or four-way contrasts, distinctive features and feature specifications of the contrasts are also discussed in this section. This will be followed by further addressing phonetic manifestations of the laryngeal contrasts, particularly focusing on a number of acoustic dimensions including voice onset time (VOT), closure duration, fundamental frequency (F0), vowel quality and voice quality.

The second aspect concerns issues of the phonetics-phonology mapping. This aspect is related to how we should better view the relations between phonological features and their phonetic correlates. One issue of particular interest here includes whether the relationship between phonology and phonetics should be transparent, i.e. whether phonological features are predictable from phonetics, or whether we should take a flexible stance with regard to the phonetics-phonology mapping particularly in conditions where phonological features do not always directly translate into their predicted phonetic correlates.

3.2 Laryngeal Features and Contrasts

Nearly all of the world's languages make at least some type of laryngeal contrast in their stops. Such languages can be broadly divided into three types on the basis of a VOT continuum: languages with a two-way laryngeal contrast, languages with a three-way laryngeal contrast and languages with a four-way laryngeal contrast (Lisker & Abramson, 1964). In this case, languages with a two-way laryngeal distinction constitute the majority (51.1%), followed by a three-way contrast (24%) and a four-way contrast (7.9%) respectively (Maddieson, 1984, p. 26)[11].

Languages with a two-way contrast can be classified into voicing and aspirating languages based on whether or not the vocal folds vibrate during the production of the phonological voiced stop in utterance-initial position, a position considered as free from contextual influence with regard to voicing (Beckman, Jessen, & Ringen, 2013; Jessen, 1996; Jessen & Ringen, 2002). In this regard, English and German are categorised as aspirating languages given that their phonological voiced stops are voiceless in utterance-initial position and their phonological voiceless stops are aspirated in utterance-initial position. In contrast, Dutch, Hungarian and Russian are considered as voicing languages since their phonological voiced stops are prevoiced in utterance-initial position and their voiceless counterparts are not aspirated in utterance-initial position.

The different laryngeal distinctions for these languages are argued to result from different phonological feature specifications. In the case of English and German, the features that are associated with their laryngeal distinctions are [spread glottis] ([sg]) for phonetically voiceless aspirated stops and [no laryngeal specification] ([Ø]) for phonetically voiceless unaspirated stops, while in the case of Dutch, Hungarian and Russian, the features which are relevant for their distinctions are [voice] for phonetically prevoiced stops and [Ø] for phonetically voiceless unaspirated stops (Beckman *et al.*, 2013; Jessen, 1996; Jessen & Ringen, 2002).

In relation to this, a number of phonologists and phoneticians have a disagreement about how to describe the two-way laryngeal contrasts that distinguish between voiced and voiceless unaspirated stops in the case of voicing languages and voiceless unaspirated and voiceless aspirated stops in the case of aspirating languages. Scholars such as Keating (1984), Lombardi (1991), and Kingston and Diehl (1994) represent the laryngeal contrast in both types of languages with the features [voice] and [Ø]. Scholars such as Beckman, Jessen and Ringen (2013), Harris (1994),

---

[11] Maddieson (1984) also mentions languages with one stop series (15.8%), five series (0.6%), and six series (0.6%).

Honeybone (2005), Iverson and Salmon (1995), and Jessen (1996) argue that the laryngeal contrast of stops in aspirating languages such as English, Icelandic and German is represented by the features [spread glottis] and [Ø] whereas the features [voice] and [Ø] are used to described the laryngeal feature of stops in voicing languages such as Dutch, Hungarian, Russian and Spanish. Furthermore, they argue that if some voicing occurs in languages with an [sg] contrast, such voicing occurs as a result of passive voicing, for example when occurring in intervocalic position.

It is worth mentioning that there is another type of languages with a two-way laryngeal contrast which was not identified by Lisker and Abramson (1964). Standard Central (SC) Swedish is such a language. Unlike other languages with a two-way contrast, which distinguish between voiced and voiceless unaspirated stops in the case of voicing languages and voiceless unaspirated and voiceless aspirated stops in the case of aspirating languages, SC Swedish distinguishes between prevoiced and voiceless aspirated stops (Beckman, Helgason, McMurray, & Ringen, 2011; Helgason & Ringen, 2008). Using speech rate as a parameter for determining which features are active, Beckman *et al.* (2011) found that both prevoicing and aspiration in Swedish increase in slow speech, suggesting that both [voice] and [sg] are active and they are, therefore, specified as the laryngeal features in that language.

A relatively similar condition may apply to languages with a three-way laryngeal contrast in terms of what laryngeal features can be used for describing them. Some of these languages have a three-way laryngeal contrast in the voiceless region particularly when the contrasting stop consonants occur in utterance-initial position. A well-known and well-studied example is Korean, which distinguishes between aspirated, lenis, and tense stops[12] (Cho, Ladefoged, & Jun, 2002; Han & Weitzman, 1970; Kang & Guion, 2008; Kang, 2014; Kim & Duanmu, 2004; Kong, Beckman, & Edwards, 2012). Another important example of languages with this type is Shanghai Chinese (Chen, 2011; Gao, 2015; Ren, 1992) where the ‘voiced’ series in this language is realised as ‘voiceless with breathy voice or voiced with aspiration’.

---

[12] Other authors use the terms ‘lax’ instead of ‘lenis’ and ‘fortis’ instead of ‘tense’. For example, Kohler (1984) uses the terms fortis and lenis to distinguish between the laryngeal contrast in stops.

Other languages such as Thai have a three-way laryngeal contrast among voiced, voiceless unaspirated and voiceless aspirated stops (Lisker & Abramson, 1964). Assuming privative features, these languages can be described by the features [sg] for voiceless aspirated stops, [voice] for voiced stops and [Ø] for voiceless unaspirated stops (Beckman *et al.*, 2013; Iverson & Salmons, 1995). These features are particularly relevant to languages with a three-way laryngeal contrast that distinguish between voiced, voiceless aspirated, and voiceless unaspirated stops. However, they are not applicable to languages such as Korean with its three series of stops all being voiceless in utterance-initial position. In this case, the relevant features for Korean would be [constricted glottis] ([cg]) for tense stops, [spread glottis] ([sg]) for aspirated stops and [Ø] for lenis stops instead (Iverson & Salmons, 1995). It could be argued that Eastern Armenian, which distinguishes between voiceless unaspirated, voiceless glottalised and voiceless aspirated stops, has similarity to Korean whereby all of its series of stops are voiceless (Maddieson, 1984). In this case, Honeybone (2005, p. 327) describes the laryngeal contrast in Eastern Armenian with the features [Ø] for voiceless unaspirated stops, [cg] for voiceless glottalised stops and [sg] for voiceless aspirated stops.

As mentioned earlier, languages with a four-way laryngeal contrast are relatively rare compared with languages with two-way and three-way laryngeal contrasts. A very well-known example of languages of this type is Hindi, which is primarily spoken in India. Hindi distinguishes voiced unaspirated, voiced aspirated, voiceless unaspirated and voiceless aspirated stops. Assuming privative features, the features that can be used to describe the four-way distinctions in Hindi are [sg] for voiceless aspirated stops, [voice] for voiced stops and [Ø] for voiceless unaspirated stops, while for voiced aspirated stops, which are also known as breathy voiced stops, the features are [voice] and [sg] (Beckman *et al.*, 2013; Iverson & Salmons, 1995).

In connection with this, there have been debates on how the value of laryngeal features should be represented phonologically. A number of scholars such as Halle and Stevens (1971), Keating (1984, 1988, 1990), Kingston and Diehl (1994), and Kingston *et al.* (2008) represent them using binary features while some others such as Beckman *et al.* (2013), Jessen and Ringen (2002), Lombardi (1991, 1995), Mester and Itô (1989), and Iverson and Salmon (1995) represent them using privative features. Those who support privativity in feature representations argue that privative features are more relevant since they not only represent the features that are actively present in phonology but also are involved in phonological processes such as laryngeal assimilation (Iverson & Salmons, 1995; Kulikov, 2012; Mester & Itô, 1989; Ringen & Kulikov, 2012).

In this case, it has been suggested that features that are considered phonologically active in languages are relatively unaffected by their environments and one way for determining which features are active in a language is by looking at segments in utterance-initial position. For example, voiced stops in languages such as Spanish, Japanese and Russian are prevoiced in utterance-initial position while their voiceless stops are voiceless unaspirated. The situation is quite different for languages such as English and German where voicing is considered to be passive since it rarely occurs in utterance-initial position (Beckman *et al.,* 2013; Iverson & Salmons, 1995; Jessen & Ringen, 2002; Kulikov, 2012). In fact, even in intersonorant position, only 62.5% of the German lenis stops show more than 90% prevoicing whereas 97% of the Russian intervocalic lenis stops are fully voiced (Beckman *et al.,* 2013). Beckman *et al.* use this finding as empirical evidence in support of the idea that the active laryngeal feature in German is the feature [sg] while that in Russian is [voice].

Thus, it appears that proponents of 'laryngeal realism' (e.g. Beckman *et al.,* 2013; Harris, 1994; Honeybone, 2005; Iverson and Salmon, 1995; and Jessen, 1996) also tend to assume a more transparent phonetics-phonology mapping, where stops that bear the feature [cg], for instance, are expected to show acoustic characteristics of being produced as [cg]. This is in contrast to positions like that of Keating (1984, 1988, 1990) or Kingston and Diehl (1994), who tend to assume a more flexible phonetics-phonology mapping. That is, the phonetic realisation of a single feature, e.g. [voice], is allowed to vary freely or vary systematically among languages.

3.3 Acoustic Correlates of Laryngeal Contrasts

In connection with the laryngeal categories discussed in the previous section, a number of studies have looked at their acoustic correlates. The following section discusses some common acoustic correlates used for distinguishing stops in languages demonstrating different laryngeal contrasts.

3.3.1 Voice onset time (VOT)

Voice onset time (VOT) is defined as the temporal interval between the release of the oral constriction for plosive production and the onset of the vibration of the vocal folds (Lisker & Abramson, 1964). Similarly, Ladefoged (2001) defines VOT as the interval between the release of a stop and the start of a following vowel. In this respect, Lisker and Abramson (1964) divide VOT into two types, namely positive VOT and negative VOT. Positive VOT, also called ‘voicing lag’, occurs when the vocal-fold activity starts after the release of the stop closure, while negative VOT, also called ‘voicing lead’, takes place prior to the closure release. On the basis of their study of initial prevocalic stops in eleven languages, Lisker and Abramson categorise stops across those languages into three types: (1) voiceless unaspirated stops (lag of 0 – 25 ms), (2) voiceless aspirated stops (lag of 60 – 100 ms) and (3) voiced stops, in which the vibration onset starts before the release of the stop closure.

On the basis of the number of stop categories, Lisker and Abramson (1964) also classifies the languages they studied into three types: (1) two-category languages such as American English, Cantonese, Dutch, Hungarian, Puerto Rican Spanish and Tamil; (3) three-category languages such as Korean, Eastern Armenian and Thai; (3) four-category languages such as Hindi and Marathi. For example, Lisker and Abramson (1964) found that Spanish has negative VOTs for /b, d, g/ and short positive VOTs for /p, t, k/ in word-initial position whereas word-initial stops /b, d, g/ and /p, t, k/ in English exhibit short and long positive VOTs respectively. The study suggests that languages can have the same number of voicing distinctions. However, they may differ in how they phonetically manifest the voicing categories associated with VOT values characterising each of the categories they have.

VOT is generally found to be much greater in velar stops than in bilabial stops whereas coronal stops have typical intermediate values. The reason for velar stops to have a longer VOT than alveolar and bilabial stops may be due to the different size of the supralaryngeal cavity behind the constriction (Abdelli-Beruh, 2009; Cho & Ladefoged, 1999). That is, for velar stops, greater air pressure builds up quickly in the vocal tract because the supraglottal cavity becomes smaller and it takes longer for the pressure to fall at the beginning of the release phase. Another factor that may also contribute to differences in VOT is vowel quality. It has been suggested that VOT is longer before tense vowels and shorter before lax vowels (Port & Rotunno, 1979). Furthermore, VOT is found to be longer before high vowels than before low vowels in other prevoiced languages such as Hungarian (Gósy, 2001) and Canadian French (Nearey & Rochet, 1994).

Cho and Ladefoged (1999) discuss VOT in a number of languages by primarily focusing on differences among voiceless unaspirated and aspirated stop consonants. Their main concern is how VOT varies with place of articulation and which among the variations are due to physiological adjustment. They found that velar stops have the longest VOTs in 13 languages that do not distinguish between velars and uvulars. They also found that there is no significant difference in VOT between dental and alveolar stops. The differences between bilabial and coronal stops are also not significant. Furthermore, there is no significant difference between the mean VOT of the unaspirated bilabial stops and that of coronal stops, consistent with what is found in other languages (e.g. Abramson & Lisker, 1971; Lisker & Abramson, 1964).

Focusing on unaspirated and aspirated velar stops in a number of languages, Cho and Ladefoged (1999) suggest that it would be possible to draw an arbitrary line at 50 ms to separate voiceless unaspirated from voiceless aspirated stops although it is not obvious that languages can choose only two phonetic categories. On the other hand, they also suggest that it would be possible to group phonetic categories into four: voiceless unaspirated (30 ms), slightly aspirated (50 ms), aspirated (90 ms), and highly aspirated (above 90 ms). No phonological reason can be suggested why there are four categories as they are not indicative of the number of voicing contrasts that individual languages may have. In fact, there are only three modal values of VOT, i.e. [voiced], [voiceless unaspirated], and [voiceless aspirated], suggesting that no languages have more than three contrasts on the basis of VOT dimension (Cho & Ladefoged, 1999; Keating, 1984; Lisker & Abramson, 1964).

3.3.2 Stop closure duration

There is a correlation between laryngeal contrasts and stop closure duration. Indeed, there is a widespread tendency in the world's languages that voiced stops and voiced obstruents in general are shorter than their voiceless counterparts (Lehiste, 1970, p. 22ff.). However, Jessen (2001, pp. 258–259) suggests that such a correlation may also depend on whether a language is a voicing language with the feature [voice], which distinguishes between voiced and voiceless unaspirated stops or an aspirating one with the feature [sg], which distinguishes between voiceless unaspirated and voiceless aspirated stops. Specifically, in a [voice] language where closure duration is particularly unambiguous because there is voicing during closure, it is reasonable that voiced stops have shorter closure duration than voiceless stops. Evidence of this can be found in languages such as French (Abdelli-Beruh, 2004; Laeufer, 1992), Arabic (Alghamdi, 1990, pp. 110–114) and Japanese (Tsuchida, 1997, pp. 111–119).

Furthermore, there is also evidence that the same pattern of closure duration can also be seen in Hindi, a language with a four-way laryngeal contrast. Both voiced unaspirated and voiced aspirated stops in this language exhibit shorter duration in comparison with their voiceless unaspirated and voiceless aspirated counterparts (Benguerel & Bhatia, 1980, p. 140; Kagaya & Hirose, 1975, p. 37). However, it is important to mention that Rome Italian shows a case where the duration of /b/ is much longer than that of /p/ (Hualde & Nadeu, 2011).

3.3.3 Stop voicing during closure

Voicing during closure has also been suggested as another correlate of laryngeal contrast in aspirating languages such as English and German and voicing languages such as Hungarian and Russian. When occurring in intervocalic or intersonorant contexts, voiceless unaspirated (lenis) stops in aspirating languages also show some voicing during closure, but their voicing is variable compared with that of intervocalic lenis stops in voicing languages.

Docherty found that in British English only 50% of word-initial lenis stops in an intervocalic context were voiced during the entire closure whereas the remainder showed interrupted or broken voicing toward the closure. Similarly, Jessen and Ringen (2002) observed that intervocalic lenis stops in German did not exhibit robust voicing during closure. Beckman *et al.* (2013) recently reported that only 62.5% of the intervocalic lenis stops in German showed voicing of over 90% during closure. They use this as evidence in support of the fact that the feature of contrast in German is [sg] instead of [voice] and attribute some voicing in the intersonorant context to passive voicing. However, such an intervocalic voicing pattern is not found in voicing languages such as Hungarian and Russian. In fact, Gósy and Ringen (2009) reported that 95.5% of the intervocalic lenis stops in Hungarian were fully voiced. Similarly, Ringen and Kulikov (2012) found that over 97% of the intervocalic lenis stops in Russian were produced with full voicing. They argue that since there is no variation in the intervocalic lenis stops in Hungarian and Russian, the feature of contrast in the languages is [voice]. That is, unlike intervocalic voicing in German lenis stops, which occurs passively, intervocalic voicing in Hungarian and Russian lenis stops is suggested to be associated with active voicing.

3.3.4 Fundamental frequency (F0)

Fundamental frequency (F0) has been considered as one of important acoustic correlates of laryngeal contrasts. That is, voiced stops are associated with lower F0 whereas voiceless stops are associated with higher F0. In fact, the covariation of VOT and F0 has been observed cross-linguistically (Hombert & Ladefoged, 1976; House & Fairbanks, 1953; Löfqvist, Baer, McGarr, & Story, 1989; Ohde, 1984).

In relation to this, two approaches have attempted to account for the relationship between VOT and F0, namely ‘automatic’ and ‘controlled’ approaches. According to the ‘automatic’ approach, the correlation between VOT and F0 is automatic and determined physiologically. That is, the effect of voicing on VOT and F0 is considered an automatic product of different articulatory and aerodynamic configurations in the production of voicing and is not as a result of the speaker’s direct control (Hombert, Ohala, & Ewan, 1979; Löfqvist *et al.,* 1989). In contrast, the ‘controlled’ approach suggests that the relationship between the two acoustic cues is in some sense deliberate and phonologically dependent (Keating, 1984; Kingston, 2007; Kingston & Diehl, 1994). This approach maintains that the F0 onset is used to perceptually enhance the difference in voicing between voiced and voiceless stops regardless of any other aspects of their phonetic realisation.

Consistent with the phonological perspective, Ohde (1984) observed that there is a covariation between voicing and F0 involving both types of stops (voiceless unaspirated and voiceless aspirated) in English even though in word-initial position they are phonetically voiceless. A similar pattern is also found in Korean where F0 is used as a cue to differentiate between lenis and aspirated stops, both of which are voiceless in utterance-initial position (Cho *et al.*, 2002; Kang, 2014; Silva, 2006).

Furthermore, Dmitrieva *et al.* (2015) provides new evidence that F0 onset in English and Spanish is determined by phonological voicing categories. They found that F0 onset across both languages was significantly higher for voiceless stops than for voiced stops. They also provide evidence that the correlation between voicing categories and F0 onset in these languages seems to follow phonological specifications rather than purely phonetic ones. Specifically, English does not distinguish between phonetically lead voicing and short lag stops and that is why there is no significant difference in F0 onset between these stop types. However, there is a significant difference in F0 onset for English short lag and long lag stops, which in fact contrast phonologically. In contrast, as Spanish distinguishes between lead voicing and short lag stops, their F0 onset does exhibit a significant difference; that is, short lag stops have higher onset F0 than lead voicing ones.

3.3.5 F1 onset

It has been established that the first formant (F1) transition and frequency at the onset of voicing are important acoustic cues for voicing distinction (Kluender, 1991; Pind, 1999; Stevens & Klatt, 1974; Summerfield & Haggard, 1977). In particular, F1 at vowel onset constitutes one acoustic parameter that can be used to distinguish laryngeal contrasts; that is, F1 is higher following aspirated stops than following voiceless unaspirated stops. Jessen (2001, p. 253) mentions two factors which contribute to F1 following voiceless aspirated stops being higher than that following voiceless unaspirated ones. The first factor is related to aspiration, which is due to the fact that aspiration partially masks the formant transitions. That is, because the onset of voicing occurs much later after the stop release in voiceless aspirated stops, the excitation of F1 will not occur until very late in the CV transition, during which the vocal tract almost reaches the vowel steady-state condition. This is not the case for voiceless unaspirated stops in which voicing usually starts relatively at the same time as, or shortly after the stop release. This allows for acoustic energy from the vibration of the vocal folds to affect F1 throughout the CV transition (Benki, 2001).

The second factor is associated with the effect of the trachea as a resonator. Such a tracheal coupling also results in a broadening in F1 bandwidth where broadening bandwidth of a formant decreases its amplitude and consequently its perceptibility. Evidence of this case comes from Danish, in which voiceless unaspirated stops have low F1 onset while voiceless aspirated counterparts have higher F1 at vowel onset (Fischer-Jørgensen, 1968). Similar evidence is found in Mandarin Chinese where voiceless aspirated stops have higher F1 than voiceless unaspirated stops (Shimizu, 1996, pp. 61–63). In the case of languages that distinguish the feature [voice], Jessen (2001) suggests that F1 onset is higher after voiceless unaspirated stops than after voiced stops. For example, F1 onset following voiceless unaspirated stops is higher than following voiced stops in French (Fischer-Jørgensen, 1968).

3.3.6 Phonation type

Phonation type has also been associated with laryngeal contrast. This makes sense considering the fact that what we perceive as laryngeal contrast is in fact the results of certain configurations of the vocal folds inside the larynx together with other associated muscles. For example, the vocal folds can be adducted or abducted, slackened or stretched to produce sounds with different laryngeal distinctions and these are made possible by the actions of intrinsic and extrinsic laryngeal muscles. For this reason, we will discuss a number of acoustic measures that have been used to measure and distinguish voice quality associated with certain phonation types which can be realised in the following vowel.

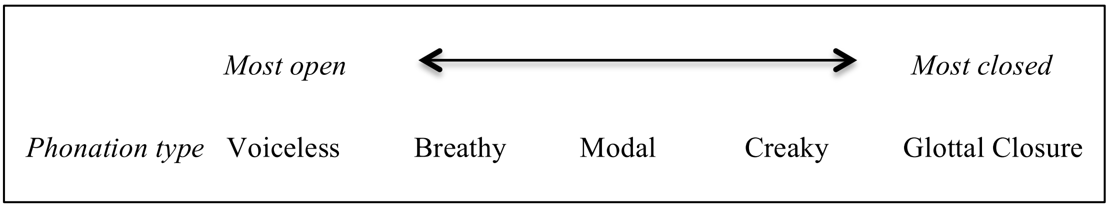


Figure 3–1. Continuum of phonation types (from Gordon & Ladefoged 2001)

Ladefoged (1971) proposes a phonation type continuum on the basis of the degree of the aperture of the glottis. This may range from the most open glottis that occurs during the production of a voiceless sound through the most closed one that occurs during the production of a glottal stop. This continuum is reproduced and shown in Figure 3–1 above. He describes the characteristics of breathy, modal and creaky phonations in the following way. Breathy phonation is characterised by the vocal folds that show both minimal adductive and little longitudinal tension, resulting in less contact during the vibration. Modal phonation is produced by the vocal folds that show regular adductive tension and longitudinal tension. Creaky phonation occurs when the vocal fold configuration shows the highest degree of closure, high adductive tension, but little longitudinal tension. In short, these three phonation types can be basically characterised by the different degree of the vocal fold aperture. That is, creaky phonation has the smallest aperture followed by modal phonation and breathy phonation respectively.

Laver (1980) provides a more detailed explanation about phonetic aspects of these and other phonation types. In addition, a review with regard to phonation types across languages can also be found in Gordon and Ladefoged (2001) and a more recent development on this topic can be found in Edmondson and Esling (2006). However, it is important to note that Edmondson and Esling (2006) call into question the phonation continuum suggested in Ladefoged (1971) and Gordon and Ladefoged (2001). They propose a set of six valves that are considered responsible for producing phonological contrast in many languages. The valves consist of vocal fold adduction and abduction (valve 1), ventricular incursion (valve 2), upward and forward sphincteric compression of the arytenoids and aryepiglottic folds (valve 3), epiglotto-pharyngeal constriction (valve 4), laryngeal raising and lowering (valve 5), and pharyngeal narrowing (valve 6). They argue that the valves do not constitute a glottal continuum but rather a representation of a ‘synergistic and hierarchical system of laryngeal articulations’.

In the following, we will discuss a number of acoustic correlates of voice quality which are commonly used for distinguishing phonation types in a variety of languages. We look at voice quality because we want to see whether voiced and voiceless aspirated stops share certain acoustic correlates of voice quality which we hope could shed light on why they pattern together in the CV co-occurrence restriction in Madurese. For this purpose, we look into acoustic parameters which are commonly and successfully used to distinguish phonation types in a number of languages. The reason why we focus on acoustics in our study because acoustic measures are easy to acquire and often correlate well with physiological parameters.

It is important to note that in addition to looking at acoustic signals taken from Fast Fourier Transform (FFT) spectra (see Figure 3–2 on the following page), there are other important methods available and quite often used for measuring voice quality. These include electroglottography, laryngoscopy and endoscopy, which all look at the glottal source directly. The use of laryngoscopic techniques, for example, is inevitably invasive and can necessarily cause physical discomfort to participants. In terms of practicality, these techniques are also difficult to be carried out on large number of participants in the field. For this reason, voice quality measurements which look at frequency distributions of vowel spectra have become an alternative and in fact a very common practice in the field of phonetics.

The technique for voice quality measurement by looking at vowel spectra has become popular and indeed widely used following the works of Klatt and Klatt (1990), Hillenbrand *et al.* (1994), Hanson (1997), Watkins (1997), Hanson and Chuang (1999), Gordon and Ladefoged (2001), and Blankenship (2002). Recently it has also been developed and used by a number of researchers such as Keating and Esposito (2006), Iseli *et al.* (2007), DiCanio (2009), Kreiman *et al.* (2010), Garellek and Keating (2011), Esposito (2010b, 2012), and Esposito and Khan (2012).

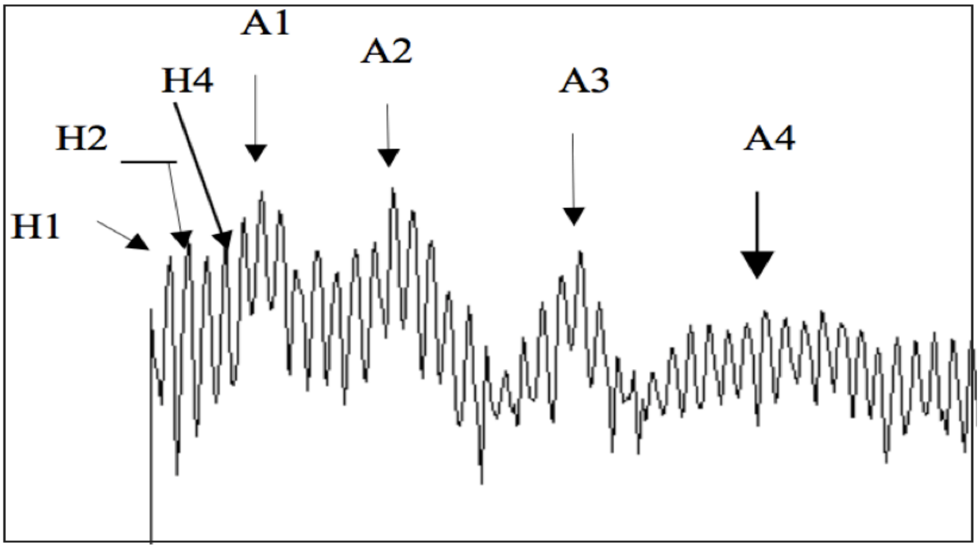


Figure 3–2. FFT of a modal vowel with labels showing A1, A2, A3, A4, H1, H2 and H4 (from Esposito 2006).

There are a number of spectral measurements which have been successfully used to measure and distinguish phonation types in general and modal versus breathy voice in particular. They include H1-H2, which is the relative difference between the amplitudes of the first harmonic (H1) and the second harmonic (H2), H1-A1, which is the relative difference between the amplitudes of the first harmonic and the most prominent harmonic in the F1 region (A1), H1-A2, the relative difference between the amplitudes of the first harmonic and the most prominent harmonic in the F2 region, and H1-A3, which is the relative difference between the amplitudes of the first harmonic and the most prominent harmonic in the F3 region. In addition, other spectral measures include H2-H4, cepstral peak prominence (CPP) and harmonic-to-noise ratio (HNR), the latter two of which constitute measures of periodicity.

Before proceeding with the discussion of these acoustic measures for voice quality and why they are affected, for example, by the breathy-modal distinction, it is important to note that there are some limitations pertaining to such spectral measures. For example, the frequency of the first formant may affect the levels of the harmonics, and they are also generally sensitive to changes in F0. However, these issues have been addressed by Hanson (1997) and later developed by Iseli and Alwan (2004) and Iseli *et al.* (2007) by making further corrections to remove or at least to minimise the effects of formant frequencies on the harmonics.

Table 8. Spectral measures distinguishing phonation types in some languages

Spectral Measures	Languages
H1-H2	!Xóõ (Ladefoged, 1983), Mazatec (Blankenship, 2002), Chanthaburi Khmer (Wayland & Jongman, 2001, 2003), Javanese (Thurgood, 2004), Takhiang Thong Chong (DiCanio, 2009), Green Hmong (Andruski & Ratliff, 2000), Hmong (Huffman, 1987), female speakers of Santa Ana del Valle Zapotec (Esposito, 2004), Gujarati (Fischer-Jørgensen, 1967), Jalapa Mazatec (Garellek & Keating, 2011), Gujarati and White Hmong (Esposito & Khan, 2012)
H1-A1	Chanthaburi Khmer (Wayland & Jongman, 2001, 2003), Jalapa Mazatec (Garellek & Keating, 2011)
H1-A2	Chong and Mazatec (Blankenship, 2002), Jalapa Mazatec (Garellek & Keating, 2011)
H1-A3	Male speakers of Santa Ana del Valle Zapotec (Esposito, 2004), Takhiang Thong Chong (DiCanio, 2009), Chanthaburi Khmer (Wayland & Jongman, 2001), Gujarati and White Hmong (Esposito & Khan, 2012)
HNR	Chanthaburi Khmer (Wayland & Jongman, 2001)
CPP	English (Hillenbrand *et al.*, 1994), Chong and Mazatec (Blankenship, 1997), Jalapa Mazatec (Garellek & Keating, 2011), Gujarati and White Hmong (Esposito & Khan, 2012).

3.3.6.1 H1-H2

H1-H2 is the relative difference between the amplitudes of the first harmonic (H1) and the second harmonic (H2). H1-H2 is an acoustic correlate of the open quotient (OQ), indicating the percentage of the glottal cycle during which the glottis is open (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995). The mechanism that may explain the correlation of OQ with H1-H2 is that the greater the open quotient (i.e. the longer the vocal folds are abducted), the greater the amplitude of the first harmonic relative to the amplitude of the second harmonic. In this case, H1-H2 for breathy vowels is expected to be greater than for modal vowels.

A number of languages that distinguish between breathy and non-breathy vowels have been shown to have different H1-H2 values. In his study on the distinction between breathy and modal vowels in !Xóõ, Ladefoged (1983) found that H1-H2 consistently distinguished between these two phonation types. That is, the breathy vowels in that language show greater H1-H2 than modal vowels. H1-H2 is also found to distinguish phonation types in Mazatec; that is, Mazatec breathy vowels have consistently higher H1-H2 compared with its other phonation types (Blankenship, 2002). H1-H2 also distinguishes breathy vs. modal, laryngealised vs. modal and breathy vs. laryngealised in Jalapa Mazatec (Garellek & Keating, 2011). A more recent study by Gao (2015), who took a look at the difference between voiced and voiceless onsets[13] indicating tone registers in Shanghai Chinese, also found that H1-H2 successfully distinguished between voiced and voiceless onsets. That is, H1-H2 for voiced onset was found to be higher than H1-H2 for voiceless onset, which corroborates the traditional description of the voiced series as ‘muddy’.

Although there are some variations among speakers, H1-H2 also successfully distinguished between breathy and clear vowels in Chanthaburi Khmer (Wayland & Jongman, 2001, 2003). Similarly, in her study on Santa Ana Del Valle Zapotec, Esposito (2010b) found that H1-H2 successfully distinguished three phonation categories (breathy, modal and creaky) for female speakers only while H1-A3 did not. In contrast, she also found that H1-A3 successfully distinguished the three phonation types for male speakers. She suggests that the successful uses of H1-H2 for females and H1-A3 for males may indicate that there is a difference in how phonation is produced between genders and this is probably associated with physiological and sociolinguistic factors. This is in line with what Blankenship (2002) shows in her study that spectral measures do not all succeed in distinguishing phonation types in the languages she examined.

In fact, the success of the spectral measures in distinguishing phonation types may depend on the language, vowel quality, dialect, tone, gender and other factors. In terms of tone, for example, H1-H2 more reliably distinguished phonation types for vowels with high tone than for those with mid or low tone in Mpi (Blankenship, 2002). Another example also comes from Javanese where H1-H2 can only distinguish breathiness for the vowel [u], but it does not distinguish [a] and [ɔ]. However, the three vowels can all successfully be distinguished by H1-A2

---

[13] Gao (2015) uses the traditional terms ‘yin’ to refer to syllables with voiceless onsets that only co-occur with tones that begin with the high F0 register and ‘yang’ to refer to syllables with voiced onsets that only co-occur with tones that start with the low F0 register.

(Thurgood, 2004). In addition, an example comes from Takhian Thong Chong where H1-H2 was found to be a reliable indicator of the distinction between tense and non-tense phonation (DiCanio, 2009). This is unexpected because H1-H2 normally indicates breathiness rather than tenseness.

3.3.6.2 H1-A1

H1-A1 is the relative difference between the amplitudes of the first harmonic (H1) and the most prominent harmonic in the F1 region (A1). This acoustic parameter (A1) indicates F1 bandwidth. Formant bandwidths have been associated with some energy losses in the vocal tract due to such factors as the yielding walls' resistance of the vocal tract, conduction of heat and losses at the walls due to frictions (Stevens & Hanson, 1995). The airflow that goes through the open glottis triggers glottal resistance and this can subsequently contribute to the loss of energy adding up significantly to the F1 bandwidth (Stevens & Hanson, 1995).

House and Stevens (1958) found an increase in bandwidth for their male subjects with the open glottis condition. Hanson (1997) suggests that the result for the F1 bandwidth measurement may provide an indirect indication of the extent to which the glottis undergoes a failure for a complete closure during a glottal cycle of the vocal fold vibration. In this case, a breathy phonation is expected to result in greater H1-A1 than a modal phonation. In other words, breathy phonation is predicted to have a relatively lower or less prominent F1 peak (or lower A1), indicating a greater F1 bandwidth. For example, this measure has been shown to successfully distinguish between breathy and clear vowels in Chanthaburi Khmer (Wayland & Jongman, 2001, 2003). H1-A1 also distinguishes between breathy vs. modal, laryngealised vs. modal and breathy vs. laryngealised in Jalapa Mazatec (Garellek & Keating, 2011).

3.3.6.3 H1-A2 and H1-A3

There are two acoustic parameters that are commonly used for measuring spectral tilt or spectral balance, namely H1-A2 and H1-A3. Stevens (1977) suggests that the slope of the source spectrum has a correlation with the abruptness or gradualness of the vocal fold closure. That is to say, the vocal folds which come together in a gradual fashion primarily causes an excitation of the lower frequencies of the vocal tract, resulting in a spectrum with a steep slope whose energy is mostly concentrated in the region close to the fundamental frequency while very little energy is found at higher frequencies. On the other hand, the vocal folds which come together simultaneously may provide a sufficient excitation on a wider range of frequencies, resulting in a less steep spectrum whose higher frequency components are relatively stronger. Since breathy phonation is characterised by the vocal folds with a gradual closure, the fundamental frequency is expected to be much higher in amplitude than the higher harmonics. That is, the measurement results of spectral slopes for breathy phonation will tend to be mostly positive. In other words, both H1-A2 and H1-A3 are expected to be higher for breathy vowels than for modal vowels.

H1-A2 was found to distinguish modal, breathy, and laryngealised vowels in Mazatec; that is, Mazatec breathy vowels have consistently higher H1-A2 than its other phonation types (Blankenship, 2002). H1-A2 also successfully distinguished between breathy vs. modal, laryngealised vs. modal and breathy vs. laryngealised in Jalapa Mazatec (Garellek & Keating, 2011). As mentioned earlier, H1-A3 successfully distinguished between phonation types in Santa Ana del Valle Zapotec only for males (Esposito, 2010b). In contrast, breathy and clear vowels in Chanthaburi Khmer can be distinguished by H1-A3 for females, but this acoustic measure does not distinguish breathy and clear vowels for males (Wayland & Jongman, 2001). Breathy and non-breathy phonation in Takhian Thong Chong can be distinguished by H1-A3 as well (DiCanio, 2009).

3.3.6.4 Cepstral peak prominence (CPP)

Cepstral peak prominence (CPP), which is a measure of the signal strength over noise across the spectrum, is another measure of periodicity and has also been used to measure breathiness. A well-defined harmonic structure indicates a high periodicity of a signal, which results in a signal having a more prominent cepstral peak than a less periodic one (Hillenbrand *et al.,* 1994). Since breathy phonation has less distinct harmonics, it is expected that it has lower CPP values than modal phonation does. The CPP measure has reliably measured the aperiodicity of breathy phonation in English (Hillenbrand *et al.*, 1994) as well as in other languages such as

Mazatec and Chong (Blankenship, 1997). It also distinguishes between breathy vs. modal and breathy vs. laryngealised in Jalapa Mazatec (Garellek & Keating, 2011).

Moreover, CPP distinguishes between post-aspirated, breathy and modal vowels in some part of the vowel in Gujarati, with post-aspirated vowels appearing to have the lowest CPP, modal vowels to have the highest and breathy vowels to be in between (Esposito & Khan, 2012). CPP also differentiates between breathy and modal vowels in White Hmong, but it does not distinguish between post-aspirated and modal vowels Interestingly, unlike in Gujarati, post-aspirated vowels have lower CPP than breathy vowels (Esposito & Khan, 2012).

3.3.6.5 Harmonic-to-noise ratios (HNR)

Breathiness can also be measured using harmonic-to-noise ratios (HNR). This acoustic measure has been associated with a glottal opening that generates noise during a breathy voice production. Higher HNR indicates modal phonation whereas lower HNR indicates breathy phonation, which is due to increased noise in the spectrum. One acoustic consequence of this should be indicated by the existence of noise or aperiodicity at higher frequency region of the spectrum (Klatt & Klatt, 1990). For example, HNR was used to measure the distinction between breathy and modal vowels in Chanthaburi Khmer (Wayland & Jongman, 2001). Wayland and Jongman found that HNRs did not distinguish breathy and modal vowels for female speakers. However, it distinguished the vowel types for male speakers although the distinction was quite unexpected because breathy vowels had higher HNRs than clear vowels. Another example comes from a recent study by Gao (2015) who looked at the difference between voiced and voiceless onsets in Shanghai Chinese. Using HNR, she found that voiced and voiceless onsets could also be distinguished by this measure of periodicity. That is, voiced onset was found to have higher HNR than voiceless onset, indicating that the latter is unexpectedly breathier than the former.

3.3.6.6 Vowel duration

Another acoustic correlate of voice quality is vowel duration. For example, in Gujarati and Jalapa Mazatec breathy vowels are longer than modal and creaky vowels (Fischer-Jørgensen, 1967; Kirk, Ladefoged, & Ladefoged, 1984). Gordon and

Ladefoged (2001) suggest that vowels produced with non-modal phonation generally show longer duration compared to vowels with modal phonation. It is also suggested that the reason why breathy vowels tend to have longer duration is in order that the listener has more time to perceive the voice quality in the vowel (Silverman, 1997). DiCanio (2009) provides a historical account on this matter, suggesting that since breathy vowels often historically develop from aspirated stops, the loss of aspiration duration is probably compensated for via vowel lengthening.

3.3.7 Voice quality and vowel height

There is a correlation between voice quality and vowel height (F1). For example, a number of scholars have shown that breathy vowels tend to be relatively higher, acoustically lower F1 while tense vowels have the tendency to be relatively lower, acoustically higher F1 (Denning, 1989; Henderson, 1952; Hombert, 1978; Huffman, 1976). Although this is not as common as the correlation between voice quality and vowel height, a correlation between voice quality and vowel fronting (F2) is also found. This correlation has been observed by Bradley in Burmese (1982), Henderson in Cambodian (1952), and Huffman (1976).

Thurgood (2007, pp. 277–278) mentions two factors that may account for these correlations, namely the vocal fold tension and larynx lowering. Breathy voice is associated with laxer tension of the vocal folds and thus with lower pitch while tense or creaky voice is associated with more constricted vocal folds and thus with higher pitch. Breathy voice is also associated with a lowered larynx. Lowering the larynx will result in a longer vocal tract and a longer vocal tract will make the formant frequencies go down even further. In contrast, tense or creaky voice is associated with a raised larynx. Raising the larynx makes the vocal tract shorter and a shorter vocal tract will make the frequency higher.

4 Methods

4.1 Introduction

There are three main related questions addressed in the present study (as laid out in Section 1.4 earlier). The first question is whether there is evidence of shared phonetic qualities between voiced and voiceless aspirated stops to the exclusion of voiceless unaspirated stops in Madurese. Previous studies have suggested that they might share some phonetic properties, but have considered a limited number of features and speakers. The second question is what the implications of the findings are for the proposals of [LL] and [ATR] features. A transparent phonetics-phonology mapping predicts that segments sharing the feature [+LL] would share some phonetic correlates such as lower F0, lower F1, and breathy voice quality that can be measured by H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* and CPP. It also predicts that segments sharing the feature [+ATR] would share some acoustic correlates including lower F1, lower F2 particularly for front vowels and breathy voice quality. The third question is what the results mean for our understanding of the phonetics-phonology interface. If these segments do not share phonetic correlates that would suggest they have a common articulation, we still want to explain how they came to pattern together phonologically.

In order to answer the questions, we carried out acoustic investigation into the CV syllables of Madurese disyllabic words focusing on the word-initial CV syllable. The acoustic parameters we looked at include voice onset time (VOT), closure duration, fundamental frequency (F0) as well as frequencies of the first (F1) and the second (F2) formants. More importantly, we also examined a number of voice quality correlates in order to find out whether voiced and voiceless aspirated stops share these acoustic properties to the exclusion of voiceless unaspirated stops.

The voice quality measures we considered in this study are H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* and Cepstral Peak Prominence (CPP). There are two reasons why we use these different measures. The first reason is related to the fact that a number of studies which look at voice quality found that not all spectral measures can successfully distinguish different voice qualities. The results may depend on, for example, languages, speakers and gender (e.g. Blankenship, 2002; Hillenbrand *et al.*, 1994; Esposito, 2010a; Esposito & Khan, 2012; Garellek & Keating, 2011; Khan, 2012). The second reason is related to the fact that this is the first study of Madurese which looks at voice quality. We believe that using different spectral measures such as these can help us determine which acoustic measures are more successful in distinguishing different categories. This subsequently will also guide us in deciding which measures can potentially be used and possibly modified for future perceptual experiments.

H1*-H2* and H2*-H4* are low-range spectral tilt measures; H1*-H2* in particular is a correlate of the open quotient, which is the percentage of the glottal cycle during which the glottis is open. H1*-A1*, H1*-A2* and H1*-A3* are mid-range measures of spectral tilt. These spectral tilt measures which show how strong or weak higher frequencies are in the spectrum have been associated with the vocal fold closing velocity, the appearance of a posterior glottal opening and the simultaneous closure of the vocal fold ligament (Hanson *et al*., 2001; Stevens, 1977). CPP is a measure of periodicity of the source spectrum. In this case, breathy phonation is expected to have higher H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* values and lower CPP values than modal phonation.

4.2 Research Participants and Location

Twenty-five native speakers of Madurese originating from across four regencies in Madura (Bangkalan, Sampang, Pamekasan and Sumenep) were recorded for the study. They included 13 males and 12 females aged between 18 and 28 and all of them were undergraduate students at Trunojoyo University in Madura. None of the participants reported to have a history of hearing and speech disorders when the recordings were made. It is important to note that although they came from different areas in Madura their speech was not noticeably different in terms of dialectical variations. This was probably because they were all relatively well-educated.

All participants were also speakers of Indonesian but particularly used the language in formal settings such as in schools and in other activities in which speakers of different local languages got involved. In addition, they also learned and spoke some English at school and university. While it is true that the participants were bilingual in Indonesian, which is a typical language situation in Indonesia, all of them grew up in dominantly Madurese-speaking households and mostly used Madurese in their daily lives. They were paid for their effort and participation in the study.

Fifteen speakers' recordings were selected for further acoustic segmentation and analysis. The recordings came from eight females (mean age 20, range 18-21) and seven males (mean age 22, range 20-28) and they were considered as the 'best' speakers for a number of reasons. First, they relatively made fewer pronunciation mistakes and did not show nervousness or hesitation in their speech during the recordings. Second, they read the stimuli relatively naturally as well as with normal speech rate. Third, the quality of their recordings was overall better than the quality of the recordings of the participants who were not selected for further acoustic analysis processes. Fourth, although the participants came from different dialect areas, we observed that their speech did not show noticeable dialect variation. This may be due to the fact that they were all well educated (see Appendix 9.3 for more information about the speakers).

4.3 Speech Material

The study uses 188 Madurese words as stimuli (see Appendix 9.1). The selection of words was done in such a way that voicing type, place of articulation and vowel type had comparable and adequate representations in the data. The word selection also took account of the stops following the vowels in the first syllables in order to make sure that they also contained stops with comparable and representative place and voicing categories. This was particularly done given the fact that vowels may also be affected by either the preceding or the following consonants.

However, it is important to note that in this study we only looked at four places of articulation: bilabial, dental/alveolar, palatal and velar. We excluded the three series of retroflex stops /ʈ, ʈʰ, ɖ/ because we were not able to find a representative sample of Madurese words beginning with this place of articulation. The voiced retroflex stop

/ɖ/ in particular is very difficult to find and as far as we are concerned we were not able to find Madurese words with /ɖ/ in either word-initial or word-medial position.

All words are disyllabic with the syllable patterns of CVCV and CVCVC except one word *dupolo* 'twenty', which has three syllables, due to the difficulty to find more words with similar place and vowel categories. It is important to note that the second syllables of the words are of two different types: open and closed syllables. Although it is true that the difference in syllable type may affect vowel duration, since only the first syllable was measured, we believe that this does not affect the consistency of the measurement results. Moreover, our main concern here focuses on investigating acoustic realisations of the three-way laryngeal contrast in Madurese stops represented particularly in the initial CV syllables. We are not comparing acoustic characteristics of stops and vowels between the first and the second syllables.

4.4 Procedure

4.4.1 Recording and data processing

Recordings were conducted in a quiet room using a Marantz PMD661 portable audio recorder with a Shure SM10A head-mounted microphone and made in mono at a sampling rate of 44,100 Hz with 16-bit resolution. The stimuli consisting of 188 disyllabic Madurese words written in Madurese orthography were embedded in the same sentence frame *Ngèrèng maos* ___ *sè saè* [ŋɛrɛŋ maɔs ___ sɛ saɛ] 'Let's read ___ well'. The use of the same sentence frame was aimed to make sure that all the stimuli were uttered in a relatively consistent manner in terms of stress, intonation and duration. The stimuli were presented in orthographic form using a web-based presentation script that was set up to randomise them in three reading blocks. Participants were instructed to read them as fluently and naturally as possible in three random repetitions. A special instruction was also given in order for participants to maintain the consistency of their intonation and particularly not to pause between words. However, when this happened, participants were instructed to repeat the sentence until an acceptable fluency of the reading was finally achieved.

The recordings were divided into three sessions, each of which lasted for approximately 20 minutes. However, some speakers had to spend more than one hour to finish the recordings because they either mispronounced the words or experienced hesitations and nervousness, which sometimes resulted in dysfluencies in their speech. There were also some cases in which participants read the stimuli very slowly, not because of their natural tendencies but possibly due to their reluctance to make mistakes. When encountering a problem of this kind, we stopped the recording and suggested the participant to take a deep breath and try reading the stimuli again as comfortably as possible.

In order to make sure that participants did not suffer tiredness when reading the stimuli, which could consequently affect the consistency of the recording quality, breaks were provided after the first and the second sessions. During the breaks, they were given something to drink to help clear their throats if necessary. Bottled water was available within the participants' reach during the recording so that they could drink any time they wanted to. They could also go out of the recording room to have fresh air outside. The next session would continue when they felt ready to do so.

4.4.2 Token exclusion

Praat (Boersma & Weenink, 2014) was used to segment and annotate the recordings and all interval boundaries were moved to the nearest zero-crossing automatically before processing. In total, we expected to have 8,460 separate sound files (15 speakers x 188 words x 3 repetitions), but due to poor recording quality, unnoticed mispronunciation or dysfluencies, 110 items were excluded, leaving us with 8,350 files instead. Such cases became obvious when the recordings were played again and the waveforms were examined in detail during the segmentation process.

The number of data points for F0, F1, F2, H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP measured at eleven timepoints from the vowel might vary from timepoint to timepoint. The acoustic data were extracted using the script *spectralTiltMaster* (Mills, 2010). It is important to note that here we used a modified version of the script as the Mills' (2010) version does not measure CPP and H2-H4. In total, we expected 91,850 spectral measurements (8,350 sound files x 11 timepoints), but we were finally able to obtain 87,597 spectral measurements. Thus, we lost as many as 4,253 data points and this is primarily because Praat could not determine an F0 candidate even after we manually changed the F0 parameter, as will be explained in Section 4.5.2.1 later. Most of the losses in spectral measurements came from the vowel [ə], which is probably due to the fact that this vowel has a very short duration. However, as this data loss only makes up of 4.9% of the whole data points, we are confident that it does not affect the reliability of the statistical analysis results later. This is particularly reasonable given that statistical data analyses in the study were carried out using linear mixed-effects models, which are known to deal well with missing data (Baayen, Davidson, & Bates, 2008).

4.5 Acoustic Measurements

4.5.1 Measurement criteria and segmentation labelling

In general, the placement of segment boundaries was decided primarily on the basis of visual inspection on spectral characteristics which can be easily observed in the wideband spectrogram, calculated using Fast Fourier Transforms (FFT) on a 5 ms Gaussian window. In addition, we also looked at waveforms which are also useful in determining segment boundaries because they usually exhibit ‘dips and rises in amplitudes’, corresponding to onsets and releases of constrictions (Turk, Nakai, & Sugahara, 2006). As Turk *et al.* (2006) point out, looking at waveforms will help us obtain a segmentation decision which is more fine-grained.

We follow Turk *et al.* (2006) who advocate oral constriction criteria in determining durations of acoustic segments because oral constriction can be used for different classes of speech sounds. However, when determining vowel duration, we did not include the aspiration portion following the stop release as belonging to the vowel interval. This is particularly relevant when later dealing with voiceless aspirated and voiceless unaspirated stops which show a certain degree of aspiration following the stop release, as shown in Figure 4–3 and Figure 4–4 on pages 87 and 88 respectively.

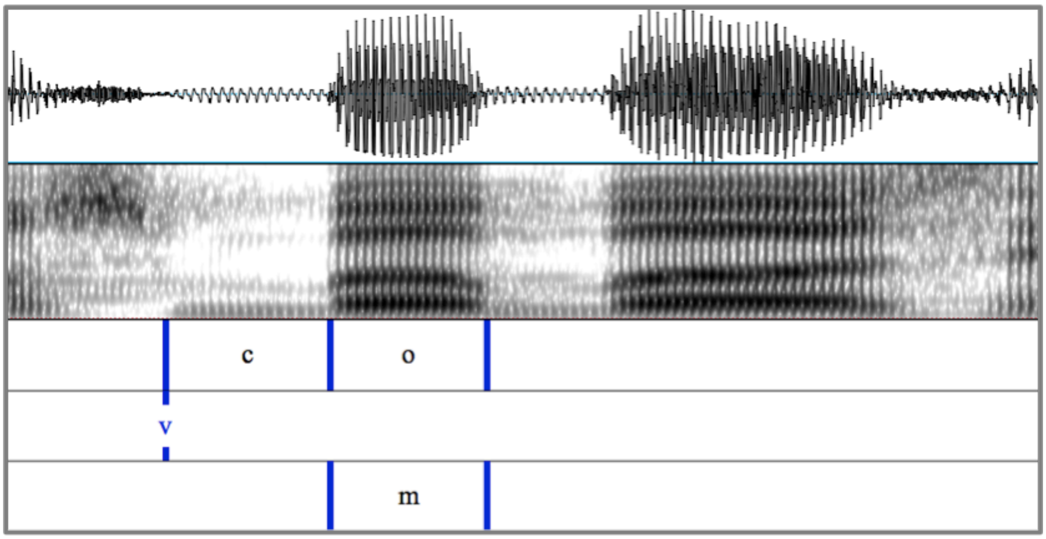


Figure 4–1. An example of fully voiced stops in word-initial position for the word *bâbâ* [bɤbɤ] 'under' produced by KA (male). ‘c’, ‘o’, ‘v’ and ‘m’ stand for oral closure, open phase, start of voicing and vowel respectively.

In these segmentations, ‘c’ stands for closure duration; ‘v’ stands for the beginning of stop voicing that usually occurs in voiced stops, where the end of stop voicing coincides with the vowel onset. ‘v’ is also used to mark the onset of voicing for vowels which occur after voiceless aspirated and voiceless unaspirated stops. Similarly, ‘o’ stands for two purposes. In the case of voiced stops, it is used to mark a time domain starting from the end of stop voicing up to the vowel offset, which also corresponds to vowel duration. In the case of voiceless unaspirated and voiceless aspirated stops, it is used to mark a period beginning from stop release up to the vowel offset. In these cases, this also includes the aspiration portion, as shown in Figure 4–3. As ‘o’ for voiced stops is different from that for voiceless unaspirated and aspirated stops, we used ‘m’ to mark vowel duration following any stop type.

The start for the stop closure is indicated by the offset of high frequency noise preceding the stops. And the end of the stop closure for voiced stops is indicated with reference to the start of the burst while that for voiceless aspirated and voiceless unaspirated stops is indicated by the burst release including the aspiration portion. As shown in Figure 4–1 above and Figure 4–2 on the following page, voiced stops appear to be either fully voiced or partially voiced respectively. Closure duration for voiced stops with partial voicing also includes the voiceless portion up to the vowel onset while closure duration for voiceless unaspirated and voiceless aspirated stops includes a voiceless time domain up to a burst marking the stop release (see Figure 4–3 and Figure 4–4).

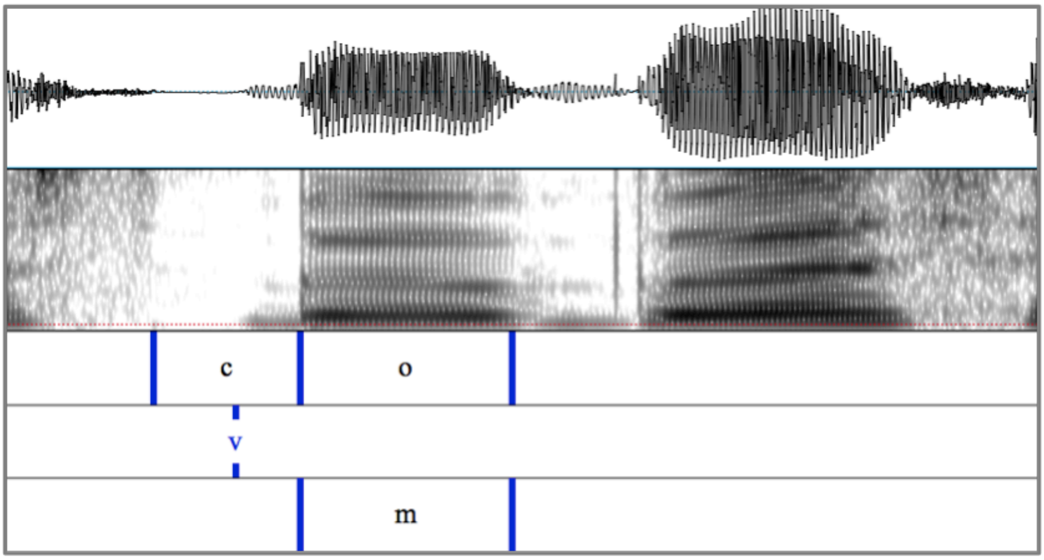


Figure 4–2. An example of partially voiced stops in word-initial position for the word *bâbâ* [bɤbɤ] 'under' produced by LH (female). 'c', 'o', 'v' and 'm' stand for oral closure, open phase, start of voicing and vowel respectively.

It is important to note that the occurrence of partial voicing such as the one shown in Figure 4–2 is very rare, making up of less than five percent of the whole voiced tokens. This is also true for voiced tokens which start with voicing and are extinguished before the closure release. For this reason, we did not treat them as a separate category from the other prevoiced tokens.

The negative VOT is determined by looking at the point in which the voice bar appears up to the point where the vowel begins. In this case, the start of the stop voicing marked with 'v' and the onset of the striations in the second formant of the vowel is the duration of the negative VOT, as shown in Figure 4–1 and Figure 4–2. In this case, the negative VOT also corresponds to voicing during closure.

Figure 4–3 and Figure 4–4 demonstrate how positive VOT values are measured for voiceless aspirated and voiceless unaspirated stops respectively. As shown in these figures, the duration of the positive VOT is measured from the point in which the burst following oral closure appears up to the start of the voicing for the following vowel indicated by the onset of the striations in the second formant of the vowel labelled with 'v'.

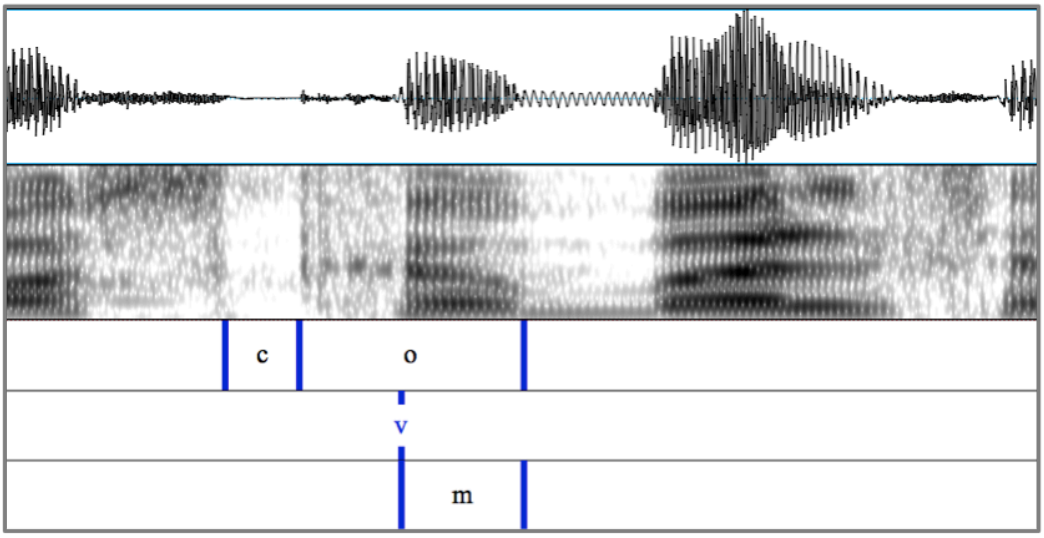


Figure 4–3. An example of voiceless aspirated stops in word-initial position for the word *ghâbâl* [kʰɤbɤl] ‘shocked’ produced by FZ (male). ‘c’, ‘o’, ‘v’ and ‘m’ stand for oral closure, open phase, start of voicing and vowel respectively.

In terms of vowel duration, the onset of the striations in the second formant of the vowel is used for defining the point in which the vowel starts while the offset of the striations in the second formant of the vowel is used for defining the point where the vowel ends. Thus, the onset and offset of the striations in the second formant determine vowel duration. As seen in Figure 4–1 and Figure 4–3, for example, voiced and voiceless stops differ with regard to how vowels are measured. This is because the open phase (‘o’) corresponds to the vowel itself for voiced stops while for voiceless stops it also includes the VOT of the preceding stop. In order to make the measurement uniform and the acoustic parameter extraction easier to calculate, we added a third tier in the *TextGrid* labelled with ‘m’ to denote the voiced portion of the vowel. It is important to note that the tier with the label ‘m’ used to extract vowel duration was also the tier from which acoustic measurements for F0, F1, F2, and a number of spectral measures including H1*, H2*, H4*, A1*, A2* and A3* were measured and extracted, as also illustrated in Figure 4–1 and Figure 4–3.

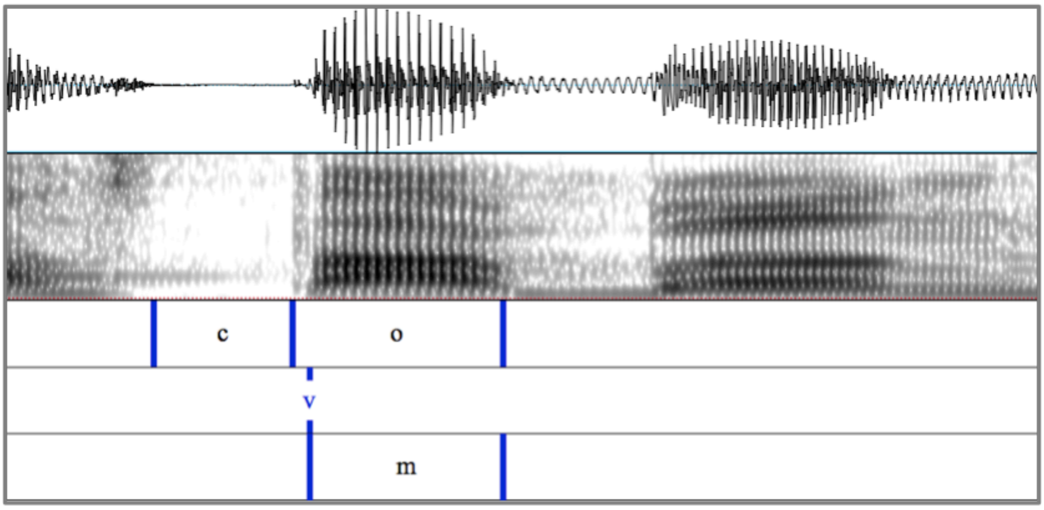


Figure 4–4. An example of voiceless unaspirated stops in word-initial position for the word *tabâng* [tabɤŋ] 'to chase' produced by HF (male). 'c', 'o', 'v' and 'm' stand for oral closure, open phase, start of voicing and vowel respectively.

4.5.2 Spectral parameters

F0 was measured applying the autocorrelation method of Boersma (1993) with a 15 ms frame duration and a 500 Hz pitch ceiling. Formants were measured by computing LPC coefficients using the implementation of the Burg algorithm in *Praat*, using a 25 ms window with pre-emphasis applied from 50 Hz, and then smoothed using the *Track...* function. Harmonic structure was determined through spectral analysis using FFT and long-term average spectra applied to 25 ms windows centred at the measurement points. The amplitudes of the first (H1), the second (H2) and the fourth (H4) harmonics were measured along with the amplitudes of the most prominent harmonics of the first (A1), the second (A2) and the third (A3) formants in order to calculate H1-A1, H1-A2, H1-A3, H1−H2 and H2-H4. These measures were subsequently corrected for the effect of the first two formants on the vocal tract transfer function (see section 4.5.2.3 below).

4.5.2.1 Measurement of fundamental frequency (F0)

F0 was measured at eleven equidistant timepoints throughout the vowel and extracted using a suite of scripts *spectralTiltMaster* created by Timothy Mills (2010). The script also provides error checking for F0 measurement; however, due to the large number of the data points we measured, we did not check any error during the process of data extraction. However, we instead carried out error checks by plotting the extraction results of F0 values using the *ggplot2* package (Wickham, 2009) in R (R Core Team, 2015). We identified errors by looking at the plot of F0 values for each speaker. In this case, if the plot showed F0 values which were not within the F0 range of the speaker, that is, the values which were either too high or too low, we identified them as errors.

When such errors were spotted, corrections were manually made on the basis of the individual measurement. Such errors primarily occurred when the algorithm used to measure F0 could not distinguish between F0 and F1 particularly when the vowels have low F1 values corresponding to high vowels. Errors were also found to occur when the vowel duration was very short resulting in an unclear formant structure. The corrections, which were all done in *Praat*, involved making adjustments to F0 range by taking gender and vowel quality into account.

4.5.2.2 Measurement of vowel quality (F1 and F2)

The measurement and extraction of F1 and F2 were also conducted using the script *spectralTiltMaster*. A number of corrections for both formant values were also done particularly when there was reason to suspect that some measurement errors had occurred. This error finding was carried out by plotting both F1 and F2 of each vowel using the *ggplot2* package (Wickham, 2009) in R (R Core Team, 2015). In this case, our knowledge on possible formant values for each vowel can help guide us in deciding whether the measurement of both formants makes sense or contains errors.

4.5.2.3 Measurement of voice quality

The measurement and extraction of H1-A1, H1-A2, H1-A3, H1−H2, H2-H4 and Cepstral Peak Prominence (CPP) were also carried out using the same *Praat* script mentioned earlier. Harmonics are known to be mostly affected by formant frequencies. For example, the first and the second harmonics, which are close to F1, are mostly amplified especially in high vowels. In order to obtain an accurate measurement for voice quality whose acoustic measures involve the measurement of harmonics, correction is necessary. This is particularly the case since the present study deals with different vowel qualities.

The asterisks shown in H1*-A1*, H1*-A2*, H1*-A3*, H1*−H2* and H2*-H4* respectively indicate that H1, H2, H4, A1, A2 and A3 were all corrected for the effects of both bandwidths and formant frequencies on the harmonics. It is important to note that Hanson's (1997) correction formula only works particularly well for non-high vowels. This is because she only corrects for the effect of the first formant on the first two harmonics (H1 and H2) but she does not take into account of the first formant bandwidth. For that reason, we used an improved correction formula suggested by Iseli and Alwan (2004) and Iseli *et al.* (2007) that corrects for all vocal tract resonances including their bandwidths (Iseli & Alwan, 2004; Iseli *et al.*, 2007). Their formula has been shown to significantly remove the effects of formant frequencies on harmonics particularly for high vowels and it can therefore be equally applied to both high and non-high vowels alike (Iseli & Alwan, 2004; Iseli *et al.*, 2007). Specifically, the script corrects for the values and bandwidths of the first two formants for all measures, and for A3, for the value/bandwidth of F3 as well, as recommended by Iseli and Alwan 2004. This subsequently facilitates comparison across different vowel types. Such corrections are relevant and crucial given the well-known covariation of stop voicing and vowel height in Madurese.

Specifically, in order to determine spectral magnitudes for H1, H2, H4, A1, A2 and A3, vowel segments, which were originally digitised in 44 kHz, were downsampled to 16 kHz. Eleven equally spaced timepoints were identified in the vowel. At each timepoint *t*, a spectrum was computed from 25 ms window centred at *t* and converted to a long-term average spectrum (LTAS). The harmonic amplitude was then determined from this spectrum by finding the maximum value in the frequency range f +/- (f/10), where f is the frequency of interest (H1, H2, H4, A1, A2 and A3).

Spectral magnitudes were then corrected in *Praat* using the method of Iseli *et al.* (2007), based on the implementation in Shue *et al.* (2011). H1, H2, H4, A1 and A2 were each corrected for the formant frequencies and bandwidths of F1 and F2, while A3 was corrected for the frequencies and bandwidths of F1, F2 and F3. Formant bandwidths were determined as the frequency of the point 3 dB below the formant peaks (i.e. the half-power point on each side of the peak).

Cepstral peak prominence (CPP) was determined using the algorithm of Hillenbrand *et al.* (1994). First, a *PowerCepstrum* object was created based on the spectrum of the audio slice. The cepstral peak was then found in the 50-500 Hz range using parabolic interpolation. To determine harmonics-to-noise (HNR) ratios, the vowel waveform was first filtered into three bands (0-500 Hz, 0-1500 Hz, and 0-2500 Hz) using a Hanning window with a smoothing frequency of 100 Hz. Short-term HNR analysis was then performed on each filtered sound, implemented in *Praat* as the *To Harmonic (cc) ...* command.

4.6 Statistical Analysis

4.6.1 General modelling

We used the *lme4* package (Bates *et al.*, 2014) run in R (R Core Team, 2015) to carry out linear mixed-effects analyses for VOT, closure duration, F0, F1, F2 and a number of voice quality measures, i.e. H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. All the models used in the study were obtained by building and comparing a sequence of nested models whereby simpler models were compared with more complex ones. Log-likelihood ratio tests were subsequently conducted using the *anova()* function in R to find out whether adding complexity to a certain model would improve its goodness of fit and was therefore justified by the data. The most complex model which converged and was justified by the data was finally selected. Model comparison of each model for each acoustic measure is addressed in more detail later in Chapter 5.

4.6.1.1 Fixed effects

Different fixed effects were included in each model to capture factors that possibly affect acoustic variables of interest in Madurese. Specifically, the VOT model includes Voicing with three levels (voiced, voiceless unaspirated and voiceless aspirated), Place with four levels (bilabial, alveolar, palatal and velar) and the interaction term for Voicing and Place as the fixed effects. The model for closure duration only includes Voicing as the only fixed effect. The models for F1 and F2 include Vowel with eight levels (a, ɛ, ə, ɔ, ɤ, i, ɨ, u) and Place as the fixed effects while the models for F0, H1*-A3*, H1*-H2*, H2*-H4* and CPP include Voicing, Gender with two levels (female and male) and the interaction term for Voicing and Gender as the fixed effects. In contrast, the models for H1*-A1* and H1*-A2* only include Voicing and Gender without their interaction term as the fixed effects.

It is important to note that we did not include Vowel as a fixed effect in all the models except for the models for F1 and F2. This is because the inclusion of Vowel creates a rank deficiency problem in computation. This problem results from insufficient or lack of information contained in the data to estimate the model due to the consonant-vowel covariation in Madurese stops. This is due to the fact that not all the voicing categories can be followed by vowels of the same height. For example, the VOT for voiceless aspirated and voiceless unaspirated stops cannot be compared as a function of vowel height because high vowels only occur after voiceless aspirated stops and voiced stops while non-high vowels only occur after voiceless unaspirated stops. In this case, we can only include Vowel as a fixed effect if we compare voiced and voiceless aspirated stops as a function of vowel height since both are followed by vowels of the same height.

In all cases, a fixed effect was considered significant at $\alpha = 0.05$. Since the *lme4* package does not provide *p*-values for either *t*- or *F*-tests due to the uncertainty and complexity with the calculation of the degrees of freedom, the *p*-values in this study were obtained using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2016) implemented in R. *P*-values in the *lmerTest* package are calculated from *F* statistics of types I - III hypotheses with two options for denominator degrees of freedom of *F* statistics, namely ‘Satterthwaite’ and ‘Kenward-Roger’. Furthermore, in order to compare differences in acoustic variables of interest, a series of post-hoc pairwise comparisons were conducted using the *lsmeans* package (Lenth, 2014) in R, which also provides *p*-values for the results of the associated tests. This package uses the *pbkrtest* package which implements the Kenward and Roger method for the degrees of freedom of *t* statistics to obtain *p*-values.

4.6.1.2 Random effects

All the models used in the study included crossed random effects for Speaker and for Word. By-speaker and by-word random intercepts were included to capture variability in relevant acoustic variables in terms of speakers and words. However, they may differ in terms of their specification for random slopes. Specifically, the VOT model includes by-speaker random slopes for Voicing and Place in order to take into account the variability in speakers' VOT productions relative to stop types and place of articulation. In contrast, the models for F1 and F2 include by-speaker random slopes for Vowel and Place to also consider the variability in speakers’ F1 and F2 relative to Vowel and Place. The models for closure duration, F0, H1*-H2*, H2*-H4* and CPP only include by-speaker random slopes for Voicing to take account of speakers’ variability in these acoustic measures relative to Voicing.

5 Results

5.1 Introduction

This chapter reports the results of statistical analyses of some acoustic measurements and consists of the following sections. Section 5.2 reports the results of VOT analyses with regard to the realisations of VOT as a function of voicing and place categories. Section 5.3 reports the results of stop closure duration analyses as a function of voicing categories. Section 5.4 presents the results of F0 analyses by focusing on the realisations of F0 at vowel onset and midpoint following voiced, voiceless unaspirated and voiceless aspirated stops taking voicing and gender into account. Sections 5.5, 5.6, 5.7, 5.8, 5.9 and 5.10 report a number of acoustic correlates of voice quality: H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4*, all of which are measures of spectral tilt, and CPP, which is a measure of periodicity of the spectrum. These analyses focus on the realisations of each of the acoustic measures at vowel onset and midpoint following each stop category, also taking voicing and gender into consideration. Section 5.11 provides the results for linear discriminant analyses, comparing a model which only uses spectral measures and a model which uses both spectral measures and VOT as predictors in assessing the three voicing categories in Madurese. Section 5.12 looks at the acoustic realisations of the eight surface vowels of Madurese by analysing their F1 and F2 values at vowel onset and midpoint. Implications of each of the acoustic findings are also given. Section 5.13 summarise the results of the analyses by identifying which acoustic properties distinguish one voicing category from another and which properties are shared by each of the voicing categories.

5.2 Voice Onset Time (VOT)

5.2.1 Descriptive statistics on VOT

Figure 5–1 on the following page shows the VOT distribution for stops in Madurese grouped by voicing type and gender. As the figure shows, there is a clear separation between voiced stops on the one hand and voiceless unaspirated (voiceless) and voiceless aspirated (aspirated) stops on the other. In contrast, the VOT values for voiceless unaspirated and voiceless aspirated stops do not look well separated from one another. If we take a closer look at Figure 5–1, there also appears to be no gender distinction in the VOT values of Madurese stops. This may indicate that female and male speakers produce voiced, voiceless unaspirated and voiceless aspirated stops with more or less similar VOT duration.

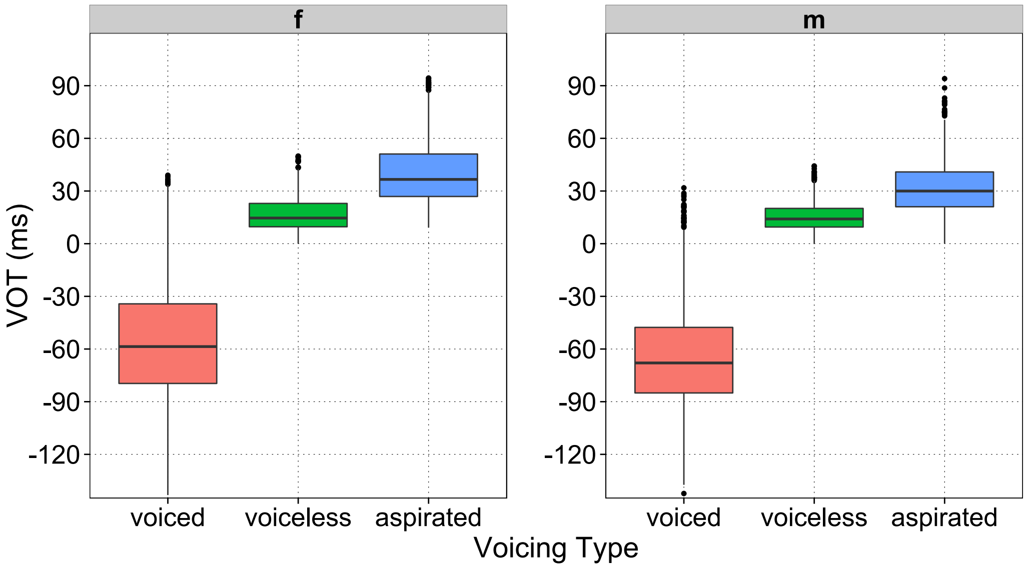


Figure 5–1. Boxplots of VOT by voicing type and gender. Females (f) are shown on the left panel and males (m) are on the right panel. The horizontal lines indicate median values and the vertical lines (whiskers) indicate the lower and the upper quartiles respectively.

Table 9 below shows the mean VOT values (ms) along with the standard deviations (in parentheses) of Madurese stops by voicing type and gender over all places of articulation averaged across speakers and repetitions. As shown in Table 9, the magnitudes in the variability in VOT values for voiced stops can also be seen in their having relatively higher standard deviation. This is also true for voiceless unaspirated and voiceless aspirated stops despite being with relatively lesser magnitudes.

Table 9. Mean VOTs (ms) and standard deviations (in parentheses) by voicing type and gender over places of articulation

Gender	Voicing Type		
	Voiced	Voiceless	Aspirated
Female	-54 (36)	17 (9)	40 (17)
Male	-65 (29)	15 (8)	32 (14)

Figure 5–2 below shows the VOT distribution of Madurese stops by voicing type and place of articulation. As we can see, there is possibly an effect of place of articulation

in the VOT of Madurese stops and the effect appears to conform to the general tendency for VOT values as expected by place of articulation for voiceless stops with positive VOT values (Abdelli-Beruh, 2009; Cho & Ladefoged, 1999). In the case of voiced stops with negative VOT values, it appears that the VOT for bilabial stops is shorter than that for alveolar stops and the VOT for palatal stops is longer than those for bilabial and alveolar stops. However, it is unexpected that the VOT for voiced velar stops looks shorter than those for voiced alveolars and for voiced palatals and it also looks very similar in duration to the VOT for voiced bilabials.

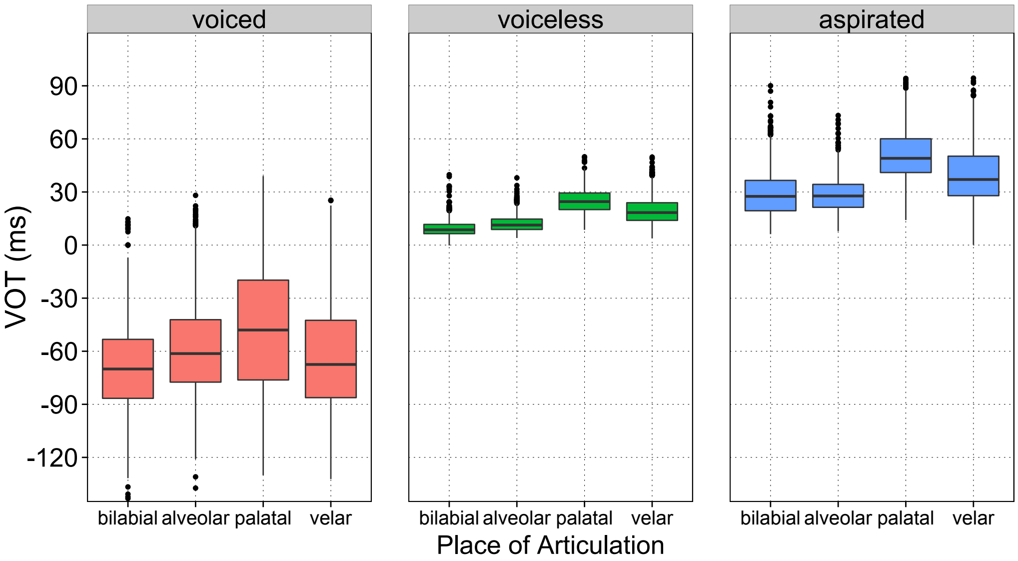


Figure 5–2. Boxplots of VOT by place of articulation and voicing type. The horizontal lines indicate median values and the vertical lines (whiskers) indicate the lower and the upper quartiles respectively.

Moreover, if we look at the VOT distribution for voiceless unaspirated stops by place of articulation in Figure 5–2 above, the stops produced at the front part of the mouth tend to exhibit shorter VOT values than those produced at the back part of the mouth, setting aside the palatal stops, which have the longest VOT probably due to affrication, as illustrated in Figure 5–3 on the following page. Specifically, the VOT for bilabial stops appears slightly shorter than that for alveolar stops while the VOT for velar stops is longer than those for bilabial and alveolar stops. Thus, we can see a relatively clear pattern in the VOT values for voiceless unaspirated stops according to place of articulation.

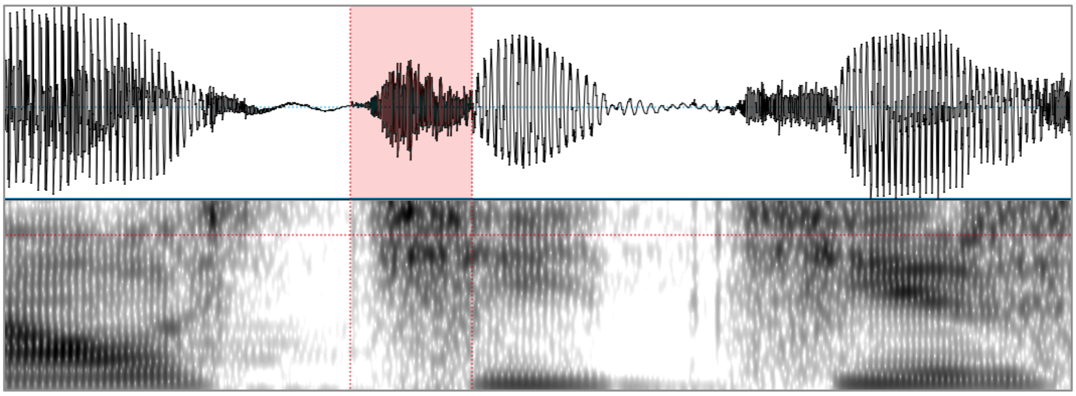


Figure 5–3. An example of an affricated voiceless aspirated palatal stop (shaded) in word-initial position for the word *jhijhir* [c^{h}ic^{h}ir] 'stand in row' produced by AF (female).

Furthermore, if we take a look at the VOT distribution for voiceless aspirated stops by place of articulation, there appears to be a different pattern from that for voiceless unaspirated stops. This is particularly due to the fact the VOT values for bilabial and alveolar stops look quite similar, as shown in Figure 5–2 earlier. However, putting aside the VOT for palatal stops, which also have the longest VOT value possibly due to affrication, the VOT for velar stops is longer than the VOTs for bilabial and alveolar stops, which are consistent with cross-linguistic and articulatory expectations of VOT values according to place of articulation.

Table 10. Mean VOTs (ms) and standard deviations (in parentheses) by voicing categories and place of articulation

Place of Articulation	Voicing Type		
	Voiced	Voiceless	Aspirated
Bilabial	-69 (27)	10 (5)	30 (13)
Alveolar	-57 (30)	12 (5)	29 (11)
Palatal	-44 (41)	25 (7)	51 (15)
Velar	-64 (32)	20 (8)	40 (16)

A further observation on Figure 5–1 shown earlier will reveal that the VOT values for voiced, voiceless unaspirated and voiceless aspirated stops in Madurese also show large ranges of variability. Specifically, the VOT for voiced stops looks relatively more variable than that for voiceless aspirated stops while the VOT for voiceless unaspirated stops appears to be the least variable. A similar observation regarding the variability in VOT values also holds if we take a look at Figure 5–2. In this regard, the ranges of variability in VOT values also take both voicing and place categories into account. This is the case if we look at the VOT values for voiced palatal and voiced velar stops in particular, which look relatively more variable than the other voiced stop categories.

Table 10 on the preceding page shows the mean VOT values (ms) along with standard deviations (in parentheses) of Madurese stops by voicing and place of articulation categories averaged across 15 speakers (7 females and 8 males) and repetitions. As shown in the table, the magnitudes in the variability in VOT values for voiced palatal and velar stops can also be clearly seen in terms of their relatively higher standard deviations. This is also true for voiceless aspirated and voiceless unaspirated palatal and velar stops although with lesser magnitudes.

The ranges of variability in VOT values will be much clearer if we look at the distribution of the VOT of each voicing category. Figure 5–4 on the following page shows that there are a number of cases in which stops that are phonologically voiced are produced with very short prevoicing or even in some cases with no clear prevoicing at all. The same also applies to voiceless unaspirated stops that are in some cases produced with relatively long lag VOT and to voiceless aspirated stops that are conversely produced with short lag VOT instead. Figure 5–4 also displays the frequency of the VOT distributions of the three voicing categories. It shows that the VOT values for voiceless unaspirated and voiceless aspirated stops in particular overlap quite extensively.

This state of affair is unexpected if Madurese is to be considered as a language with a three-way laryngeal contrast given that its VOT distributions appear to be bimodal, which is similar to the VOT distributions for voiced and voiceless stops instead. Thus, the VOT distributions in Madurese are different from other languages that have been well-known to have a three-way laryngeal contrast such as Thai and Eastern Armenian. Unlike Madurese, these languages show clear trimodal VOT distributions that correspond to voiced, voiceless unaspirated and voiceless aspirated stops (Lisker & Abramson, 1964).

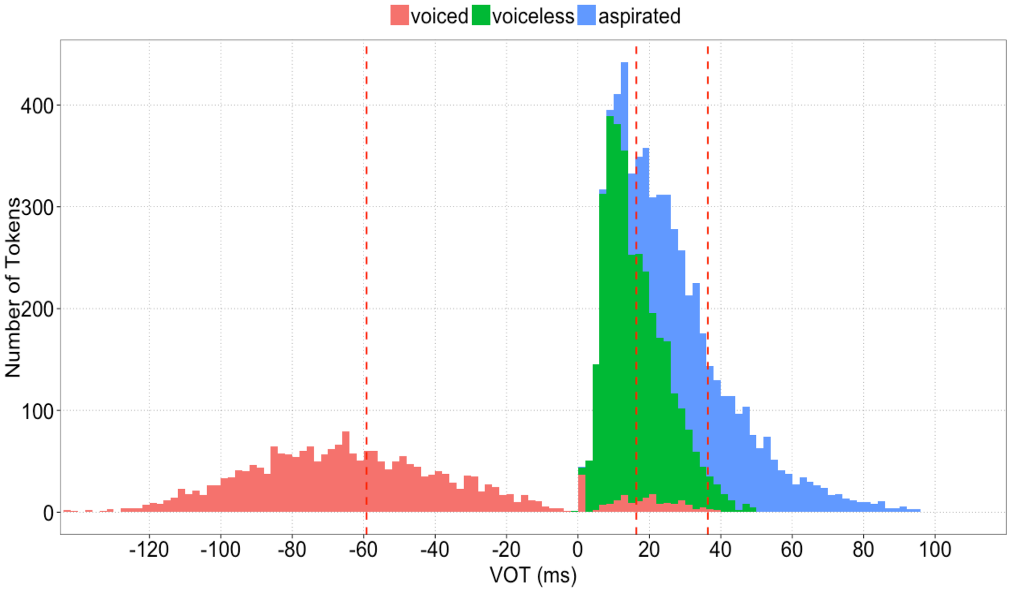


Figure 5–4. Frequency of VOT distribution by voicing categories. Red vertical dashed lines indicate mean values.

A further examination of the voiced stops with positive VOT values that comprise 157 of the total 2306 phonologically voiced tokens yields the following distributions. Females make up the majority of speakers who produce voiced stops with positive VOT values. Specifically, females produced 118 voiced tokens with no prevoicing while males produced 39 of the cases. In terms of place of articulation, palatal stops account for 90 tokens, alveolars 40, velars 17, and finally bilabials 10 tokens.

Moreover, if we look at individual variation in the realisation of VOT for the three-voicing categories, we can also find that there are some cases where female and male speakers show similar patterns. This will be clearer if we look at Figure 5–5, which shows the VOT plots for each of the speakers. It appears that most of the variations come from voiceless unaspirated and voiceless aspirated stops.

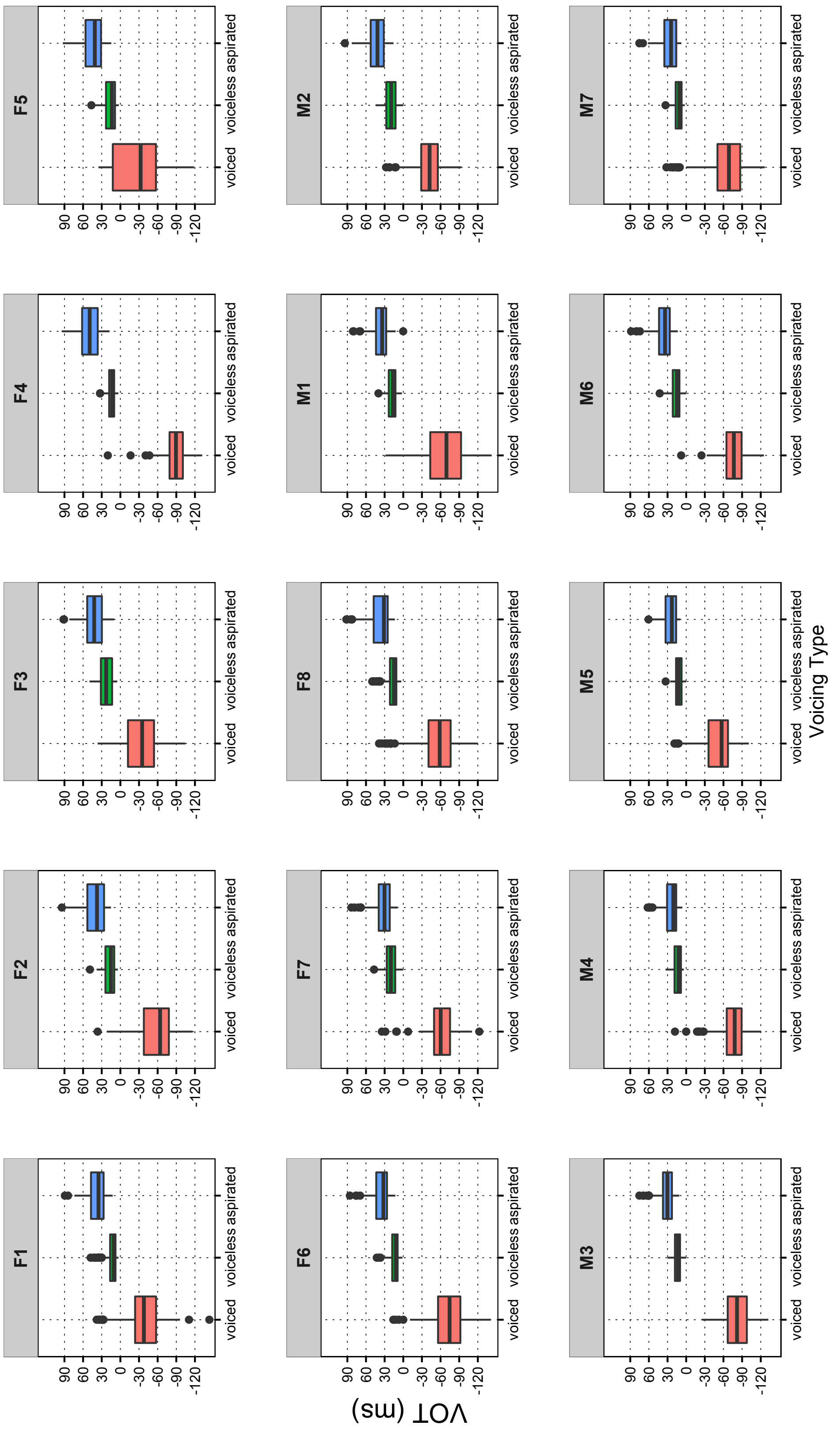


Figure 5–5. Individual variation in the realisation of VOT for voiced, voiceless unaspirated and voiceless aspirated stops. F stands for female speakers and M for male speakers.

As shown in Figure 5–5, there is some individual variation in the realisation of VOT for voiced, voiceless unaspirated and voiceless aspirated stops. The variation is particularly evident when we look at the VOT values for voiceless unaspirated and voiceless aspirated stops and it can be grouped into two main groups. The first group consists of speakers F1, F2, F3, F4, F5, F6, F8, M1, M2, M3 and M6. These speakers appear to have relatively different VOT values for voiceless unaspirated and voiceless aspirated stops. The second group consists of speakers F7, M4, M5, M6 and M7. These speakers appear to have relatively similar values for these two voiceless stop categories. However, despite such variations, all speakers show a consistent prevoicing for voiced stops. That is, the VOT for voiced stops is robustly separated from that for voiceless unaspirated and voiceless aspirated stops.

5.2.2 Model comparison for VOT

Based on the observation above, we were particularly interested in testing the effects of voicing and place of articulation categories on the realisation of VOT in Madurese stops. To this end, we built a number of linear mixed-effects models and carried out model comparison on them in order to obtain the final model for VOT. Log-likelihood ratio tests were subsequently conducted using the *anova()* function in R to find out whether adding complexity to a certain model improved its goodness of fit and was therefore justified by the data. The following five models were considered:

vot1: VOT~Voicing + (1 + Voicing | Speaker) + (1 | Word)
vot2: VOT~Voicing + Place + (1 +Voicing | Speaker) + (1 | Word)
vot3: VOT~Voicing + Place + (1 +Voicing + Place | Speaker) + (1 | Word)
vot4: VOT~Voicing + Place + Gender + (1 + Voicing + Place | Speaker) + (1 | Word)
vot5: VOT~Voicing * Place + (1 + Voicing + Place | Speaker) + (1 | Word)

Using the model vot1 as an example for description purposes, VOT as the dependent variable appears to the left of the tilde operator (~), which means ‘as a function of’. The fixed effect, Voicing, is specified to the right of the tilde. The random effects for Speaker are specified as (1 + Voicing | Speaker). This notation means that we introduce by-speaker adjustments to the intercept (denoted by 1) and by-speaker adjustments to Voicing. The random intercept for Word is specified as (1 + Word), which we can read as a random effect which introduces adjustments to the intercept (denoted by 1) grouped by Word. In other words, the model includes by-speaker and by-word random intercepts and by-speaker random slopes for Voicing.

Table 11. Log-likelihood results for VOT model comparison

	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
vot1	11	69983	70060	-34981			
vot2	14	69834	69932	-34903	155.4835	3	< 2.2e-16 ***
vot3	29	69535	69739	-34739	328.1596	15	< 2.2e-16 ***
vot4	30	69535	69746	-34738	2.1922	1	0.1387
vot5	35	69505	69751	-34718	39.9952	5	1.497e-07 ***

As shown in Table 11, adding a main effect of Place in vot2 provides a significantly better VOT model; that is, vot2 is better compared to vot1, and adding by-speaker random slopes for Place in vot3 also improves the model significantly. In contrast, adding a main effect of Gender in vot4 does not significantly produce a better model. For this reason, Gender was removed and interaction terms for Voicing and Place were added in vot5 instead. This results in a significant improvement in vot5 as shown by the log-likelihood tests above. Therefore, vot5 was chosen as the model for VOT since it is relatively most complex and has no convergence issues. The model includes main effects of Voicing and Place as well as the interaction of Voicing and Place as the fixed effects. It also includes by-speaker and by-word random intercepts as well as by-speaker random slopes for Voicing and Place as the random effects.

It is important to note that because VOT values may vary with vowel types cross-linguistically (Gósy, 2001; Nearey & Rochet, 1994; Port & Rotunno, 1979), we have also attempted to include Vowel as a factor. However, as Vowel by itself cannot be included as a predictor in the model due to the problem of rank deficiency (i.e. there is insufficient information in the data to estimate the model given the consonant-vowel covariation), a new variable VowelPair were instead designed to deal with it. The variable consists of four levels, i.e. pair a~ɤ, pair ɛ~i, pair ə~ɨ and pair ɔ~u. As the log-likelihood ratio test confirms that the inclusion of VowelPair does not yield a better goodness of fit for the VOT model (χ^2 (2) = 2.82, p = 0.24), we removed it from the model. The result does not provide any evidence that vowels contribute to the VOT differences in Madurese. However, it is likely that their effects, if any, have been confounded by the fact that the value for each pair actually derives from the average of high and non-high vowels. Therefore, in actual fact it is difficult to tease apart the effect of individual vowels on the VOT of Madurese stops, if any, due to the CV co-occurrence restriction.

5.2.3 Inferential statistics on VOT as a function of Voicing and Place

Table 12. The output of a linear mixed-effects model for VOT. Voiceless unaspirated is the reference category for Voicing and bilabial is the reference category for Place. *P* values were obtained using the *lmerTest* package.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	9.519	1.654	79.34	5.755	**< .0001**
VoicingVoiced	-78.686	4.799	21.76	16.404	**< .0001**
VoicingAspirated	20.139	2.497	66.74	8.064	**< .0001**
PlaceAlveolar	2.907	2.031	169.85	1.432	0.1541
PlacePalatal	15.765	3.015	45.91	5.229	**< .0001**
PlaceVelar	10.285	2.042	134.82	5.037	**< .0001**
VoicingVoiced:PlaceAlveolar	8.864	2.945	184.04	3.010	**0.0029**
VoicingAspirated:PlaceAlveolar	-3.674	2.761	185.60	-1.331	0.1849
VoicingVoiced:PlacePalatal	9.936	3.074	184.50	3.232	**0.0015**
VoicingAspirated:PlacePalatal	6.213	2.956	184.09	2.102	**0.0369**
VoicingVoiced:PlaceVelar	-4.223	2.924	184.30	-1.444	0.1504
VoicingAspirated:PlaceVelar	0.324	2.766	184.18	0.117	0.9069

Table 12 above shows the results of a linear mixed-effects model for VOT. In this model, voiceless unaspirated is used as the reference level for Voicing and bilabial is the reference level for Place. Voicing is treatment-coded in order to facilitate the comparison between voiceless aspirated and voiceless unaspirated stops. Comparing these two voiceless stops is of particular interest given the fact that, as shown in Figure 5–1 and Figure 5–2, they have a relatively small VOT difference and indeed show some overlap in their VOT values.

The results show that there was a significant difference between the mean VOT values for voiced and voiceless unaspirated bilabial stops ($p < .0001$). The mean VOT value for voiced bilabial stops was estimated to be -69 ms (miliseconds). The difference between the mean VOT values for voiceless aspirated and voiceless unaspirated bilabial stops was also found to be statistically significant ($p < .0001$). The mean VOT value for voiceless aspirated bilabial stops was estimated to be 30 ms or around 20 ms longer than that for their voiceless unaspirated counterparts.

Moreover, there was no significant difference between the mean VOT values for voiceless unaspirated alveolar and unaspirated bilabial stops (*p* = 0.15). However, there was a significant difference between the mean VOT values for voiceless unaspirated palatal and voiceless unaspirated bilabial stops (*p* < .0001). The mean VOT value for voiceless unaspirated palatal stops was estimated to be 25 ms or about 16 ms longer than that for voiceless unaspirated bilabial stops. The difference between the mean VOT values for voiceless unaspirated velar and voiceless unaspirated bilabial stops was also statistically significant (*p* < .0001). The mean VOT value for voiceless unaspirated velar stops was approximately 20 ms or around 10 ms longer than that for voiceless unaspirated bilabial stops.

The results in Table 12 also show that there was a significant difference between the mean VOT values for voiced alveolar and bilabial stops (*p* = 0.003). The mean VOT value for voiced alveolar stops was estimated to be -60 ms. The difference between the mean VOT values for voiced palatal and bilabial stops was also found to be statistically significant (*p* = 0.001). The mean VOT value for voiced palatal stops was estimated to be around -59 ms. However, the mean VOT value for voiced velar stops was not significantly different from that for voiced bilabial stops (*p* = 0.15).

As also shown in Table 12, there was no significant difference between the mean VOT values for voiceless aspirated alveolar and bilabial stops (*p* = 0.18). The mean VOT value for voiceless aspirated alveolar stops was estimated to be 26 ms. The difference between the mean VOT values for voiceless aspirated palatal and bilabial stops was also found to be statistically significant (*p* = 0.04). The mean VOT value for voiceless aspirated palatal stops was estimated to be about 36 ms. However, the difference between the mean VOT values for voiceless aspirated velar and bilabial stops did not reach statistical significance (*p* = 0.91).

In order to find out whether the differences between the VOT values for voiced, voiceless unaspirated and voiceless aspirated stops were significant across places of articulation, a series of post hoc tests with the Tukey method for a family of three means (voiced, voiceless and aspirated) were conducted. As seen in Table 13, the results indicate that the differences between the VOT values for voiced, voiceless unaspirated and voiceless aspirated stops were significant for all place categories.

Table 13. Results of post-hoc within-place pairwise comparisons for VOT by voicing categories. *P* values are adjusted based on the Tukey method for a family of 3 means.

Place	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Bilabial	aspirated – voiced	98.824	5.046	22.78	19.59	< **.0001**
	aspirated – voiceless	20.138	2.564	70.48	7.85	< **.0001**
	voiced – voiceless	-78.686	4.952	23.20	-15.89	< **.0001**
Alveolar	aspirated – voiced	86.286	5.110	23.96	16.89	< **.0001**
	aspirated – voiceless	16.464	2.639	77.28	6.24	< **.0001**
	voiced – voiceless	-69.822	4.999	24.10	-13.97	< **.0001**
Palatal	aspirated – voiced	95.101	5.191	25.49	18.32	< **.0001**
	aspirated – voiceless	26.351	2.852	96.34	9.240	< **.0001**
	voiced – voiceless	-68.749	5.081	25.68	-13.53	< **.0001**
Velar	aspirated – voiced	103.371	5.172	25.12	19.99	< **.0001**
	aspirated – voiceless	20.462	2.645	77.66	7.74	< **.0001**
	voiced – voiceless	-82.909	4.987	23.86	-16.62	< **.0001**

Furthermore, another series of post hoc tests with the Tukey method for a family of four means (bilabial, alveolar, palatal and velar) were carried out to find out whether there were significant differences between the VOT values of voiced, voiceless unaspirated and voiceless aspirated stops within places of articulation. As seen in Table 14 on the following page, the results show that there were no significant differences between the VOT values for voiced alveolar and velar stops (p = 0.11) or for bilabial and velar stops (p = 0.08). Similarly, no significant differences were found between the VOT values for voiceless unaspirated alveolar and bilabial stops (p = 0.5) and also for palatal and velar stops (p = 0.2). Moreover, there was no significant difference between the VOT values for voiceless aspirated alveolar and bilabial stops (p = 0.98). Other than these, all the pairwise comparisons within place of articulation reached statistical significance.

Table 14. Results of post-hoc within-voicing pairwise comparisons for VOT by place categories. *P* values are adjusted based on the Tukey method for a family of 4 means.

Voicing	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Voiced	alveolar – bilabial	11.772	2.369	191.54	4.97	**< .0001**
	alveolar – palatal	-13.923	3.050	82.52	-4.57	**0.0001**
	alveolar – velar	5.709	2.522	189.96	2.26	0.1103
	bilabial – palatal	-25.701	3.248	57.14	-7.91	**< .0001**
	bilabial – velar	-6.062	2.503	177.75	-2.42	0.0767
	palatal – velar	19.639	3.045	90.39	6.45	**< .0001**
Voiceless	alveolar – bilabial	2.907	2.081	180.16	1.38	0.5027
	alveolar – palatal	-12.857	2.839	66.41	-4.53	**0.0001**
	alveolar – velar	-7.378	2.052	164.29	-3.59	**0.0024**
	bilabial – palatal	-15.764	3.098	48.46	-5.09	**< .0001**
	bilabial – velar	-10.285	2.093	143.82	-4.91	**< .0001**
	palatal – velar	5.479	2.730	64.93	2.01	0.1960
Aspirated	alveolar – bilabial	-0.7667	2.124	184.85	-0.37	0.9838
	alveolar – palatal	-22.744	2.906	71.63	-7.83	**< .0001**
	alveolar – velar	-11.376	2.310	183.37	-4.92	**< .0001**
	bilabial – palatal	-21.978	3.131	50.28	-7.02	**< .0001**
	bilabial – velar	-10.609	2.307	164.25	-4.59	**< .0001**
	palatal – velar	11.369	2.932	81.06	3.88	**0.0012**

5.2.4 Summary and implication of results for VOT

We have established that there is a significant difference in VOT values between voiced and voiceless unaspirated stops as well as between voiceless aspirated and voiceless unaspirated stops for all places of articulation. The VOT difference between voiced stops on the one hand and voiceless aspirated and voiceless unaspirated stops on the other is large and VOT alone clearly distinguishes them. However, this does not seem to be the case when it comes to distinguishing between voiceless aspirated and voiceless unaspirated stops. As we have seen, the difference in VOT between these two voiceless stops is not large in comparison with that between voiced stops on the one hand and voiceless aspirated and voiceless unaspirated stops on the other. That is, the VOT values for voiceless unaspirated stops are only about 23 ms and 17 ms longer than for voiceless aspirated stops for females and males respectively. These results confirm and are consistent with previous findings (Cohn, 1993a; Cohn & Ham, 1998; Cohn & Lockwood, 1994).

The findings also establish that the stops in Madurese can be primarily classified into two main phonetic categories by VOT, i.e voiced and voiceless stops. The voiceless category consists of two types: voiceless unaspirated and slightly aspirated stops. Regarding the question if voiced and voiceless aspirated stops share phonetic properties, we can confirm that they show completely different phonetic properties in their VOT values. However, although voiceless unaspirated and voiceless aspirated stops have phonetically similar VOT values, as discussed in Section 1.2, Madurese is best described as a language with a three-way laryngeal contrast. This suggests that the phonetics-phonology relationship is not always straightforward.

5.3 Stop Closure Duration

5.3.1 Descriptive statistics on stop closure duration

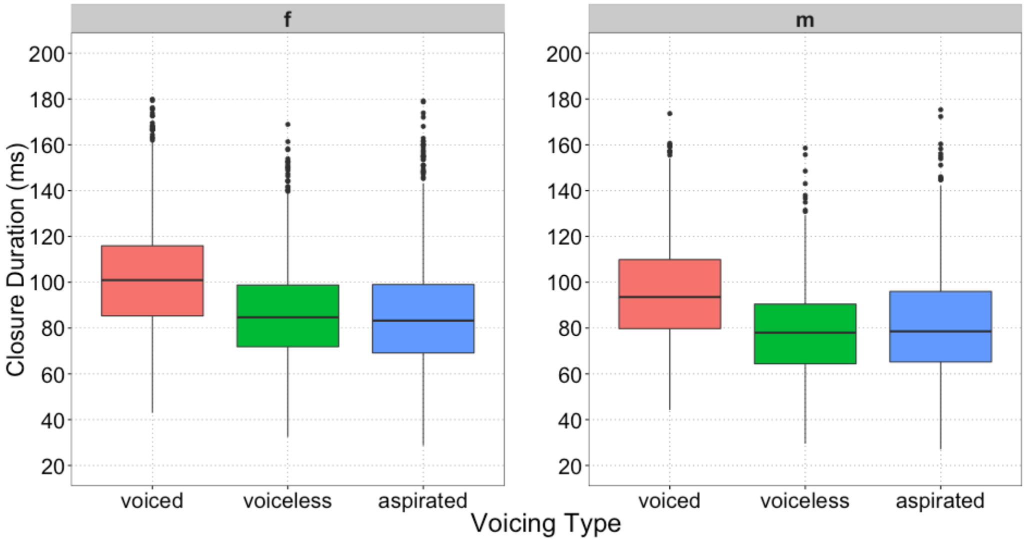


Figure 5–6. Boxplots of closure duration by voicing type and gender averaged across four places of articulation. Females (f) are shown on the left panel and males (m) are on the right panel. The horizontal lines indicate median values and the vertical lines (whiskers) indicate the lower and the upper quartiles respectively.

Figure 5–6 above shows the closure duration for voiced, voiceless unaspirated and voiceless aspirated stops in Madurese grouped by voicing categories and gender. As we can see, there seem to be no differences in closure duration between voiceless unaspirated and voiceless aspirated stops. The closure durations of these two voicing categories appear to overlap considerably. In contrast, the closure durations for voiceless unaspirated and voiceless aspirated stops look slightly shorter from the closure duration for voiced stops. Figure 5–6 also shows that there appears to be no gender distinction in the closure duration of Madurese stops. This may suggest that female and male speakers produce voiced, voiceless unaspirated and voiceless aspirated stops with more or less similar closure duration.

It is worth mentioning that some voiced stops are not fully voiced. There are two types of these: the first type is one where the start of closure for voiced stops begins without voicing but ends with voicing that coincides with the onset of the following vowel and the second type is one where the start of closure for voiced stops begins with voicing but is subsequently extinguished before the closure release. The later type can generally be observed in voiced palatal and velar stops. We examined both cases but found that they occurred very rarely, accounting for just about five percent of the data. For this reason, we did not group them into different categories of analysis nor treat them differently from the rest of the voiced categories.

Figure 5–7 on the following page shows some individual variations in closure duration for the three voicing categories. In this regard, speakers F1, F2, F3, F5, F6, F7, M1, M2, M5 and M6 show similar patterns in which their voiced stops have longer closure durations than their voiceless unaspirated and voiceless aspirated counterparts and where these two voiceless stops categories show similar closure durations. In contrast, speakers F4 and F8 show relatively different closure durations between the three voicing categories. Specifically, the closure duration for voiced stops of these two speakers is longer than for voiceless unaspirated stops while their closure duration for voiceless aspirated stops is the shortest.

However, speakers M3 and M7 show a different pattern from the speakers mentioned previously. The closure duration for voiced stops of these two speakers is longer than for voiceless unaspirated and voiceless aspirated stops whereas the closure duration for voiceless aspirated stops is longer than for voiceless unaspirated stops. Moreover, speaker M4 also shows a different pattern from the other speakers. This is because the closure durations for voiced and voiceless unaspirated stops of the speaker are similar in comparison with the closure duration for voiceless unaspirated stops. In this case, the closure durations for voiced and voiceless unaspirated stops are longer than for voiceless aspirated stops.

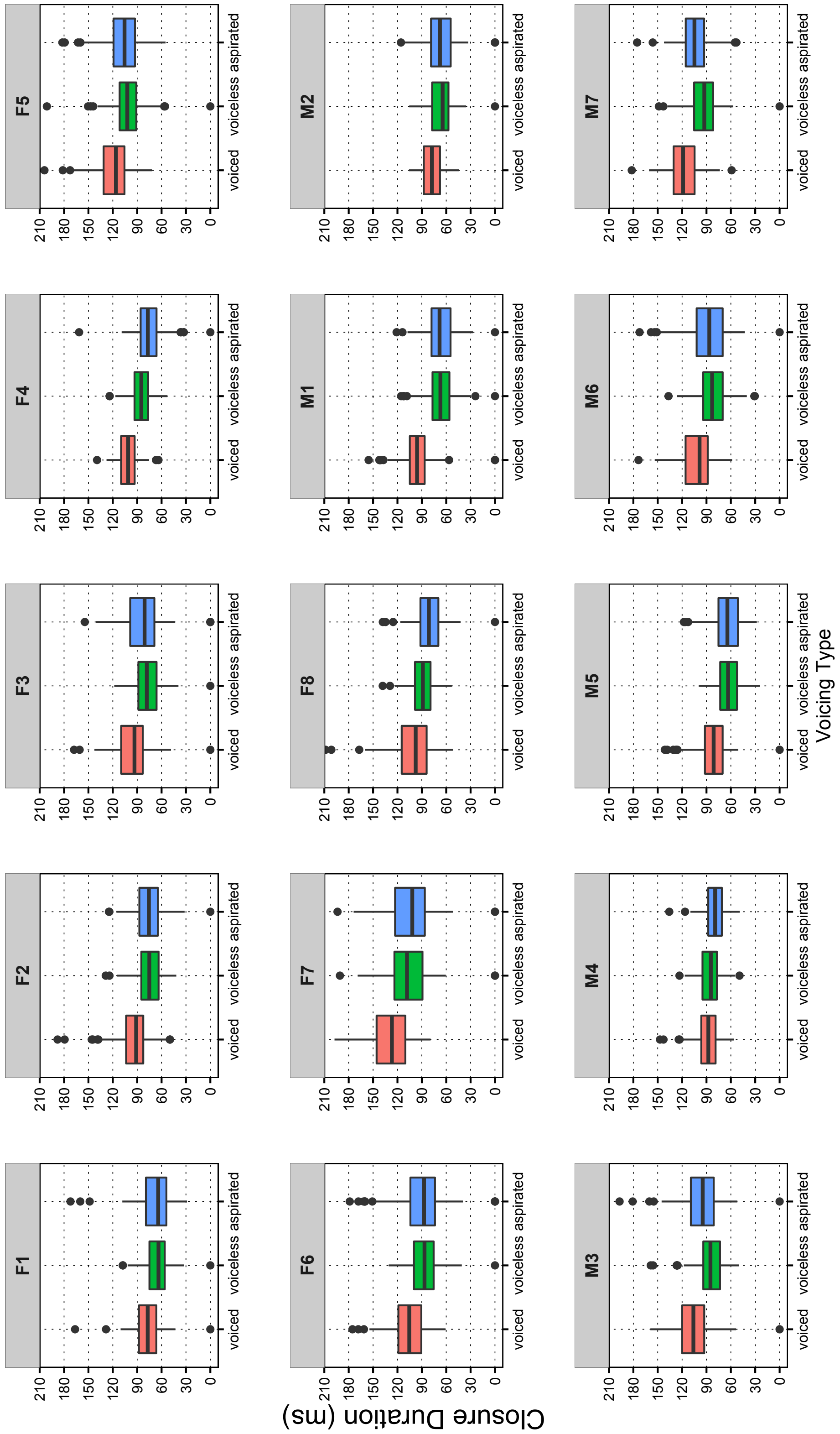


Figure 5–7. Individual variation in the realisation of closure duration for voiced, voiceless unaspirated and voiceless aspirated stops. F stands for females and M for males.

5.3.2 Model comparison for closure duration

Our main concern with regard to closure duration was to find out whether voiced, voiceless unaspirated and voiceless aspirated stops in Madurese had different closure durations. For this purpose, we compared three mixed-effects models to estimate the differences in closure duration between the voicing categories. The models we compared for closure duration were:

clo1: ClosDur ~ Voicing + (1 | Speaker) + (1 | Word)
clo2: ClosDur ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
clo3: ClosDur ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 |Word)

Table 15. Log-likelihood results for closure duration model comparison

	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
clo1	6		70654	-35300			
clo2	11	70354	70431	-35166	268.4193	5	< 2e-16***
clo3	26	69827	70009	-34887	557.1697	15	0.1706

Log-likehood ratio tests were subsequently conducted using the *anova()* function in R for the model comparison in order to find out whether adding complexity to a certain model would improve its goodness of fit and was therefore justified by the data. Table 15 confirms that the model clo2 was the maximal model justified by our data for closure duration. The model includes Voicing as the fixed effect and it also contains by-word random intercepts and by-speaker random slopes for Voicing and Place as the random effects. The test also confirms that there is no gender effect in closure duration, indicated by the fact that adding Gender to the model clo3 does not result in a significantly improved model.

5.3.3 Inferential statistics on stop closure duration

Table 16. The output of a linear mixed-effects model for closure duration. Voiced is the reference level for Voicing.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	99.717	3.806	18.18	26.199	< **.0001**
VoicingVoiceless	-17.324	2.202	49.31	-7.869	< **.0001**
VoicingAspirated	-15.668	2.229	52.37	-7.028	< **.0001**

Table 16 shows the results of a linear mixed-effects model for closure duration. In this model, voiced stops are used as the reference level for Voicing and Voicing is treatment-coded. As shown in the table, the difference between the mean closure duration values for voiceless unaspirated and voiced stops was found to be significant (*p* < .0001). The mean closure duration for voiceless unaspirated stops was estimated to be 82 ms or about 17 ms shorter than for voiced stops. The difference between the mean closure duration values for voiceless aspirated and voiced stops was also statistically significant (*p* < .0001). The mean closure duration for voiceless aspirated stops was estimated to be 84 ms or around 16 ms shorter than that for voiced stops. In order to compare closure durations for voiceless unaspirated and voiceless aspirated stops, Voiceless was set as the reference level. As expected from the boxplots shown in Figure 5–6 earlier, the difference between voiceless unaspirated and aspirated stops was not statistically significant (*p* = 0.43), suggesting that they are very similar in closure durations.

5.3.4 Implication of results

The results are in line with the previous study by Cohn and Ham (1998), who also observe that voiceless unaspirated and voiceless aspirated stops have similar closure durations. However, unlike their study, which observes voiced stops to be shorter than voiceless unaspirated and voiceless aspirated stops, our finding indicates that it is voiced stops that have longer duration than voiceless unaspirated and aspirated stops. With regard to our research question on whether voiced and voiceless aspirated stops share closure duration to the exclusion of voiceless unaspirated stops, our finding confirms that they do not share this acoustic property. In fact, what we found here is that it is voiceless unaspirated and voiceless aspirated stops that pattern together in closure duration. Therefore, our findings on closure duration are similar to those on VOT in which voiceless unaspirated and voiceless aspirated stops also pattern together to the exclusion of voiced stops.

5.4 Fundamental Frequency (F0)

5.4.1 Descriptive statistics on F0

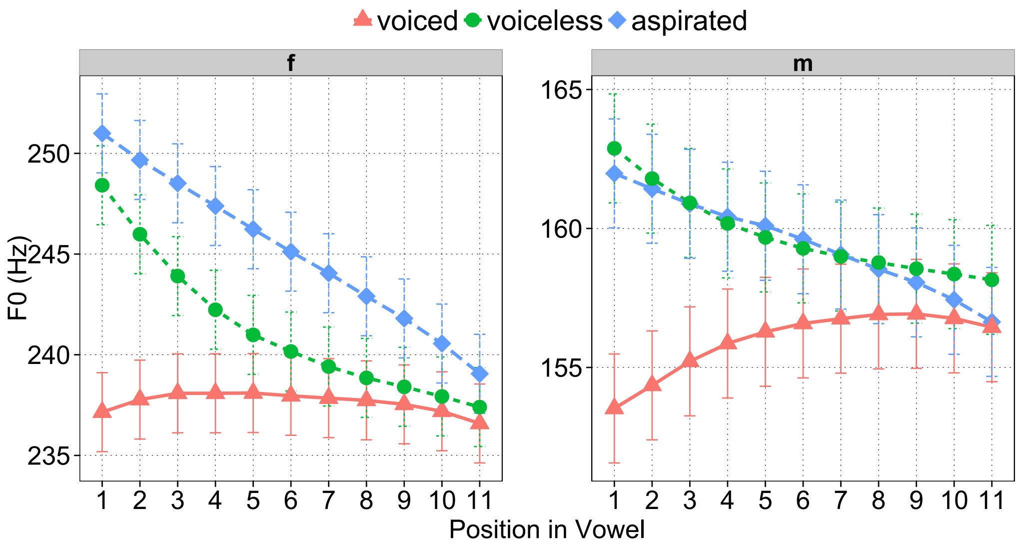


Figure 5–8. Mean F0 of vowels following voiced, voiceless aspirated and voiceless unaspirated stops measured at 11 equidistant timepoints; female on the left panel and male on the right panel. Error bars represent 95% confidence interval.

Figure 5–8 shows the mean F0 values for vowels following voiced, voiceless unaspirated and voiceless aspirated stops in Madurese averaged across speakers, places of articulation and repetitions measured at eleven equidistant timepoints into the vowels. Female and male speakers are plotted separately given the fact that there is a general tendency for female speakers to have higher F0. It is true that comparing between females and males is also possible through F0 normalisation. However, the fact that we observe quite distinct F0 trajectory patterns in Madurese with respect to gender has also contributed to the decision to plot them separately.

This is particularly clear if we take a look at Figure 5–8 above, where female speakers appear to have a relatively different F0 trajectory pattern to male speakers particularly with respect to voiceless unaspirated and voiceless aspirated stops. Specifically, the F0 values for the two voiceless stop categories appear to be relatively separated from each other for female speakers but they appear to overlap considerably for male speakers.

Despite the different behaviours of female and male speakers with regard to the F0 trajectory patterns following the three voicing categories, we can still see a general trend that the F0 for voiced stops appears to be slightly higher than that for voiceless unaspirated and voiceless aspirated stops (Hombert *et al.*, 1979). This is particularly evident when we look at the first quarter of the vowel duration.

However, if we look at the F0 plot for each speaker in Figure 5–9 on the following page, it turns out that the F0 patterns are in fact not based on gender such as the one shown in Figure 5–8 previously. This is because some female and male speakers appear to show similar patterns. For example, two female speakers F7 and F8 have similar F0 patterns to male speakers shown in Figure 5–8. In contrast, male speakers such as M3 and M5 also show similar F0 patterns to female speakers shown in Figure 5–8. Thus, this could be that the patterns we have seen in Figure 5–8 previously may be simply due to the result of mathematical averaging. However, it is important to note that in the following statistical analysis, we still include gender as a factor because, as we will see, it turns out that its inclusion is justified by our data.

Figure 5–8 also shows that there is a clear difference in F0 values between female and male speakers. That is, females appear to have higher F0 values than males do. Gender differences in F0 such as this is very common cross-linguistically. This may result from the fact that anatomically females and males have a different vocal tract size. In general, the female vocal tract is around 15% shorter than the male vocal tract (Goldstein, 1980). In this case, as male speakers have a larger vocal tract than female speakers, we would expect that males will have a lower F0 than females. In fact, there is evidence in the literature that the female F0 is about 1.7 times higher than that of the male (Peterson & Barney, 1952).

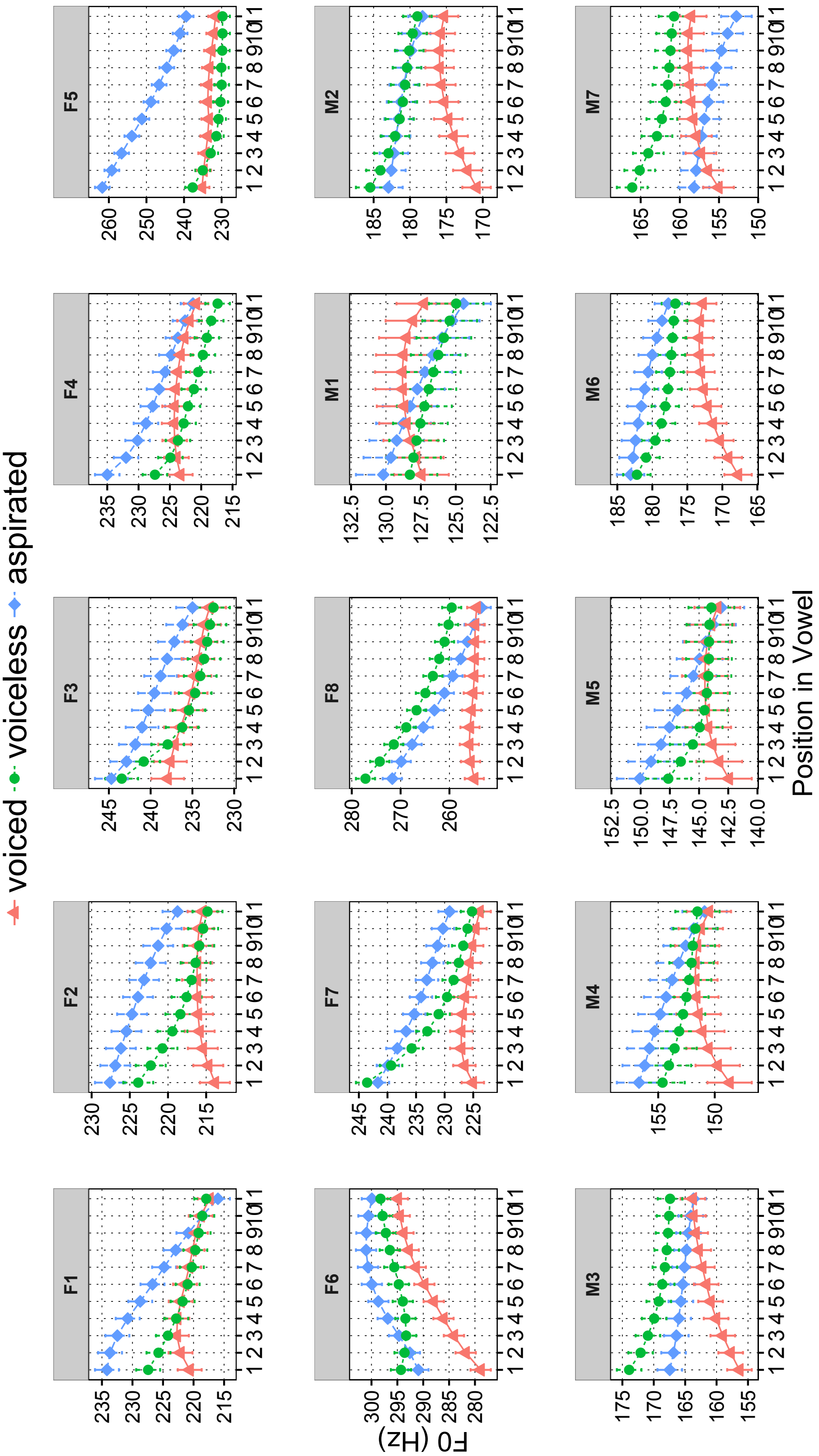


Figure 5–9. Individual speakers' F0 values of fifteen speakers. F stands for females and M stands for males.

Figure 5–9 on the preceding page shows individual variation in F0 realisation for 15 speakers of Madurese. There are eight female speakers and seven male speakers. In terms of F0 patterns, we can classify the speakers into three groups. Group 1 consists of speakers where the F0 following voiced, voiceless unaspirated and voiceless aspirated stops are relatively separated from one another. This group includes speakers F1, F2 and M3.

Group 2 consists of speakers where F0 following voiced stops is lower than that following voiceless unaspirated and voiceless aspirated stops, but the two voiceless stops do not seem to differ from one another. This group includes F3, F6, F7, F8, M2, M4, M5 and M6. As can be seen, Group 2 itself can be further divided into two categories, namely those in which voiceless unaspirated stops have relatively higher F0 and those in which voiceless aspirated have relatively higher F0. However, since they look overlapping, we do not attempt to interpret this variability further.

Group 3 consists of speakers where F0 following voiceless aspirated stops is higher than that following voiced and voiceless unaspirated stops, but where voiced and voiceless unaspirated stops show similar F0 values. This group consists of two speakers, F4 and F5. In addition, two speakers do not appear to belong to any of the groups mentioned above. That is, M1 where there appears to be no distinction in his F0 following the three stop categories and M7 where his F0 for voiceless unaspirated stops is higher than for voiced and voiceless aspirated stops.

The degree of variability in the realisation of F0 among speakers may suggest that they implement (de)voicing in different manners. For example, speakers F3, F4, F5 and M1 tend to do a lot of devoicing indicated by overlapping values between the three categories. In contrast, speakers F7, F8, M2 and M8 appear to maximise the distinction between voiced stops on the one hand and voiceless unaspirated and aspirated stops on the other. Moreover, the fact that the distinction in the F0 realisation appears to be robust only at vowel onset may suggest that this is due to the consonantal effects. However, despite individual variations, there is a general tendency that the F0 following voiced stops is lower than that following voiceless unaspirated and voiceless aspirated stops. As we can see, such variations mostly come from the F0 realisations for voiceless unaspirated and voiceless aspirated stops.

5.4.2 Model comparison for F0

To estimate the effects of voicing and gender categories on the realisation of F0 following voiced, voiceless unaspirated and voiceless aspirated stops in Madurese we compared the following linear mixed-effects models.

f01: F0 ~ Voicing + (1 | Speaker) + (1 | Word)
f02: F0 ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
f03: F0 ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 | Word)
f04: F0 ~ Voicing * Gender + (1 + Voicing | Speaker) + (1 | Word)

Table 17. Log-likelihood results for F0 model comparison

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
f01	6	258932	258989	-129460			
f02	11	255990	256094	-127984	2951.91	5	< 2.2e-16 ***
f03	12	255966	256079	-127971	26.19	1	3.099e-07 ***
f04	14	255964	256095	-127968	6.41	2	0.0405 *

In all of these model comparisons, F0 values were averaged across all eleven timepoints. As we can see in Table 17, Model f04 is the maximal model justified by our F0 data and we used the model to analyse the data. Specifically, the model includes Voicing, Gender and an interaction term of Voicing and Gender as fixed effects. It also includes by-speaker and by-word random intercepts as well as by-speaker random slopes for Voicing as the random effects.

In the following we present the results of F0 analysis by considering F0 in two regions of the vowel: at vowel onset (the average of timepoints 1-3) and vowel midpoint (the average of timepoints 5-7). Vowel onset is chosen for analysis because it is the part of vowels closest to the stops where the F0 perturbation is expected to be more pronounced. Vowel midpoint is also selected for analysis to see the extent to which the effect of the preceding stop on the fundamental frequency of the following vowel continues or persists into the following vowel. To allow for comparison across speakers and genders, F0 in Hertz was converted to semitones (St) using the f2st function in the *hqmisc* package (Quené, 2014) implemented in R. We used 100 Hz as the base frequency for all speakers.

5.4.3 Inferential statistics on F0 at vowel onset

Table 18. The output of a linear mixed-effects model of F0 at vowel onset. Voiceless unaspirated is the reference category for Voicing and female is for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	15.3494	0.6551	15.044	23.43	**< .0001**
VoicingVoiced	-0.4951	0.1675	16.659	-2.956	**0.0089**
VoicingAspirated	0.4127	0.1864	16.191	2.215	**0.0415**
GenderMale	-7.2215	0.9583	15.002	-7.536	**< .0001**
VoicingVoiced:GenderMale	-0.3153	0.2387	14.984	-1.321	0.2065
VoicingAspirated:GenderMale	-0.4762	0.2676	15.005	-1.779	0.0954

Table 18 summarises the results of a linear mixed-effects model of F0 as a function of Voicing and Gender at vowel onset. Voiceless unaspirated is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment coded in order to facilitate comparisons between voiced and voiceless unaspirated stops and between voiceless aspirated and voiceless unaspirated stops in particular.

As shown in Table 18, there was a significant difference between the mean females' F0 values for voiced and voiceless unaspirated stops at vowel onset (*p* = 0.01). The mean females' F0 for voiced stops was estimated to be 14.85 St. The mean females' F0 values for voiceless aspirated and voiceless unaspirated stops also turned out to be significantly different (*p* = 0.042). The females' F0 value for voiceless aspirated stops was estimated to be 15.76 St.

Table 19. Results of post-hoc within-gender pairwise comparisons for F0 by voicing categories at vowel onset. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	0.7497	0.1588	18.54	4.722	**0.0005**
	aspirated - voiceless	0.3343	0.1574	18.30	2.124	0.1359
	voiced - voiceless	-0.4154	0.1548	18.57	-2.683	**0.0441**
Male	aspirated - voiced	0.7369	0.1694	18.42	4.350	**0.0011**
	aspirated - voiceless	-0.0468	0.1680	18.22	-0.278	0.9899
	voiced - voiceless	-0.7837	0.1651	18.43	-4.745	**0.0005**

In order to find out whether the differences between the F0 values for voiced, voiceless unaspirated and voiceless aspirated stops at vowel onset were significant for female and male speakers, a series of post hoc tests were conducted. Table 19 shows the post-hoc, within-gender pairwise comparisons by voicing at vowel onset.

The difference in F0 between voiceless aspirated and voiced stops was significant for either gender (females: *p* = 0.001, males: *p* = 0.001). However, the difference between voiceless aspirated and voiceless unaspirated stops was not significant for either gender (females: *p* = 0.13, males: *p* = 0.99). Moreover, there was a significant difference between the F0 values for voiced and voiceless unaspirated stops for either gender (females: *p* = 0.044, males: *p* = 0.001).

5.4.4 Inferential statistics on F0 at vowel midpoint

Table 20. The output of a linear mixed-effects model F0 at vowel midpoint. Voiceless unaspirated is the reference category for Voicing and female is the reference category for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	14.9874	0.6567	15.056	22.823	**< .0001**
VoicingVoiced	-0.0888	0.1170	19.874	-0.759	0.4569
VoicingAspirated	0.4383	0.1389	17.895	3.155	**0.0055**
GenderMale	-7.0996	0.9605	15.003	-7.392	< .0001
VoicingVoiced:GenderMale	-0.2184	0.1594	14.961	-1.371	0.1907
VoicingAspirated:GenderMale	-0.4636	0.1945	14.995	-2.384	**0.0308**

Table 20 provides the results of a linear mixed-effects model of F0 as a function of Voicing and Gender at vowel midpoint. Voiceless unaspirated is the reference level for Voicing and female is the reference level for Gender. Voicing is also treatment coded in order to allow for comparisons between voiced and voiceless unaspirated stops and in particular between voiceless aspirated and voiceless unaspirated stops.

As seen in Table 20, there was no significant difference between the mean females' F0 values for voiced and voiceless unaspirated stops at vowel midpoint (*p* = 0.46). The mean females' F0 for voiced stops was estimated to be 14.9 St. However, the mean females' F0 values for voiceless aspirated and voiceless unaspirated stops was significantly different (*p* = 0.006). The females' F0 value for voiceless aspirated stops was estimated to be 15.43 St.

Table 21. Results of post-hoc within-gender pairwise comparison for F0 by voicing at vowel midpoint. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	0.4331	0.1235	20.85	3.508	**0.0063**
	aspirated - voiceless	0.4001	0.1145	20.45	3.495	**0.0067**
	voiced - voiceless	-0.0331	0.0989	23.09	-0.335	0.9826
Male	aspirated - voiced	0.2453	0.1313	20.43	1.869	0.2112
	aspirated - voiceless	-0.0234	0.1218	20.12	-0.192	0.9966
	voiced - voiceless	-0.2688	0.1049	22.38	-2.562	**0.0520**

To find out whether the differences between the F0 values for voiced, voiceless unaspirated and voiceless aspirated stops at vowel midpoint were significant for female and male speakers, a series of post hoc tests were also conducted. Table 21 shows the post-hoc, within-gender pairwise comparisons by voicing at vowel midpoint. As we can see, there was a significant difference between the F0 values for voiceless aspirated and voiced stops for females (p = 0.006), but the difference was not significant for males (p = 0.21). The difference in F0 between voiceless aspirated and voiceless unaspirated stops was also significant for females (p = 0.007), but it was not significant for males (p = 0.99). In contrast, the difference in F0 between voiced and voiceless unaspirated stops was not significant for females (p = 0.98), but it turned out to be significant for males (p = 0.05).

5.4.5 Summary and implication of results for F0

We have presented the results of a linear mixed-effects model of F0 as a function of Voicing and Gender at vowel onset and vowel midpoint. At vowel onset, female and male speakers show significant differences between the F0 values for voiced and voiceless unaspirated stops and also between F0 values for voiced and voiceless aspirated stops. However, no significant differences in F0 values were found between voiceless aspirated and voiceless unaspirated stops at vowel onset for either gender.

Slightly different results as compared to F0 values at vowel onset were found for F0 values at vowel midpoint. There is no significant difference in females' F0 values between voiced and voiceless unaspirated stops, but there is a significant difference between their males' counterparts. In contrast, there are significant differences in females' F0 values between voiceless aspirated and voiceless unaspirated stops as well as between voiced and voiceless aspirated stops, but no significant differences were found between their males' counterparts.

It is not always clear why the results of F0 analysis at vowel midpoint are gender-specific. As shown in Table 21, female speakers distinguish F0 between voiceless aspirated and voiced stops and between aspirated and voiceless unaspirated stops while male speakers do not distinguish F0 for these two sets of categories. While male speakers distinguish F0 between voiced and voiceless unaspirated stops, female speakers do not. One explanation with regard to the gender-specific results may be that F0 at vowel midpoint is not an important cue for distinguishing the voicing categories in Madurese while it is at vowel onset.

Another possible reason may be related to language change. In this case, there are studies which show that females and males have a different tendency to lead sound changes. For example, in her corpus study on Seoul Korean stops, Kang (2014) found that her female speakers tend to put less distinction in VOT for aspirated and lenis stops and they distinguish them more by F0 instead. A slightly different example comes from a study by Abramson *et al*. (2007) on voice register in Khmu'. They found that male speakers do not distinguish between Register 1 and Register 2 by harmonic intensity ratios while female speakers do. They associate the gender-related finding to the fact that females spend most of their time in the village while males tend to travel away and therefore become more exposed to linguistic diversity along the way. As the present study does not happen to have any sociolinguistic data, we could only speculate that these could also take place in Madurese.

With regard to the question whether voiced and voiceless aspirated stops share phonetic properties, our F0 findings indicate that they do not share this acoustic property. Instead, F0 robustly distinguishes between voiced and voiceless aspirated stops. With respect to the phonetic prediction of whether voiced and voiceless aspirated stops share a feature [+LL], the F0 results also suggest that voiceless aspirated stops cannot be considered as having the feature whereas voiced stops can. This is because voiceless aspirated stops have higher F0 values as opposed to lower F0 values as predicted for the feature [+LL]. In other words, the F0 results suggest that the feature [+LL] cannot be considered as the phonological feature shared by voiced and voiceless aspirated stops in Madurese.

5.5 H1*-A1*

5.5.1 Descriptive statistics on H1*-A1*

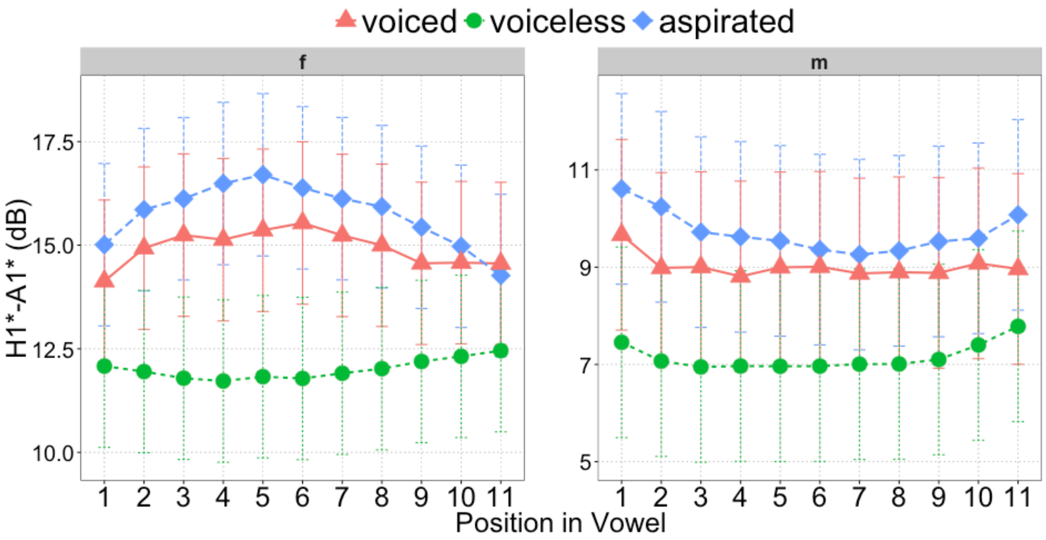


Figure 5–10. Mean H1*-A1* of vowels following voiced, voiceless unaspirated and voiceless aspirated measured at 11 equidistant timepoints; female on the left panel and male on the right panel. Error bars represent 95% confidence interval.

Figure 5–10 shows plots for mean H1*-A1* of vowels following voiced, voiceless unaspirated and voiceless aspirated stops measured at 11 equally spaced timepoints into the vowels. As we can see, the H1*-A1* values for voiceless unaspirated stops appear to be lower than those for voiced and voiceless aspirated stops. In this case, voiceless aspirated stops appear to have the greatest values, but the values overlap quite extensively with those for voiced stops. More importantly, voiced and voiceless aspirated stops seem to pattern together in their H1*-A1* values excluding those for voiceless unaspirated stops. As Figure 5–10 shows, this pattern looks fairly consistent across the vowel duration as well as across genders.

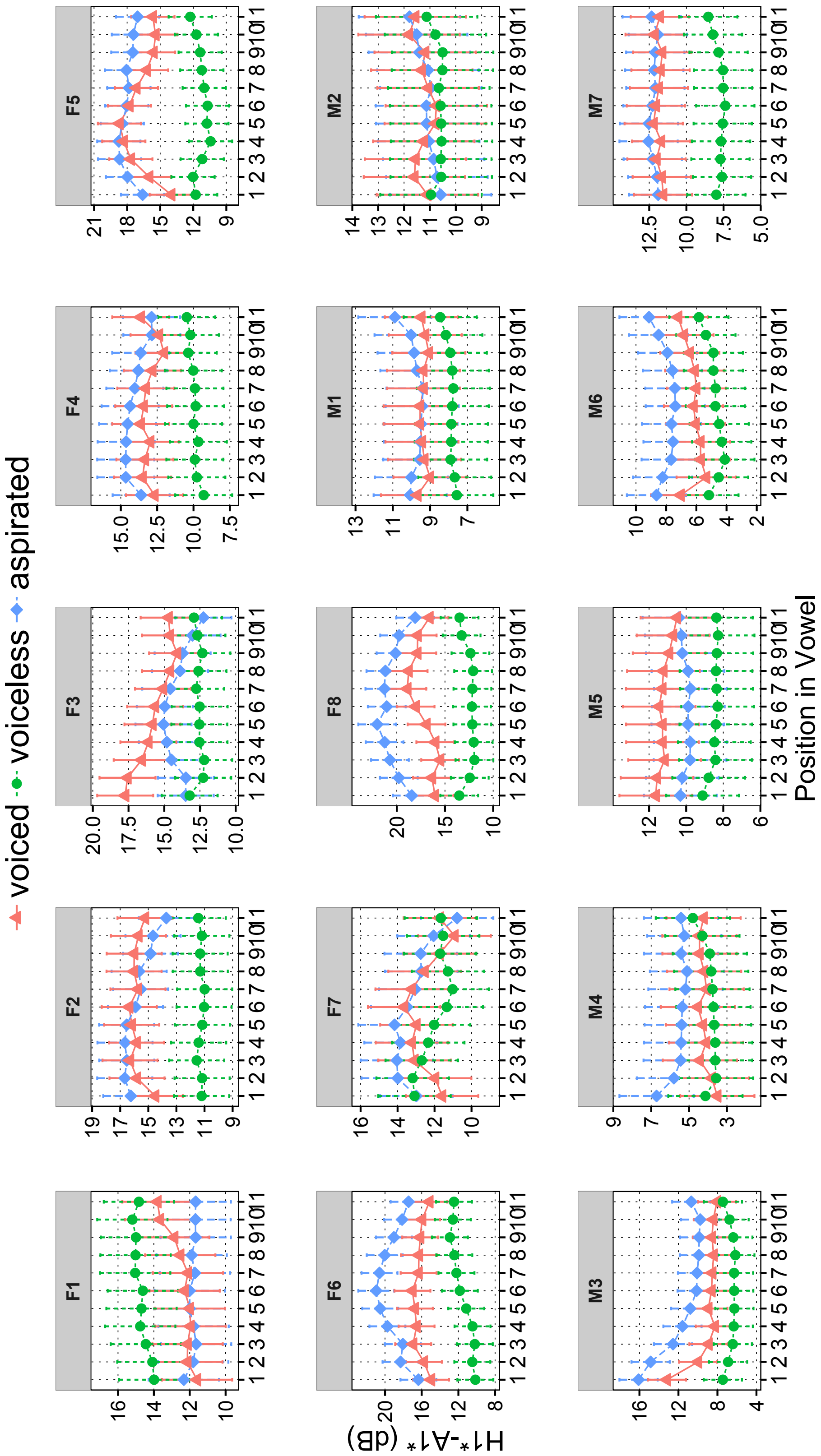


Figure 5–11. Individual speakers' H1*-A1* values of fifteen speakers. F stands for females and M stands for males.

Figure 5–11 shows individual variation in H1*-A1* realisation for 15 speakers of Madurese. There are eight female speakers and seven male speakers. In terms of H1*-A1* patterns, we can classify the speakers into three groups. Group 1 consists of speakers where the H1*-A1* following voiced, voiceless unaspirated and voiceless aspirated stops are relatively separated from one another. The group includes F6, F8, M3, M5 and M6. Group 2 consists of speakers where H1*-A1* following voiceless unaspirated stops is lower than that following voiceless voiced and aspirated stops and where voiced and aspirated stops do not seem to differ. The group includes F2, F4, F5, M1 and M7. Finally, group 3 consists of speakers where H1*-A1* following the three stop categories does not seem to differ, which includes F7, M1, M2 and M4. However, one speaker F1 does not belong to any of the groups mentioned above. This is due to the fact that the H1*-A1* for voiceless unaspirated stops of this speaker is higher than that for voiced and voiceless aspirated stops and where these latter two categories do not seem to differ from one another.

Despite individual differences observed above, we can see a general picture that the H1*-A1* values for voiceless aspirated and voiced stops look higher than for voiceless unaspirated stops. Moreover, in most of the cases, voiceless aspirated and voiced stops appear to pattern together in this spectral property.

5.5.2 Model comparison for H1*-A1*

We estimated the effects of voicing and gender on the realisation of H1*-A1* for voiced, voiceless unaspirated and aspirated stops. We compared the following linear mixed-effects models to find the maximal model justified by our data:

a1a: H1*-A1* ~ Voicing + (1 | Speaker) + (1 | Word)
a1b: H1*-A1* ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
a1c: H1*-A1* ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 | Word)
a1d: H1*-A1* ~ Voicing * Gender + (1 + Voicing | Speaker) + (1 | Word)

Table 22. Log-likelihood results for H1*-A1* model comparison

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
a1a	6	594777	594834	-297383			
a1b	11	591209	591312	-295593	3578.7289	5	< 2.2e-16***
a1c	12	591192	591305	-295584	18.6202	1	1.60E-05***
a1d	14	591195	591327	-295583	1.1914	2	0.5512

In all of these model comparisons, H1*-A1* values were averaged across all eleven timepoints. The log-likelihood ratio test shows that the model a1c was the maximal model that was justified by our H1*-A1* data. Specifically, the model includes Voicing and Gender as the fixed effects. It also includes by-speaker and by-word random intercepts as well as by-speaker random slopes for Voicing as the random effects. It is worth noting that in spite of the plot shown in Figure 5–10, the inclusion of the interaction term for Voicing and Gender was not justified.

In the following we present the results of H1*-A1* analysis by considering H1*-A1* at vowel onset and midpoint that were obtained by averaging timepoints 1-3 and timepoints 5-7 respectively.

5.5.3 Inferential statistics on H1*-A1* at vowel onset

Table 23. The output of a linear mixed-effects model of H1*-A1* at vowel onset. Voiceless unaspirated is the reference category for Voicing and female is the reference category for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	12.0076	0.6103	18.65	19.676	**< .0001**
VoicingVoiced	2.4636	0.5976	18.42	4.122	**0.0006**
VoicingAspirated	3.3740	0.7559	16.82	4.464	**0.0004**
GenderMale	-4.8810	0.7989	15.01	-6.108	**< .0001**

Table 23 summarises the results of a linear mixed-effects model of H1*-A1* as a function of Voicing and Gender at vowel onset. Voiceless unaspirated is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment coded, which is in this case this means comparing voiced stops with voiceless unaspirated stops and voiceless aspirated stops with voiceless unaspirated stops. The table shows that there was a significant difference between the mean females' H1*-A1* values for voiced and voiceless unaspirated stops at vowel onset ($p < 0.001$). The mean females' H1*-A1* value for voiced stops was estimated to be 14.47 dB. The mean females' H1*-A1* values for voiceless aspirated and voiceless unaspirated stops were also significantly different ($p < 0.001$). The female H1*-A1* value for voiceless aspirated stops was estimated to be 15.38 dB. Moreover, the main effect of Gender was also found to be significant ($p < .0001$). In this case, female speakers had higher H1*-A1* values than male speakers did, suggesting that they produced breathier voice quality.

Table 24. Results of post-hoc pairwise comparisons for H1*-A1* by voicing categories at vowel onset. *P* values are adjusted based on the Sidak method for 3 tests.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
aspirated - voiced	0.9105	0.5207	20.92	1.749	0.2588
aspirated - voiceless	3.3741	0.7797	17.65	4.327	**0.0013**
voiced - voiceless	2.4636	0.6149	19.12	4.006	**0.0022**

Table 24 shows the post-hoc pairwise comparisons by voicing averaged across genders at vowel onset. The difference in H1*-A1* between voiceless aspirated and voiced stops was not significant (*p* = 0.26). However, the differences in H1*-A1* values between voiceless aspirated and voiceless unaspirated stops and between voiced and voiceless aspirated stops were significant (*p* < at least 0.01 in both cases).

5.5.4 Inferential statistics on H1*-A1* at vowel midpoint

Table 25. The output of a linear mixed-effects model of H1*-A1* at vowel midpoint. Voiceless unaspirated is the reference category for Voicing and female is the reference category for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	12.0954	0.6292	18.25	19.223	**< .0001**
VoicingVoiced	2.7961	0.6662	19.44	4.197	**0.0005**
VoicingAspirated	3.5244	0.8518	17.26	4.138	**0.0007**
GenderMale	-5.3555	0.8448	15.01	-6.339	**< .0001**

Table 25 above summarises the results of a linear mixed-effects model of H1*-A1* as a function of Voicing and Gender at vowel midpoint. Voiceless unaspirated is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment coded. The results show that there was a significant difference in the mean females' H1*-A1* values between voiced and voiceless unaspirated stops at vowel midpoint (*p* < 0.001). The mean females' H1*-A1* value for voiced stops was estimated to be 14.89 dB. The mean females' H1*-A1* values for voiceless aspirated and voiceless unaspirated stops were also significantly different (*p* < 0.001). The females' H1*-A1* value for voiceless aspirated stops was estimated to be 15.62 dB. Similar to what we see in vowel onset, the main effect of Gender was also significant at vowel midpoint (*p* < .0001).

Table 26. Results of post-hoc pairwise comparisons for H1*-A1* by voicing categories at vowel midpoint. *P* values are adjusted based on the Sidak method for 3 tests.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
aspirated - voiced	0.7189	0.4401	30.48	1.633	0.3014
aspirated - voiceless	3.4287	0.8667	18.10	3.956	**0.0027**
voiced - voiceless	2.7098	0.6662	20.45	4.067	**0.0017**

Table 26 shows the post-hoc pairwise comparisons by voicing averaged across genders at vowel midpoint. The results show that there was no significant difference in H1*-A1* values between voiceless aspirated and voiced stops ($p = 0.30$). However, there were significant differences in H1*-A1* values between voiceless aspirated and voiceless unaspirated stops as well as between voiced and voiceless aspirated stops ($p <$ at least 0.01 in both cases).

5.5.5 Summary and implication of results for H1*-A1*

We have presented the results of a linear mixed-effects model of H1*-A1* as a function voicing averaged across genders. We found that at both vowel onset and midpoint, there are no significant differences between H1*-A1* values for voiced and voiceless aspirated stops. However, there are significant differences in H1*-A1* values between voiceless aspirated and voiceless unaspirated stops and between voiced and voiceless unaspirated stops at both vowel onset and midpoint.

With regard to the question whether voiced and voiceless aspirated stops share acoustic properties to the exclusion of voiceless unaspirated stops, our findings indicate that they share the acoustic property of H1*-A1*. Furthermore, with respect to whether the results for H1*-A1* are in line with the prediction that voiced and voiceless aspirated stops have the feature [+ATR] or [+LL], the findings also suggest that they are. This is because both [+ATR] and [+LL] predict that vowels produced with an advanced tongue root or a lowered larynx would be expected to be breathy, as indicated by greater H1*-A1* values for voiced and voiceless aspirated stops.

5.6 H1*-A2*

5.6.1 Descriptive statistics on H1*-A2*

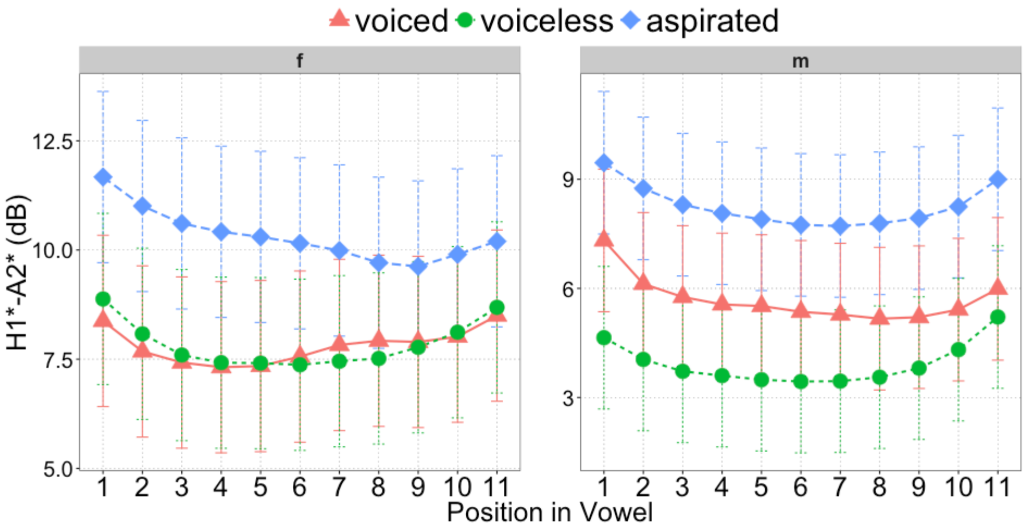


Figure 5–12. Mean H1*-A2* of vowels following voiced, voiceless unaspirated and voiceless aspirated measured at 11 equidistant timepoints; female on the left panel and male on the right panel. Error bars represent 95% confidence interval.

Figure 5–12 shows plots for mean H1*-A2* of vowels following voiced, voiceless unaspirated and voiceless aspirated stops measured at eleven equally spaced timepoints. As we can see, female and male speakers exhibit a different pattern with respect to particularly voiced and voiceless unaspirated stops. Specifically, female speakers do not seem to distinguish between voiced and voiceless unaspirated stops, indicated by a complete overlap in their H1*-A2* values. However, they appear to distinguish between the two voicing categories and voiceless aspirated stops by this spectral measure. Unlike female speakers, male speakers seem to distinguish each of the three voicing categories by H1*A2*. The similarity in H1*-A2* between female and male speakers resides in the fact that the values for voiceless unaspirated stops are consistently lower than those for voiceless aspirated stops.

Note however that if we look at the plots for individual speakers in Figure 5–13, we can see that the patterns that are seemingly based on gender as shown in Figure 5–12 do not paint the whole picture. This is because there are a number of cases where females and males pattern together in this spectral measure. That is, what we see in Figure 5–12 where females and males have different patterns in their H1*-A2* values for the three voicing categories probably results from averaged values, similar to what happens when we plot F0 by gender shown earlier.

Figure 5–13 shows individual variation in H1*-A2* realisation for 15 speakers of Madurese. In terms of H1*-A2* patterns, we can classify the speakers into four groups. Group 1 consists of speakers where the H1*-A2* following voiced, voiceless unaspirated and voiceless aspirated stops are relatively separated from one another. This group includes F6, M1 and M3. Group 2 consists of speakers where H1*-A2* following voiceless aspirated stops is higher than that following voiceless unaspirated and voiced stops and where the latter two categories pattern together. This group includes F2, F4, F5, F6, M4 and M6. Group 3 consists of speakers where H1*-A2* following the three stop categories does not seem to differ. This group includes F3, F7, and F8. Group 4 consists of speakers where the H1*-A2* for voiced and voiceless aspirated stops pattern together excluding voiceless unaspirated stops. This group includes M5 and M7. Speaker F1, however, does not belong to any of the groups because the speaker's H1*-A2* for voiceless unaspirated stops is higher than that for voiced and voiceless aspirated stops.

In spite of individual variations, there is still a general consistency that the H1*-A2* values for voiceless aspirated stops appear to be higher than for voiceless unaspirated or voiced stops. The higher values in H1*-A2* for voiceless aspirated stops are also consistent with the spectral measures presented earlier, namely F0 and H1*-A1*, suggesting that they could be potentially strong acoustic correlates for this category.

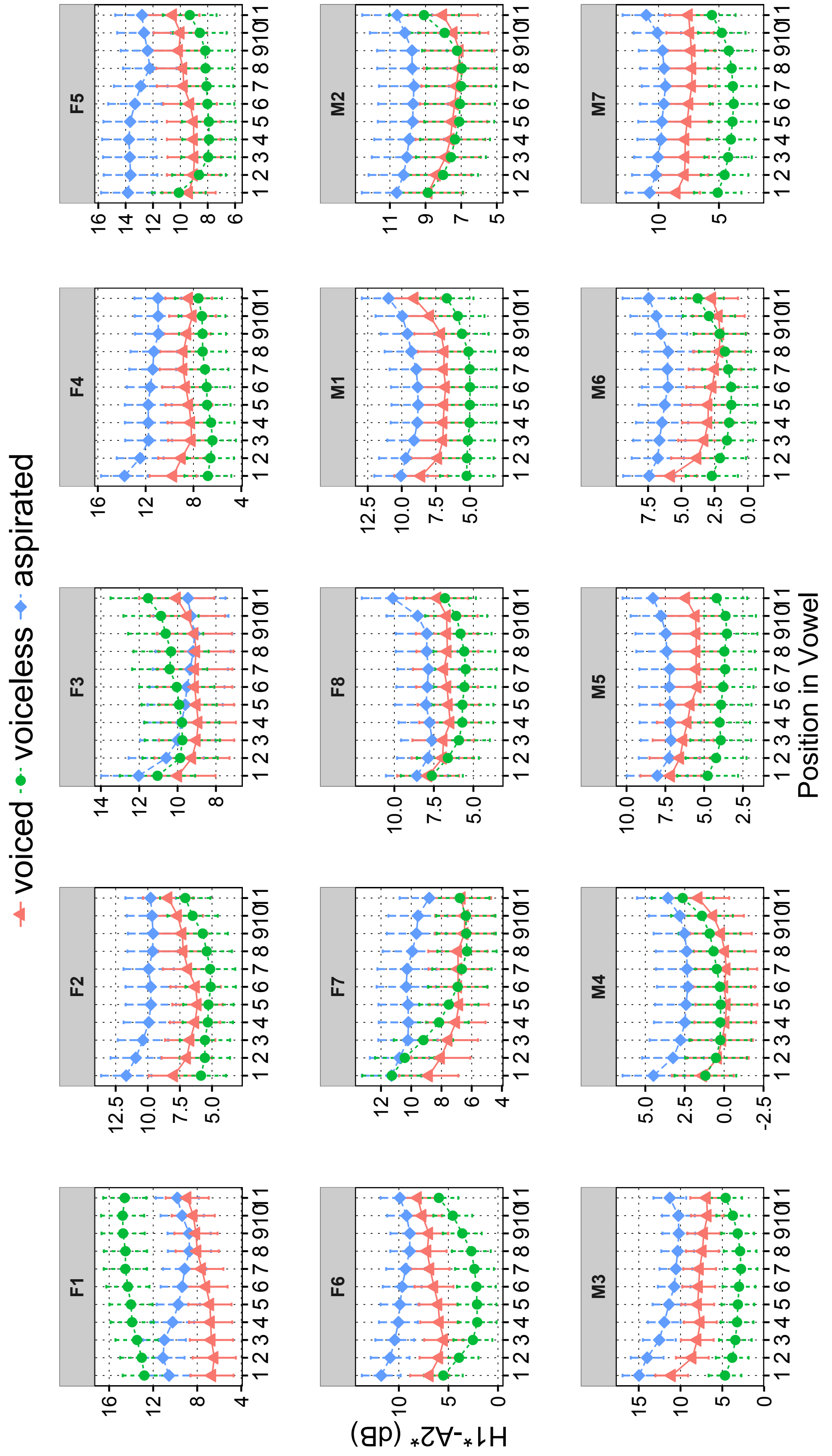


Figure 5–13. Individual speakers' H1*-A2* values of fifteen speakers. F stands for females and M stands for males.

5.6.2 Model comparison for H1*-A2*

In order to estimate the differences in H1*-A2* values for voiced, voiceless unaspirated and voiceless aspirated stops, we compared the following linear mixed-effects models to find the maximal model justified by our data:

a2a: H1*-A2* ~ Voicing + (1 | Speaker) + (1 | Word)
a2b: H1*-A2* ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
a2c: H1*-A2* ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 | Word)
a2d: H1*-A2* ~ Voicing * Gender + (1 + Voicing | Speaker) + (1 | Word)

Table 27. Log-likelihood results for H1*-A2* model comparison

	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
a2a	6	539838	539895	-269913			
a2b	11	532357	532461	-266167	7491.3961	5	< 2e-16***
a2c	12	532354	532467	-266165	4.8188	1	0.02815*
a2d	14	532356	532488	-266164	2.1183	2	0.34674

In all of these model comparisons, H1*-A2* values were averaged across all eleven timepoints. The result shows that the model a2c was the maximal model that was justified by our H1*-A2* data. The model includes Voicing and Gender as the fixed effects. It also includes by-speaker and by-word random intercepts as well as by-speaker random slopes for Voicing as the random effects. It is important to note that in spite of the plot shown in Figure 5–12, the inclusion of the interaction term for Voicing and Gender was not justified either.

In the following we present the results of H1*-A2* analysis by considering H1*-A2* in two regions of the vowel: at vowel onset (the average of timepoints 1-3) and vowel midpoint (the average of timepoints 5-7).

5.6.3 Inferential statistics on H1*-A2* at vowel onset

Table 28. The output of a linear mixed-effects model of H1*-A2* at vowel onset. Voiceless unaspirated is the reference category for Voicing and female is the reference category for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	7.4543	0.9359	26.65	7.965	**< .0001**
VoicingVoiced	0.8267	0.9454	48.78	0.875	0.3861
VoicingAspirated	3.7368	0.976	38.85	3.829	**0.0005**
GenderMale	-2.3761	1.0126	14.95	-2.347	**0.0331**

Table 28 on the preceding page summarises the results of a linear mixed-effects model of H1*-A2* as a function of Voicing and Gender at vowel onset. Voiceless unaspirated is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment coded. As seen in the table above, there was no significant difference between the mean females' H1*-A2* values for voiced and voiceless unaspirated stops ($p = 0.39$), but there was a significant difference between the mean females' H1*-A2* values for voiceless aspirated and voiceless unaspirated stops ($p < 0.001$). The mean females' H1*-A2* value for voiced stops was estimated to be 8.28 dB while that for voiceless aspirated stops was estimated to be 11.19 dB. In addition, there was also a significant main effect of Gender ($p = 0.033$).

Table 29. Results of post-hoc pairwise comparisons for H1*-A2* by voicing categories at vowel onset. *P* values are adjusted based on the Sidak method for 3 tests.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
aspirated - voiced	2.9101	0.7598	146.61	3.830	**0.0006**
aspirated - voiceless	3.7367	0.9907	38.13	3.772	**0.0017**
voiced - voiceless	0.8267	0.9578	47.95	0.863	0.7757

Table 29 shows the post-hoc pairwise comparisons by voicing averaged for the levels of gender at vowel onset. As we can see, the differences in H1*-A2* values between voiceless aspirated and voiced stops as well as between voiceless aspirated and voiceless unaspirated stops were significant ($p <$ at least 0.001 in both cases). However, there was no significant difference in H1*-A2* values between voiced and voiceless aspirated stops ($p = 0.78$).

5.6.4 Inferential statistics on H1*-A2* at vowel midpoint

Table 30. The output of a linear mixed-effects model of H1*-A2* at vowel midpoint. Voiceless unaspirated is the reference category for Voicing and female is the reference category for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	6.9054	1.002	28.97	6.892	**< .0001**
VoicingVoiced	0.8998	0.9928	53.48	0.906	0.3688
VoicingAspirated	3.4352	1.0321	40.97	3.328	**0.0019**
GenderMale	-2.6759	1.019	14.94	-2.626	**0.0191**

Table 30 summarises the results of a linear mixed-effects model of H1*-A2* as a function of Voicing and Gender at vowel midpoint. Voiceless unaspirated is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment coded. As shown in the table above, there was no significant difference between the mean females' H1*-A2* values for voiced and voiceless unaspirated stops (*p* = 0.37), but there was a significant difference between the mean females' H1*-A2* values for voiceless aspirated and voiceless unaspirated (*p* = 0.002). The mean females' H1*-A2* value for voiced stops was estimated to be 7.81 dB while that for voiceless aspirated stops was estimated to be 10.34 dB. Moreover, the result shows that the main effect of Gender was significant at vowel midpoint (*p* < 0.01).

Table 31. Results of post-hoc pairwise comparisons for H1*-A2* by voicing categories at vowel midpoint. *P* values are adjusted based on the Sidak method for 3 tests.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
aspirated - voiced	2.4575	0.7576	193.74	3.244	**0.0042**
aspirated - voiceless	3.3617	1.0469	40.24	3.211	**0.0078**
voiced - voiceless	0.9043	0.9978	53.95	0.906	0.7485

Table 31 shows the post-hoc pairwise comparisons by voicing averaged for the levels of genders at vowel midpoint. As we can see, there were significant differences in H1*-A2* values between voiceless aspirated and voiced stops and between voiceless aspirated and voiceless unaspirated stops (*p* < at least 0.01 in both cases). However, there was no significant difference in H1*-A2* values between voiced and voiceless aspirated stops (*p* = 0.75).

5.6.5 Summary and implication of results for H1*-A2*

We have presented the results of a linear mixed-effects model of H1*-A2* as a function of voicing averaged for the levels of gender. We found that at both vowel onset and midpoint, there are significant differences in H1*-A2* values between voiceless aspirated and voiced stops and between voiceless aspirated and voiceless unaspirated stops. However, there are no significant differences in H1*-A2* values between voiced and voiceless unaspirated stops at both vowel onset and midpoint.

Unlike the results for H1*-A1*, where we observe voiced and voiceless aspirated stops pattern together to the exclusion of voiceless unaspirated stops, the results for H1*-A2* shows a different pattern. This is because voiced and voiceless unaspirated stops pattern together to the exclusion of voiceless aspirated stops. The question is how this could happen. One possible answer to the question is that it is possible that this acoustic property is not relevant for distinguishing certain voicing categories in Madurese. In fact, cases such as these are also quite common in a number of studies that look at voice quality distinctions using spectral measures. For example, in his study on register distinctions in Takhian Thong Chong, DiCanio (2009) found that H1-A3 differentiates between breathy and non-breathy voice while H1-H2 does not. He also found that H1-H2 distinguishes between tense and non-tense voice instead.

With regard to the question whether voiced and voiceless aspirated stops share phonetic properties to the exclusion of voiceless unaspirated stops, our findings indicate that they do not share the acoustic property of H1*-A2*. This is because voiceless aspirated stops have significantly higher H1*-A2* values than voiced stops. In this case, it is voiced and voiceless unaspirated stops that share this phonetic property. Furthermore, with respect to whether the results for H1*-A2* are in line with the prediction that voiced and voiceless aspirated stops have either the feature [+ATR] or [+LL], the findings suggest that either feature is only consistent with voiceless aspirated stops. Recall that vowels produced with either an advanced tongue root or a lowered larynx are predicted to show greater H1*-A2* values. However, this prediction is in contrast with what we observe for voiced stops.

5.7 H1*-A3*

5.7.1 Descriptive statistics on H1*-A3*

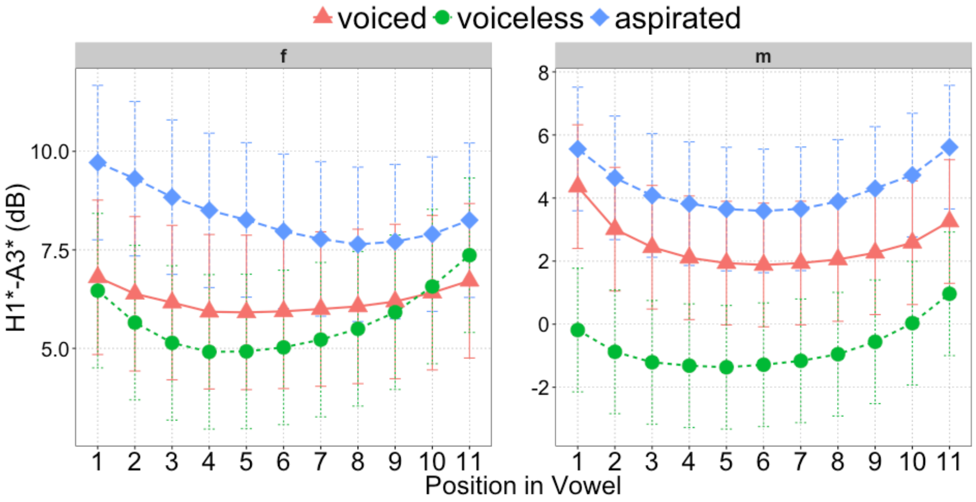


Figure 5–14. Mean H1*-A3* of vowels following voiced, voiceless aspirated and voiceless unaspirated stops measured at 11 equidistant timepoints; female on the left panel and male on the right panel. Error bars represent 95% confidence interval.

The same procedure is performed for displaying the plots of H1*-A3* for female and male speakers separately. Figure 5–14 above shows that females' H1*-A3* values for voiced and voiceless unaspirated stops appear to overlap extensively and their values are lower than for voiceless aspirated stops. In contrast, male speakers show a different pattern where their H1*-A3* values for voiced and voiceless aspirated stops overlap and look higher than their H1*-A3* values for voiceless unaspirated stops. However, both genders have a similarity in the way that their H1*-A3* values for voiceless unaspirated stops are consistently lower than for voiceless aspirated stops.

If we examine the H1*-A3* plot for each speaker in Figure 5–15 on the following page, the H1*-A3* patterns cannot be based on gender such as the one shown in Figure 5–14. This is due to the fact that some female and male speakers share similar patterns. For example, two female speakers F3 and F7 pattern with male speakers M1 and M4. Also, F2 and F8 have similar patterns with M3, M5, M6 and M7. Thus, this suggests that the H1*-A3* patterns shown in Figure 5–14 do not really provide the whole picture, but rather it may result from averaging the values. The degree of individual variation in this measure will be clear if we look at Figure 5–15.

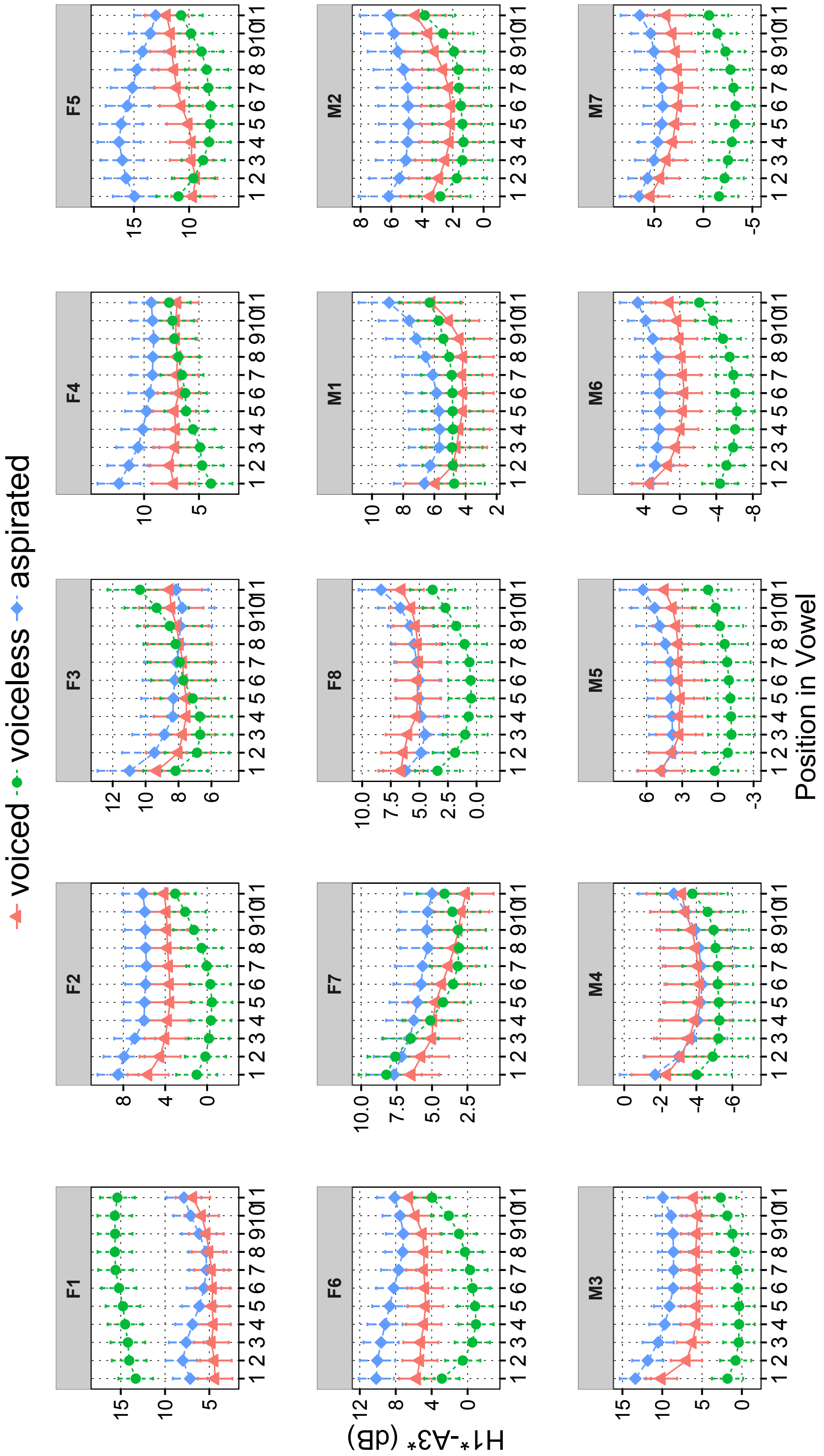


Figure 5–15. Individual speakers' H1*-A3* values of fifteen speakers. F stands for females and M stands for males.

Figure 5–15 shows individual variation in H1*-A3* realisation for 15 speakers of Madurese (8 females and 7 males). In terms of H1*-A3* patterns, we can classify the speakers into three groups. Group 1 consists of speakers where the H1*-A3* following voiced, voiceless unaspirated and voiceless aspirated stops are relatively similar. This group includes F3, F7, M1 and M4. Group 2 consists of speakers where H1*-A3* following voiced and voiceless aspirated stops is higher than that following voiceless unaspirated stops. This group includes F2, F8, M3, M5, M6 and M7.

Group 3 consists of speakers where H1*-A3* following voiceless aspirated stops is higher than that following voiced and voiceless unaspirated stops while these latter two voicing categories appear to pattern together. This group includes F4, F5 and M2. Two speakers F1 and F6, however, do not belong to any of the groups. This is because the H1*-A3* for voiceless unaspirated stops of speaker F1 is higher than for voiced and voiceless aspirated stops. In contrast, speaker F6 shows relatively different values for the three voicing categories.

Even though there are a lot of individual variations, still there is a general picture that the H1*-A3* values for voiceless aspirated stops appear to be higher than for either voiceless aspirated or voiceless unaspirated stops. More importantly, there is a fairly high trend that voiceless aspirated and voiced stops pattern together in this measure. For example, this can be seen in the plots of speakers M3, M5, M6, F2, and F8.

5.7.2 Model comparison for H1*-A3*

In order to estimate the differences between H1*-A3* values following each stop type, we compared the following linear mixed-effects models to find the maximal model justified by our data:

a3a: H1*-A3* ~ Voicing + (1 | Speaker) + (1 | Word)
a3b: H1*-A3* ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
a3c: H1*-A3* ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 | Word)
a3d: H1*-A3* ~ Voicing * Gender + (1 + Voicing | Speaker) + (1 | Word)

Table 32. Log-likelihood results for H1*-A3* model comparison

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
a3a	6	577575	577631	-288781			
a3b	11	566804	566907	-283391	10780.7892	5	< 2e-16***
a3c	12	566799	566912	-283388	6.4228	1	0.01127*
a3d	14	566801	566933	-283387	2.2402	2	0.32625

In all of these models, H1*-A3* values were averaged across all eleven timepoints. As shown in Table 32 above, the model a3c was the maximal model that was justified by our H1*-A3* data. However, we decided to use the model a3d because it better reflects what we observe in the plot shown in Figure 5–14. For example, using the model a3c there was a significant difference between the males' H1*-A3* for voiced and voiceless aspirated stops, but this effect was no longer significant when using the model a3d. This model includes Voicing, Gender and their interaction as the fixed effects. It also includes by-speaker and by-word random intercepts as well as by-speaker random slopes for Voicing as the random effects. We present the results of H1*-A3* analysis by considering H1*-A3* at vowel onset and midpoint, which were obtained by averaging timepoints 1-3 and timepoints 5-7 respectively.

5.7.3 Inferential statistics on H1*-A3* at vowel onset

Table 33. The output of a linear mixed-effects model of H1*-A3* at vowel onset. Voiced is the reference category for Voicing and female is for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	6.4770	0.9115	17.97	7.106	**< .0001**
VoicingVoiceless	-0.6659	1.3521	17.24	-0.492	0.6286
VoicingAspirated	2.8306	0.7316	25.99	3.869	**0.0007**
GenderMale	-3.1996	1.2748	14.99	-2.510	**0.0240**
VoicingVoiceless:GenderMale	-3.2385	1.9111	14.99	-1.695	0.1108
VoicingAspirated:GenderMale	-1.2769	0.9289	14.90	-1.375	0.1895

Table 33 shows the results of a linear mixed-effects analysis of H1*-A3* as a function of Voicing and Gender at vowel onset. Voiced is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment-coded in order to facilitate the comparisons between voiceless unaspirated and voiced stops as well as between voiceless aspirated and voiced stops.

As shown in Table 33, there was no significant difference in females' H1*-A3* values between voiceless unaspirated and voiced stops (*p* = 0.63). The result is not surprising as they overlap considerably in their H1*-H2* values (see Figure 5–14). In contrast, there was a significant difference in females' H1*-H2* values between voiceless aspirated and voiced stops at vowel onset (*p* = 0.001). The mean females' H1*-H2* for voiceless aspirated stops was estimated to be 9.31 dB at vowel onset.

Table 34. Results of post-hoc within-gender pairwise comparisons for H1*-A3* by voicing categories at vowel onset. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	2.8306	0.7670	27.10	3.690	**0.003**
	aspirated - voiceless	3.4964	1.5784	18.56	2.215	0.1138
	voiced - voiceless	0.6659	1.4415	19.15	0.462	0.9569
Male	aspirated - voiced	1.5536	0.8084	25.73	1.922	0.1846
	aspirated - voiceless	5.4579	1.6829	18.39	3.243	**0.0132**
	voiced - voiceless	3.9043	1.5354	18.89	2.543	**0.0586**

Table 34 shows the post-hoc within-gender pairwise comparisons by voicing at vowel onset. As we can see, there was a significant difference in H1*-A3* values between voiceless aspirated and voiced stops for females (*p* = 0.003), but there was no significant difference for males (*p* = 0.18). In contrast, the difference in H1*-A3* values between voiceless aspirated and voiceless unaspirated stops was not significant for females (*p* = 0.11), but it was significant for males (*p* = 0.013). Similarly, there was no significant difference in H1*-A3* values between voiced and voiceless unaspirated stops for females (*p* = 0.96), but there was a marginal significant difference for males (*p* = 0.058).

5.7.4 Inferential statistics on H1*-A3* at vowel midpoint

Table 35. The output of a linear mixed-effects model of H1*-H2* at vowel midpoint. Voiced is the reference category for Voicing and female is for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	5.9635	0.9571	17.27	6.231	< **.0001**
VoicingVoiceless	-0.8727	1.3969	16.78	-0.625	0.5405
VoicingAspirated	2.0573	0.5820	32.96	3.535	**0.0012**
GenderMale	-4.0340	1.3523	15.00	-2.983	**0.0093**
VoicingVoiceless:GenderMale	-2.1788	1.9882	15.00	-1.096	0.2904
VoicingAspirated:GenderMale	-0.2647	0.6932	14.89	-0.382	0.7080

Table 35 provides the output of a linear mixed-effects model of H1*-A3* as a function of Voicing and Gender at vowel midpoint. Similar to the model of H1*-A3* at vowel onset, voiced is the reference level for Voicing and female is the reference level for Gender. Voicing is also treatment coded for the same reason.

The results show that the difference in females' H1*-A3* values between voiceless unaspirated and voiced stops at vowel midpoint was not significant (*p* = 0.54). However, consistent with the result at vowel onset, there was a significant difference between females' H1*-A3* for voiceless aspirated and voiced stops at vowel midpoint (*p* = 0.001). The mean females' H1*-A3* value for voiceless aspirated stops was estimated to be 8 dB.

Table 36. Results of post-hoc within-gender pairwise comparisons for H1*-A3* by voicing categories at vowel midpoint. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	2.8307	0.7670	27.1	3.69	**0.003**
	aspirated - voiceless	3.4964	1.5784	18.56	2.215	0.1138
	voiced - voiceless	0.6659	1.4415	19.15	0.462	0.9569
Male	aspirated - voiced	1.5536	0.8084	25.73	1.922	0.1846
	aspirated - voiceless	5.4579	1.6829	18.39	3.243	**0.0132**
	voiced - voiceless	3.9043	1.5354	18.89	2.543	**0.0586**

Table 36 shows the post-hoc within-gender pairwise comparisons by Voicing at vowel midpoint. As can be seen, there was a significant difference in H1*-A3* values between voiceless aspirated and voiced stops for females (*p* = 0.003), but there was no significant difference for males (*p* = 0.18). In contrast, the difference in H1*-A3* values between voiceless aspirated and voiceless unaspirated stops was not significant for females (*p* = 0.11), but it was significant for males (*p* = 0.013). Similarly, there was no significant difference in H1*-A3* values between voiced and voiceless unaspirated stops for females (*p* = 0.96), but there was a significant difference for males (*p* = 0.05).

5.7.5 Summary and implication of results for H1*-A3*

We have presented the results of analysis for H1*-A3* as a function of Voicing and Gender at vowel onset and midpoint. We found that at vowel onset and midpoint there is a significant difference in H1*-A3* values between voiceless aspirated and voiced stops for females, but there is no significant difference in H1*-A3* values between voiceless aspirated and voiced stops for males. The difference in H1*-A3* values between voiceless aspirated and voiceless unaspirated at vowel onset and midpoint is not significant for females, but it turns out to be significant for males. Similarly, there is no significant difference in H1*-A3* values between voiced and voiceless unaspirated for females, but the difference is significant for males.

With respect to the question whether voiced and voiceless aspirated stops share phonetic properties to the exclusion of voiceless unaspirated stops, our findings show that they share this acoustic property of H1*-A3* for male speakers. Regarding the question if they are in line with the prediction that voiced and voiceless aspirated stops are either [+ATR] or [+LL], the findings suggest that they are. This is because both [+ATR] and [+LL] predict that vowels produced with either an advanced tongue root or a lowered larynx will be breathy, as indicated by lower H1*-A3* values.

It is not always clear why there is a gender-based difference in the realisation of this feature. That is, why only male speakers show a consistent patterning of voiced and voiceless aspirated stops while female speakers do not. However, if we look at other studies which examine voice quality contrasts using spectral measures, these differences in gender-related findings are in fact fairly common. For example, while H1-A3 distinguishes between phonation types in Santa Ana del Valle Zapotec only for males (Esposito, 2010b), it distinguishes between breathy and clear vowels in Chanthaburi Khmer only for females (Wayland & Jongman, 2001).

Wayland and Jongman (2001) argue that such a gender-specific difference may be related to the fact that females maintain ‘the historical breathy and clear phonation distinction in Khmer’, but such a contrast is probably disappearing in males, who realise it as a tense-lax contrast instead. However, they do not provide any reason why female speakers tend to be conservative to sound change such as this. However, a study by Abramson *et al.* (2007) on voice registers in Khmu' suggests that one possible reason why females are more conservative in this case is because they spend most of their time in their village while males tend to travel away and therefore become exposed to linguistic differences. Whether this can also account for what we observe in Madurese remains an open question, which will certainly become an interesting area for further investigation in future studies.

5.8 H1*-H2*

5.8.1 Descriptive statistics on H1*-H2*

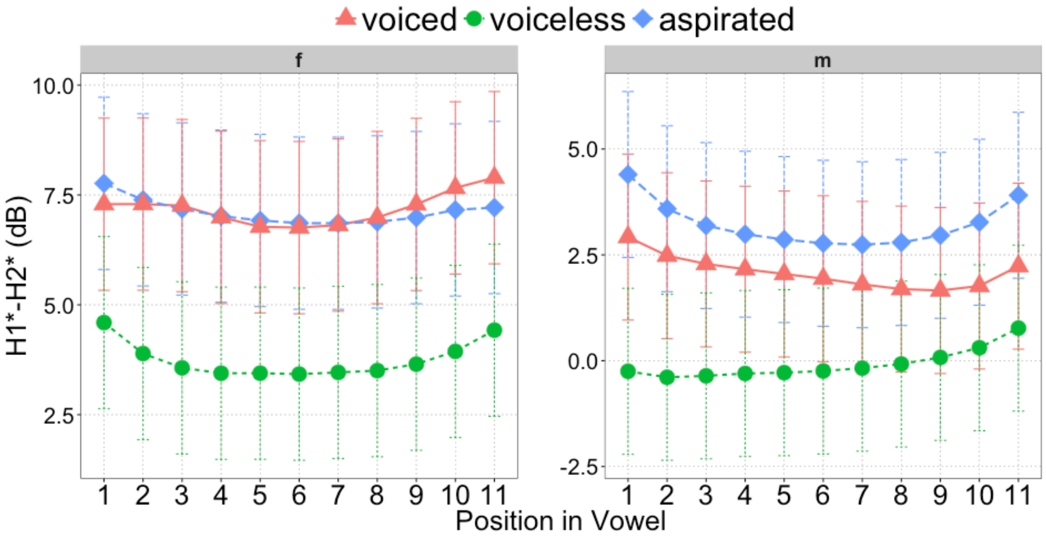


Figure 5–16. Mean H1*-H2* of vowels following voiced, voiceless aspirated and voiceless unaspirated stops measured at 11 equidistant timepoints; female on the left panel and male on the right panel. Error bars represent 95% confidence interval.

Figure 5–16 shows the mean H1*-H2* values for vowels following voiced, voiceless unaspirated and voiceless aspirated stops in Madurese. Here, H1*-H2* is plotted separately by gender in order to show whether both genders have a similar pattern in their H1*-H2*. It is also meant to demonstrate the extent to which they demonstrate similarities and/or differences on the basis of this acoustic parameter.

The figure shows that H1*-H2* appears to distinguish voiceless unaspirated stops from voiceless aspirated and voiced stops throughout the vowel timecourse for both genders. However, it does not seem to distinguish voiced and voiceless aspirated stops as indicated by a lot of overlaps in their H1*-H2* values. Females in particular show a complete overlap in H1*-H2* values for voiced and voiceless aspirated stops. In general, the pattern that emerges from this acoustic measure is that females show higher values of H1*-H2* regardless of voicing categories. More important is the fact that both genders appear to pattern together in this measure consistently.

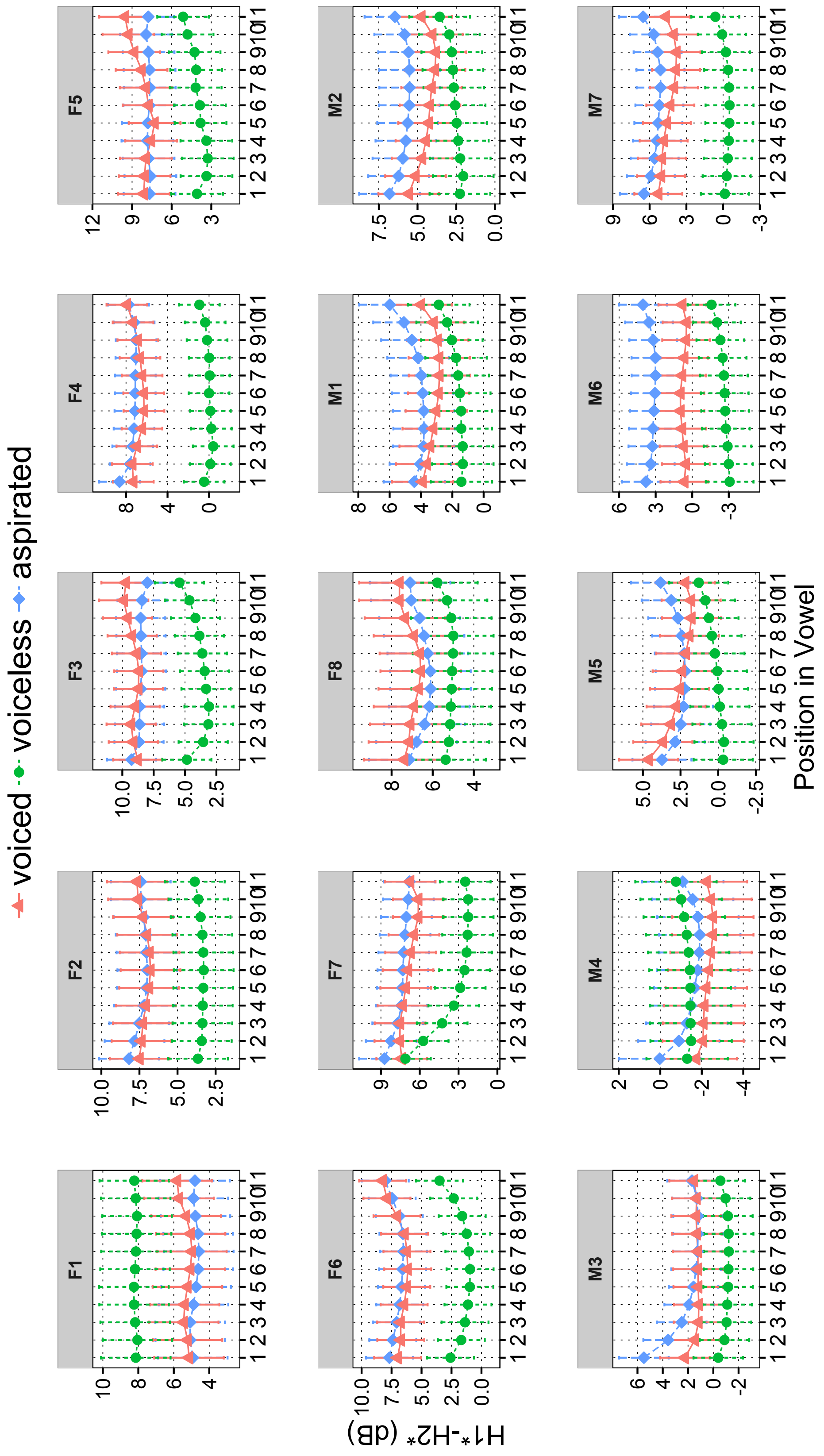


Figure 5–17. Individual speakers' H1*-H2* values of fifteen speakers. F stands for females and M stands for males.

Figure 5–17 shows individual variation in H1*-H2* realisation for 15 speakers of Madurese. There are eight female speakers and seven male speakers. In terms of H1*-H2* patterns, we can group the speakers into two. Group 1 consists of speakers where the H1*-H2* following voiced and voiceless aspirated stops pattern together excluding that following voiceless unaspirated stops. This group includes F2, F3, F4, F5, F6, M2, M5, M6 and M7. Group 2 consists of speakers where the H1*-H2* following the three stop categories is relatively similar. This includes F8, M1, M3, and M4. F1 is the only speaker who does not belong to any of the groups. This is because voiceless unaspirated stops have higher H1*-H2* than voiced and voiceless aspirated stops. The individual variations in H1*-H2* values may simply reflect individual differences in the phonetic implementation of the voicing categories.

5.8.2 Model comparison for H1*-H2*

In order to estimate the differences in H1*-H2* values for voiced, voiceless unaspirated and voiceless aspirated stops, we compared the following models:

h2a: H1*-H2* ~ Voicing + (1 | Speaker) + (1 | Word)
h2b: H1*-H2* ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
h2c: H1*-H2* ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 | Word)
h2d: H1*-H2* ~ Voicing * Gender + (1 + Voicing | Speaker) + (1 | Word)

Table 37. Log-likelihood results for H1*-H2* model comparison

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
h2a	6	476586	476642	-238287			
h2b	11	463433	463536	-231705	13163.183	5	< 2.2e-16***
h2c	12	463427	463540	-231701	7.911	1	0.004914**
h2d	14	463418	463550	-231695	13.018	2	0.001490**

H1*-H2* values were averaged across all eleven timepoints. As seen in Table 37, the result of the log-likelihood ratio test indicates that the model h2d was the maximal model justified by our H1*-H2* data. The model includes Voicing, Gender and the interaction term for Voicing and Gender as the fixed effects. It also includes by-speaker and by-word random intercepts and by-speaker random slopes for Voicing as the random effects. Here we present the results of H1*-H2* analysis by considering H1*-H2* in two regions of the vowel: at vowel onset (the average of timepoints 1-3) and vowel midpoint (the average of timepoints 5-7).

5.8.3 Inferential statistics on H1*-H2* at vowel onset

Table 38. The output of a linear mixed-effects model of H1*-H2* at vowel onset. Voiced is the reference category for Voicing and female is for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	7.2979	0.6494	15.89	11.237	< **.0001**
VoicingVoiceless	-3.3274	0.8617	15.86	-3.862	**0.0014**
VoicingAspirated	0.1441	0.3318	23.01	0.434	0.6682
GenderMale	-4.7330	0.9371	15.00	-5.051	**0.0001**
VoicingVoiceless:GenderMale	0.4536	1.2439	15.00	0.365	0.7205
VoicingAspirated:GenderMale	1.0153	0.4349	14.92	2.335	**0.0339**

Table 38 shows the results of a linear mixed-effects analysis of H1*-H2* as a function of Voicing and Gender at vowel onset. Voiced is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment-coded in order to facilitate the comparisons between voiceless unaspirated and voiced stops as well as between voiceless aspirated and voiced stops. As shown in Table 38, there was a significant difference between the mean females' H1*-H2* values for voiceless unaspirated and voiced stops ($p = 0.001$). The mean females' H1*-H2* for voiceless unaspirated stops was estimated to be 3.97 dB at vowel onset. The result is not surprising if we look at Figure 5–16, in which these two stop categories exhibit a well separation in their H1*-H2* values. However, there was no significant difference in females' H1*-H2* values for voiceless aspirated stops and voiced stops ($p = 0.66$). This is also indicated by a considerable overlap in H1*-H2* values for voiced and voiceless aspirated stops shown in Figure 5–16 above.

Table 39. Results of post-hoc within-gender pairwise comparisons for H1*-H2* by voicing categories at vowel onset. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	0.1441	0.3489	23.93	0.413	0.9683
	aspirated - voiceless	3.4715	0.9907	17.80	3.504	**0.0077**
	voiced - voiceless	3.3274	0.9226	17.97	3.607	**0.0061**
Male	aspirated - voiced	1.1594	0.3688	23.00	3.143	**0.0136**
	aspirated - voiceless	4.0332	1.0581	17.74	3.812	**0.0039**
	voiced - voiceless	2.8739	0.9849	17.88	2.918	**0.0274**

Table 39 above shows the post-hoc within-gender pairwise comparisons by voicing at vowel onset. As we can see, there was no significant difference in H1*-H2* values between voiceless aspirated and voiced stops for females ($p = 0.97$), but there was a significant difference for males (*p* = 0.014). In contrast, the difference in H1*-H2* values between voiceless aspirated and voiceless unaspirated stops was significant for both genders (females: *p* = 0.008, males: *p* = 0.004). Similarly, the difference in H1*-H2* values between voiced and voiceless unaspirated stops was also significant for both genders (females: *p* = 0.006, males: *p* = 0.027).

5.8.4 Inferential statistics on H1*-H2* at vowel midpoint

Table 40. The output of a linear mixed-effects model of H1*-H2* at vowel midpoint. Voiced is the reference category for Voicing and female is the reference category for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	6.7913	0.6049	17.85	11.226	**< .0001**
VoicingVoiceless	-3.3524	0.8614	17.35	-3.892	**0.0011**
VoicingAspirated	0.0879	0.3164	65.30	0.278	0.7821
GenderMale	-4.8585	0.8476	14.99	-5.732	**< .0001**
VoicingVoiceless:GenderMale	1.2157	1.2157	15.00	1.000	0.3331
VoicingAspirated:GenderMale	0.7875	0.3070	14.72	2.565	**0.0218**

Table 40 provides the output of a linear mixed-effects model of H1*-H2* as a function of Voicing and Gender at vowel midpoint. Similar to the model of H1*-H2* at vowel onset, voiced is the reference level for Voicing and female is the reference level for Gender. Voicing is also treatment coded for the same reason. The results show that the difference between the mean females' H1*-H2* values for voiceless unaspirated and voiced stops at vowel midpoint was significant (*p* = 0.001). The mean females' H1*-H2* value for voiceless unaspirated stops was estimated to be 3.44 dB. However, consistent with the result at vowel onset, there was no significant difference between females' H1*-H2* for voiceless aspirated and voiced stops at vowel midpoint (*p* = 0.78).

Table 41. Results of post-hoc within-gender pairwise comparisons for H1*-H2* by voicing categories at vowel midpoint. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	0.0386	0.3257	64.90	0.118	0.9992
	aspirated - voiceless	3.4271	1.0236	18.62	3.348	**0.0103**
	voiced - voiceless	3.3885	0.9174	19.36	3.694	**0.0045**
Male	aspirated - voiced	0.9351	0.3363	58.57	2.780	**0.0217**
	aspirated - voiceless	2.9726	1.0913	18.44	2.724	**0.0406**
	voiced - voiceless	2.0375	0.9769	19.07	2.086	0.1445

Table 41 shows the post-hoc within-gender pairwise comparisons by voicing at vowel midpoint. As we can see, there was no significant difference in H1*-H2* values between voiceless aspirated and voiced stops for females ($p = 0.99$), but there was a significant difference for males ($p = 0.022$). On the other hand, the difference in H1*-H2* values between voiceless aspirated and voiceless unaspirated stops was significant for both genders (females: $p = 0.01$, males: $p = 0.041$). While the difference in H1*-H2* values between voiced and voiceless unaspirated stops was significant for females ($p = 0.005$), it was not significant for males ($p = 0.145$).

5.8.5 Summary and implication of results for H1*-H2*

We have presented the results of a linear mixed-effects model of H1*-H2* as a function of Voicing and Gender. We found that there are no significant differences in H1*-H2* values between voiceless aspirated and voiced stops at vowel onset and midpoint for females but there are for males. However, there are significant differences in H1*-H2* values between voiceless aspirated and voiceless unaspirated stops at vowel onset and midpoint for both genders. The difference in H1*-H2* values between voiced and voiceless unaspirated stops at vowel onset and midpoint is significant for females, but for males it is only significant at vowel onset.

It is also not clear why there is a gender-based difference in the realisation of this feature, i.e. why only female speakers demonstrate a consistent patterning of voiced and voiceless aspirated stops to the exclusion of voiceless unaspirated stops while male speakers do not. However, if we look at related studies that look into voice quality differences, gender-related findings such as this one are also relatively common. For example, in her study on Santa Ana Del Valle Zapotec, Esposito (2010b) found that H1-H2 successfully distinguishes three phonation categories (breathy, modal and creaky) only for female speakers.

This finding is interesting because voiced and voiceless aspirated stops pattern together in their H1*-A3* for male speakers while for female speakers it is voiced and voiceless aspirated stops that pattern together in their H1*-H2*. Thus, there seems to be a different mechanism that may be used by females and males in the realisation of the feature. In response to the different ways in which females and males distinguish their phonation, Esposito (2010b) suggests that the successful uses of H1-H2 for females and H1-A3 for males may indicate that there is a difference in how phonation is realised between genders and this is probably associated with some physiological and sociolinguistic factors. This statement is also in agreement with what Blankenship (2002) shows in her study that spectral measures do not all succeed in distinguishing phonation types in some languages she has examined.

With regard to the question whether voiced and voiceless aspirated stops share phonetic properties to the exclusion of voiceless unaspirated stops, our findings indicate that they share the acoustic property of H1*-H2*. Furthermore, with respect to the question whether they are in line with the phonetic prediction that voiced and voiceless aspirated stops have the feature [+ATR] or [+LL], the findings also suggest that they are. This is because both [+ATR] and [+LL] predict that vowels produced with either an advanced tongue root or a lowered larynx are expected to be breathy, which is indicated by greater H1*-H2* values in this case.

5.9 H2*-H4*

5.9.1 Descriptive statistics on H2*-H4*

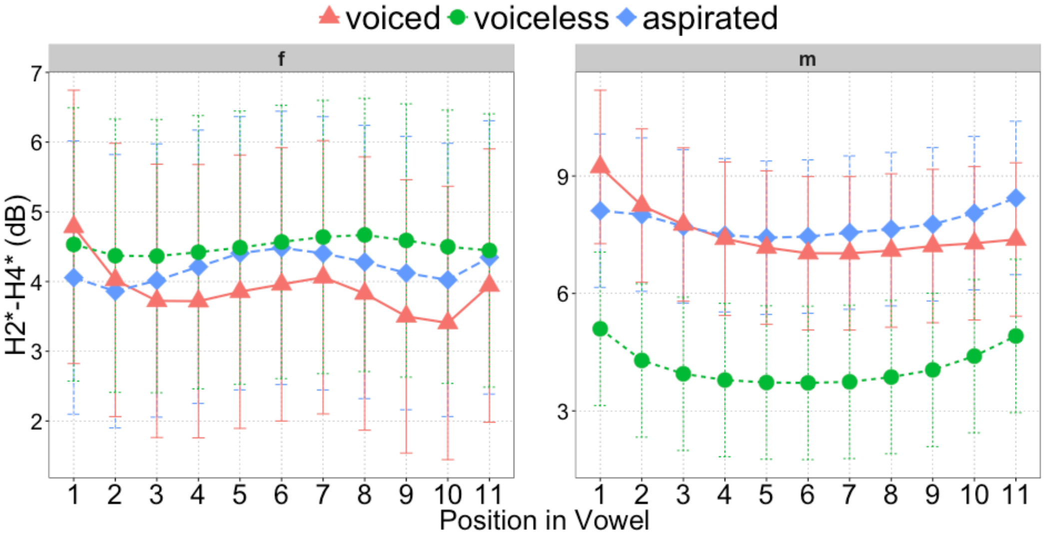


Figure 5–18. Mean H2*-H4* of vowels following voiced, voiceless unaspirated and voiceless aspirated measured at 11 equidistant timepoints; female on the left panel and male on the right panel. Error bars represent 95% confidence interval.

Figure 5–18 shows plots for mean H2*-H4* of vowels following voiced, voiceless unaspirated and voiceless aspirated stops measured at eleven equally spaced timepoints. As we can see, female and male speakers exhibit a very different pattern with regard to their H2*-H4* values. Specifically, female speakers do not seem to distinguish between voiced, voiceless unaspirated and voiceless aspirated stops as indicated by overlap in their H2*-H4* values. In contrast, male speakers seem to distinguish between the H2*-H4* values for voiceless unaspirated stops on the one hand and voiced and voiceless aspirated stops on the other. However, their H2*-H4* values for voiced and voiceless aspirated stops completely overlap with one another.

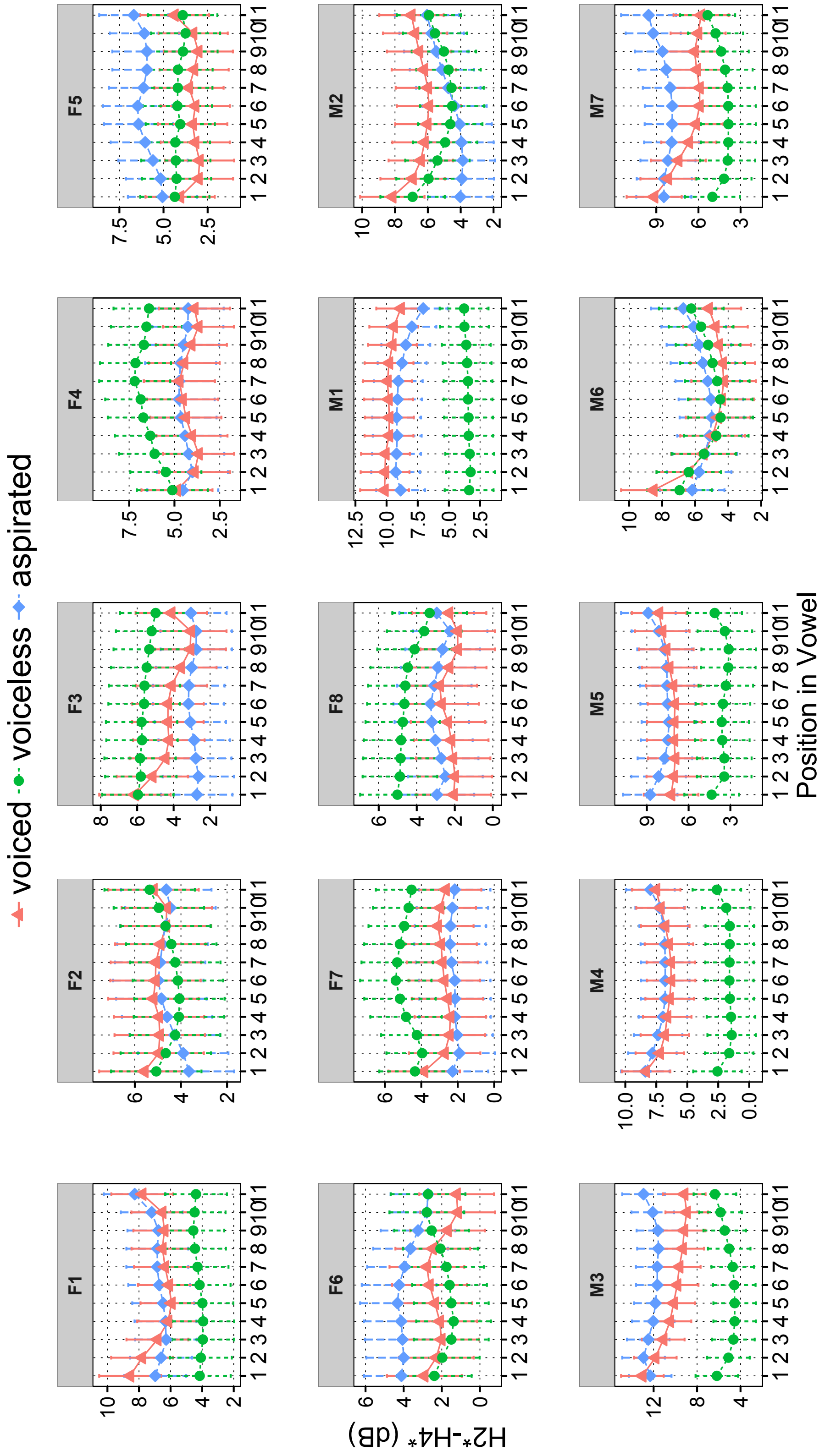


Figure 5–19. Individual speakers’ H2*-H4* values of fifteen speakers. F stands for females and M stands for males.

Figure 5–19 on the preceding page shows individual variation in H2*-H4* realisation for 15 speakers of Madurese. There are eight female speakers and seven male speakers. In terms of H2*-H4* patterns, we can classify them into three groups. Group 1 consists of speakers where the H2*-H4* following voiced and voiceless aspirated stops is higher than that following voiceless unaspirated stops. This group includes F1, M1, M3, M4, M5, and M7. Group 2 consists of speakers where the H2*-H4* following voiceless unaspirated stops is higher than that following voiced and voiceless aspirated stops. This group includes F4, F5, F6, F7 and F8. Group 3 consists of speakers where the H2*-H4* following the three categories does not seem to differ from one another. This group includes F2 and M6. However, F5 does not belong to any of the three groups due to the fact that voiceless unaspirated and voiced stops pattern together to the exclusion of voiceless aspirated stops.

As we can also see in Figure 5–19, female speakers contribute to a lot of variations. It does not come to a surprise that their whole pattern for this acoustic measure is rather unpredictable. This is not the case for male speakers; despite the fact that there is some individual variation, most of these speakers demonstrate a relatively consistent pattern as expected from this measure. We can see that voiced and voiceless aspirated stops pattern together in their H2*-H4* to the exclusion of voiceless unaspirated stops. See, for example, speakers M1, M3, M4, M5 and M7.

5.9.2 Model comparison for H2*-H4*

In order to estimate the differences between H2*-H4* values following each stop type, we compared the following linear mixed-effects models to find the maximal model justified by our data:

h4a: H2*-H4* ~ Voicing + (1 | Speaker) + (1 | Word)
h4b: H2*-H4* ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
h4c: H2*-H4* ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 | Word)
h4d: H2*-H4* ~ Voicing * Gender + (1 + Voicing | Speaker) + (1 | Word)

Table 42. Log-likelihood results for H2*-H4* model comparison

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
h4a	6	514664	514720	-257326			
h4b	11	502909	503012	-251443	11765.2478	5	< 2.2e-16***
h4c	12	502909	503022	-251442	1.6788	1	0.195086
h4d	14	502901	503033	-251437	11.6134	2	0.003007**

In all of these models, H2*-H4* values were averaged across all eleven timepoints. The test result shows that the model h4d was the maximal model justified by our H2*-H4* data. The model includes Voicing, Gender and their interaction as the fixed effects. It also includes by-speaker and by-word random intercepts as well as by-speaker random slopes for Voicing as the random effects.

In the following we present the results of H2*-H4* analysis by considering H2*-H4* at vowel onset and midpoint, which were obtained by averaging timepoints 1-3 and timepoints 5-7 respectively.

5.9.3 Inferential statistics on H2*-H4* at vowel onset

Table 43. The output of a linear mixed-effects model of H2*-H4* at vowel onset. Voiced is the reference level for Voicing and female is for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	4.1775	0.6456	16.832	6.471	**< .0001**
VoicingVoiceless	0.2899	0.7630	17.276	0.380	0.7087
VoicingAspirated	-0.2029	0.5382	20.514	-0.377	0.7101
GenderMale	4.2344	0.9181	14.996	4.612	**0.0004**
VoicingVoiceless:GenderMale	-4.2621	1.0779	14.994	-3.954	**0.0013**
VoicingAspirated:GenderMale	-0.2392	0.7272	14.948	-0.329	0.7467

Table 43 summarises the results of a linear mixed-effects analysis of H2*-H4* as a function of Voicing and Gender at vowel onset. Voiced is the reference level for Voicing and female is the reference level for Gender. Voicing is treatment-coded in order to facilitate the comparisons between voiceless unaspirated and voiced stops as well as between voiceless aspirated and voiced stops. As shown in Table 43 above, no significant differences in females' H2*-H4* values were found between voiceless unaspirated and voiced stops (p = 0.71) as well as between voiceless aspirated and voiced stops (p = 0.71), as expected from the plots in Figure 5–18.

Table 44. Results of post-hoc within-gender pairwise comparisons for H2*-H4* by voicing categories at vowel onset. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	-0.2029	0.5688	21.82	-0.357	0.9792
	aspirated - voiceless	-0.4927	1.0288	18.22	-0.479	0.9524
	voiced - voiceless	-0.2899	0.8133	19.15	-0.356	0.9793
Male	aspirated - voiced	-0.4421	0.6031	21.18	-0.733	0.8524
	aspirated - voiceless	3.5301	1.0977	18.09	3.216	**0.0142**
	voiced - voiceless	3.9722	0.8663	18.9	4.585	**0.0006**

Table 44 shows the post-hoc within-gender pairwise comparisons by voicing at vowel onset. As we can see, there was no significant difference in H2*-H4* values between voiceless aspirated and voiced stops for either gender (females: $p = 0.98$, males: $p = 0.85$). While the difference in H2*-H4* values between voiceless aspirated and voiceless unaspirated stops was not significant for females ($p = 0.95$), the difference turned out to be significant for males ($p = 0.014$). Similarly, the difference in H2*-H4* values between voiced and voiceless unaspirated stops was not significant for females ($p = 98$), but it was significant for males ($p = 0.001$).

5.9.4 Inferential statistics on H2*-H4* at vowel midpoint

Table 45. The output of a linear mixed-effects model of H2*-H4* at vowel midpoint. Voiced is the reference level for Voicing and female is for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	3.9543	0.5696	16.169	6.942	**< .0001**
VoicingVoiceless	0.6382	0.6864	16.371	0.93	0.3661
VoicingAspirated	0.4719	0.4379	18.979	1.078	0.2948
GenderMale	3.1229	0.8183	15.002	3.817	**0.0017**
VoicingVoiceless:GenderMale	-3.9569	0.9829	14.996	-4.026	**0.0011**
VoicingAspirated:GenderMale	-0.0506	0.6034	14.952	-0.084	0.9343

As shown in Table 45 above, no significant differences in females' H2*-H4* values were found between voiceless unaspirated and voiced stops ($p = 0.37$) as well as between voiceless aspirated and voiced stops ($p = 0.29$) at vowel midpoint, as also expected from the plots in Figure 5–18 shown earlier.

Table 46. Results of post-hoc within-gender pairwise comparisons for H2*-H4* by voicing categories at vowel midpoint. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	0.4676	0.4571	20.52	1.023	0.6832
	aspirated - voiceless	-0.2296	0.8812	17.91	-0.261	0.9917
	voiced - voiceless	-0.6972	0.7294	18.36	-0.956	0.7273
Male	aspirated - voiced	0.4543	0.4855	20.07	0.936	0.7385
	aspirated - voiceless	3.7448	0.9409	17.84	3.980	**0.0027**
	voiced - voiceless	3.2905	0.7781	18.22	4.229	**0.0015**

Table 46 shows the post-hoc within-gender pairwise comparisons by voicing at vowel midpoint. As we can see, the difference in H2*-H4* values between voiceless aspirated and voiced stops was not significant for either gender (females: $p = 0.68$, males: $p = 0.74$). While there was no significant difference in H2*-H4* values between voiceless aspirated and voiceless unaspirated stops for females ($p = 0.99$), there was a significant difference for males ($p = 0.003$). Likewise, the difference in H2*-H4* values between voiced and voiceless unaspirated stops was not significant for females ($p = 0.73$), but it was significant for males ($p = 0.002$).

5.9.5 Summary and implication of results for H2*-H4*

We have presented the results of a linear mixed-effects model of H2*-H4* as a function of Voicing and Gender. We found that there are no significant differences in H2*-H4* values between voiceless aspirated and voiced stops at vowel onset and midpoint for either gender. While the differences in H2*-H4* between voiceless unaspirated and voiceless aspirated stops as well as between voiced and voiceless unaspirated stops at vowel onset and midpoint are not significant for females, they are for males. Thus, similar to the results for H1*-A3* and H1*-H2* discussed earlier, we also found variation by gender for H2*-H4*.

The only consistency that we can see from the results for female and male speakers is that their H2*-H4* does not distinguish between voiced and voiceless aspirated stops, suggesting the two stop categories share this acoustic property. However, the results for female speakers in particular need to be considered with care due to variations within these speakers themselves (see Figure 5–19 on page 149).

With regard to the question whether voiced and voiceless aspirated stops share phonetic properties to the exclusion of voiceless unaspirated stops, our findings show that they share the acoustic property of H2*-H4*. Furthermore, with respect to whether the results for H2*-H4* are in line with the prediction that voiced and voiceless aspirated stops have either the feature [+ATR] or [+LL], the findings also suggest that they are. This is because both [+ATR] and [+LL] predict that vowels produced with an advanced tongue root or a lowered larynx will be breathy, and here it is indicated by greater H2*-H4* values.

5.10 Cepstral Peak Prominence (CPP)

Cepstral peak prominence (CPP) is a measure of periodicity of the source spectrum. In this case, breathy phonation is expected to have lower CPP values than modal phonation. Unlike the other voice quality measures, the results for CPP are particularly important in the context of Madurese given that this measure does not require F0 analysis and therefore theoretically it is not correlated with vowel height.

5.10.1 Descriptive statistics on CPP

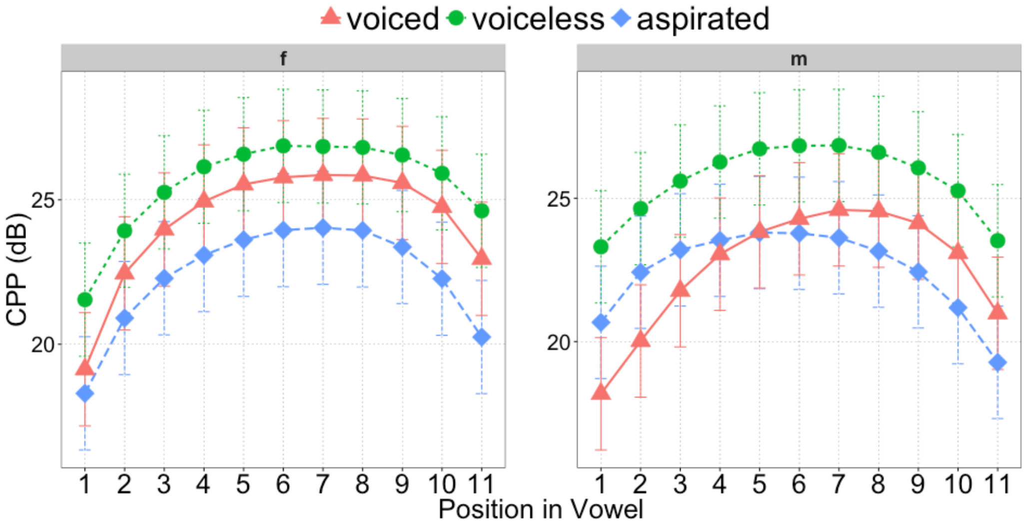


Figure 5–20. Mean CPP of vowels following voiced, voiceless unaspirated and voiceless aspirated measured at 11 equidistant timepoints; female on the left panel and male on the right panel. Error bars represent 95% confidence interval.

Figure 5–20 above shows the mean CPP values for female and male speakers averaged across speakers, places of articulation and repetitions and measured at eleven equidistant timepoints into the vowel. As we can see, female and male speakers exhibit a relatively different pattern in their CPP values. Specifically, females' CPP values for voiceless aspirated stops look consistently lower than those for voiced and voiceless unaspirated stops respectively. In contrast, males' CPP values for voiced stops are lower than those for voiceless aspirated and unaspirated stops respectively, particularly if we look at the first three timepoints. However, the pattern changes at the last three timepoints into the vowel offset where it is males' CPP values for voiceless aspirated stops that appear to be lower than those for voiced and voiceless unaspirated stops respectively. Furthermore, female and male speakers also show some differences in terms of which voicing categories pattern together in this measure; for females, it is voiced and voiceless unaspirated stops while for males it is voiced and voiceless aspirated stops.

Furthermore, if we examine the CPP plot for each speaker in Figure 5–21 on the following page, it appears that the CPP patterns cannot be based on gender such as the one shown in Figure 5–20 earlier. This is due to the fact that there are a number of cases where we can also observe that some females share similar patterns with males as well. For example, four female speakers F1, F2, F6 and F8 pattern with six male speakers M1, M3, M4, M5, M6 and M7. Similarly, three female speakers F3, F4 and F7 have similar patterns with one male speaker M2. This suggests that the CPP patterns shown in Figure 5–20 do not show the whole picture, but rather it may result from averaging the values for either gender. However, it is important to note that, as we will see later, we include gender as a variable in the model because it was justified by the data and it may not be related to this trajectory pattern.

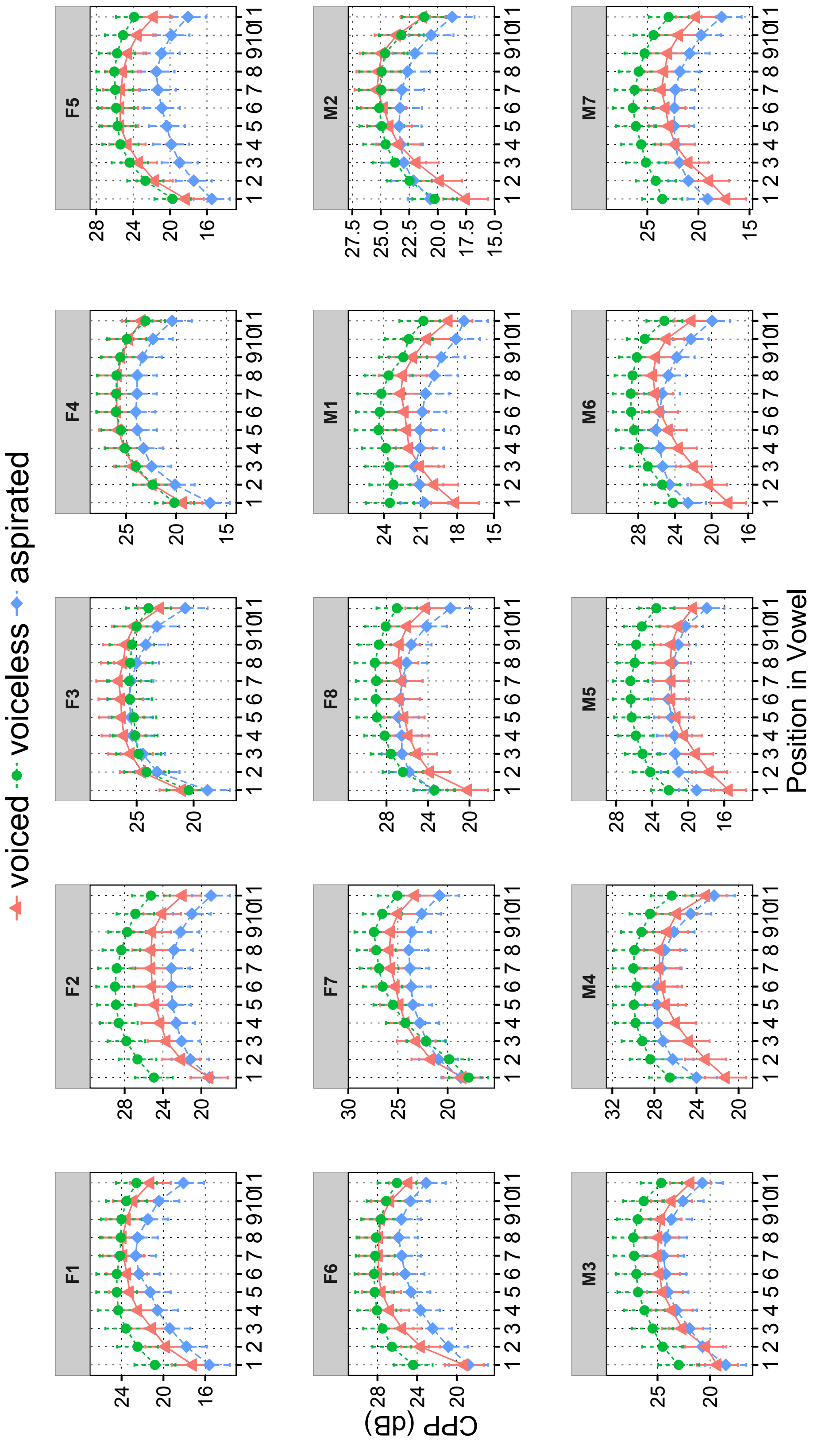


Figure 5–21. Individual speakers' CPP values of fifteen speakers. F stands for females and M stands for males.

Figure 5–21 on the preceding page shows individual variations in CPP realisation for 15 speakers of Madurese. There are eight female speakers and seven male speakers. In terms of CPP patterns, we can classify the speakers into two groups. Group 1 consists of speakers where the CPP following voiced and voiceless aspirated stops is lower than that following voiceless unaspirated stops. This group includes F1, F2, F6, F8, M1, M3, M4, M5, M6 and M7. Group 2 consists of speakers where the CPP following the three categories does not seem to differ from one another. This group includes F3, F4, F7 and M2. Speaker F5 is the only speaker who does not belong to any of the groups because voiced and voiceless unaspirated stops pattern together to the exclusion of voiceless aspirated stops.

5.10.2 Model comparison for CPP

In order to estimate the differences in CPP values for voiced, voiceless unaspirated and voiceless aspirated stops, we compared the following linear mixed-effects models to find the maximal model justified by our data:

cp1: CPP ~ Voicing + (1 | Speaker) + (1 | Word)
cp2: CPP ~ Voicing + (1 + Voicing | Speaker) + (1 | Word)
cp3: CPP ~ Voicing + Gender + (1 + Voicing | Speaker) + (1 | Word)
cp4: CPP ~ Voicing * Gender + (1 + Voicing | Speaker) + (1 | Word)

Table 47. Log-likelihood results for CPP model comparison

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
cp1	6	492418	492474	-246203			
cp2	11	488474	488578	-244226	3953.3898	5	< 2.2e-16***
cp3	12	488472	488585	-244224	4.342	1	0.037183*
cp4	14	488466	488598	-244219	9.7295	2	0.007714**

In all of these models, CPP values were averaged across all eleven timepoints. The result of the test shows that the model cp4 was the maximal model justified by our CPP data. The model includes Voicing, Gender and their interaction as the fixed effects. It also includes by-speaker and by-word random intercepts as well as by-speaker random slopes for Voicing as the random effects.

In the following we present the results of CPP analysis by considering CPP in two regions of the vowel: at vowel onset (the average of timepoints 1-3) and vowel midpoint (the average of timepoints 5-7).

5.10.3 Inferential statistics on CPP at vowel onset

Table 48. The output of a linear mixed-effects model of CPP at vowel onset. Voiced is the reference level for Voicing and female is for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	21.8439	0.5513	20.95	39.623	**< .0001**
VoicingVoiceless	1.6434	0.6659	22.24	2.468	**0.0218**
VoicingAspirated	-1.3525	0.6088	25.18	-2.221	**0.0355**
GenderMale	-1.8540	0.7412	14.96	-2.501	**0.0245**
VoicingVoiceless:GenderMale	2.7901	0.8813	14.93	3.166	**0.0064**
VoicingAspirated:GenderMale	3.4162	0.7799	14.89	4.380	**0.0005**

Table 48 summarises the results of a linear mixed-effects analysis of CPP as a function of Voicing and Gender at vowel onset. Voiced is the reference level for Voicing and female is the reference level for Gender. Table 48 shows that there was a significant difference between the mean females' CPP values for voiceless unaspirated and voiced stops at vowel onset ($p = 0.022$). The difference in females' CPP values for voiceless aspirated and voiced stops at vowel onset also turned out to be significant ($p = 0.04$).

Table 49. Results of post-hoc within-gender pairwise comparisons for CPP by voicing categories at vowel onset. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	-1.3525	0.6386	26.08	-2.118	0.1259
	aspirated - voiceless	-2.9959	0.7702	21.26	-3.890	**0.0025**
	voiced - voiceless	-1.6434	0.7015	23.33	-2.343	0.0818
Male	aspirated - voiced	2.0637	0.6735	24.84	3.064	**0.0155**
	aspirated - voiceless	-2.3699	0.8171	20.71	-2.910	**0.0257**
	voiced - voiceless	-4.4335	0.7422	22.48	-5.973	**< .0001**

Table 49 shows the post-hoc within-gender pairwise comparisons by voicing at vowel onset. As we can see, there was no significant difference in CPP values between voiceless aspirated and voiced stops for females ($p = 0.13$), but there was a significant difference for males ($p = 0.016$). However, the difference in CPP values between voiceless aspirated and voiceless unaspirated stops was significant for both genders (females: p = 0.003, males: p = 0.026). Moreover, while the difference in CPP values between voiced and voiceless unaspirated stops did not reach statistical significance for females (p = 0.082), it was highly significant for males (p < .0001).

5.10.4 Inferential statistics on CPP at vowel midpoint

Table 50. The output of a linear mixed-effects model of CPP at vowel midpoint. Voiced is the reference level for Voicing and female is the reference level for Gender.

	Estimate	Std. Error	d.f.	*t*-value	*p*-value
(Intercept)	25.7282	0.6046	24.98	42.554	**< .0001**
VoicingVoiceless	0.9529	0.6102	37.54	1.562	0.1268
VoicingAspirated	-1.8901	0.5497	55.42	-3.438	**0.0011**
GenderMale	-1.5012	0.7769	14.95	-1.932	0.0725
VoicingVoiceless:GenderMale	1.4055	0.7012	14.75	2.004	0.0637
VoicingAspirated:GenderMale	1.3110	0.5649	14.68	2.321	**0.0351**

Table 50 summarises the results of a linear mixed-effects analysis of CPP as a function of Voicing and Gender at vowel midpoint. Voiced is the reference level for Voicing and female is the reference level for Gender. As Table 50 shows, the difference between the mean females' CPP values for voiceless unaspirated and voiced stops at vowel midpoint was not significant (p = 0.13). In contrast, there was a significant difference in females' CPP values for voiceless aspirated and voiced stops at vowel midpoint (p = 0.001).

Table 51. Results of post-hoc within-gender pairwise comparisons for CPP by voicing categories at vowel midpoint. *P* values are adjusted based on the Sidak method for 3 tests.

Gender	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
Female	aspirated - voiced	-1.9028	0.5465	63.78	-3.482	**0.0027**
	aspirated - voiceless	-2.8304	0.6112	36.36	-4.631	**0.0001**
	voiced - voiceless	-0.9276	0.6199	39.58	-1.496	0.3696
Male	aspirated - voiced	-0.8011	0.5647	57.33	-1.419	0.4103
	aspirated - voiceless	-3.0219	0.6389	33.67	-4.729	**0.0001**
	voiced - voiceless	-2.2208	0.6469	36.42	-3.433	**0.0045**

Table 51 shows the post-hoc within-gender pairwise comparisons by voicing at vowel midpoint. The table shows that there was a significant difference in CPP values between voiceless aspirated and voiced stops for females (p = 0.003), but there was no significant difference for males (p = 0.41). In contrast, the difference in CPP values between voiceless aspirated and voiceless unaspirated stops was significant for both genders (females: *p* = 0.0001, males: *p* = 0.0001). Moreover, the difference in CPP values between voiced and voiceless unaspirated stops was not significant for females (*p* = 0.37), but it was significant for males (*p* = 0.005).

5.10.5 Summary and implication of results for CPP

We have presented the results of a linear mixed-effects model of CPP as a function of Voicing and Gender. We found that the differences in CPP values between voiceless aspirated and voiced stops as well as between voiced and voiceless unaspirated stops at vowel onset are not significant for females, but they are significant for males. While the difference in CPP values between voiceless aspirated and voiced stops at vowel midpoint is significant for females, it is not significant for males. In contrast, the difference in CPP values between voiced and voiceless unaspirated stops at vowel midpoint is not significant for females, but it is for males. However, the differences in CPP values between voiceless aspirated and voiceless unaspirated stops at vowel onset and midpoint are significant for both genders.

With respect to how these gender-related findings may happen, no studies on voice quality show gender-specific differences in CPP. It is true that females may have lower CPP values than males, suggesting that they produce speech with more high-frequency aperiodic components for some reason. However, CPP has previously been found to successfully distinguish a set of phonation type regardless of gender. For example, Garellek and Keating (2011) found that CPP distinguishes between breathy and modal phonation in Jalapa Mazatec for both genders.

With regard to the question whether voiced and voiceless aspirated stops share phonetic properties to the exclusion of voiceless unaspirated stops, our findings indicate that to some extent they share the acoustic property of CPP. Furthermore, with respect to whether the results for CPP are in line with the prediction that voiced and voiceless aspirated stops have the feature [+ATR] or [+LL], the findings also suggest that they are. This is because both [+ATR] and [+LL] predict that vowels produced with either an advanced tongue root or a lowered larynx will be breathy, which can be indicated by lower CPP values. It is important to note, however, that due to the fact the results for CPP depend on gender as well as on where in the vowel the measurement was taken, this acoustic property cannot be considered as a strong acoustic correlate for either an advanced tongue root or a lowered larynx here.

5.10.6 Interim summary

The three stop types in Madurese show a three-way distinction in VOT values but voiceless (unaspirated and aspirated) stops show a lot of overlap. F0 also suggests a two-way patterning (voiced versus voiceless stops). The voice quality measures are more variable in terms of patterning; H1*-A1* and H1*-H2* distinguish voiceless unaspirated stops from voiced and voiceless aspirated stops for both genders while H1*-A3* and H2*-H4* distinguish voiceless unaspirated stops from voiced and voiceless aspirated stops only for male speakers. CPP distinguishes voiceless unaspirated stops from voiced and voiceless aspirated stops at vowel onset for female speakers but at vowel midpoint for male speakers. Overall, the three stop types also fall into two categories by voice quality measures.

5.11 Linear Discriminant Analysis (LDA) for Spectral Measures

In addition, we carried out a linear discriminant analysis based on a number of acoustic variables, namely F0, H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, CPP and VOT, as predictors. This is done in order to assess how well voiced, voiceless unaspirated and voiceless aspirated stops may be separated from one another on the basis of these acoustic measures or how much each of these variables may contribute to the distinction between the three voicing categories. For the sake of brevity, we only looked at vowel onset, i.e. the average for timepoints 1-3 and midpoint, i.e. the average for timepoints 5-6.

There are two models we compared in this analysis. In the first model, we tried to predict voicing categories, i.e. voiced, voiceless unaspirated and voiceless aspirated based on seven spectral measures, i.e. F0, H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* and CPP. In the second model, we tried to predict voicing categories, i.e. voiced, voiceless unaspirated and voiceless aspirated on the basis of these seven spectral measures plus VOT. The reasons why we tested these two models are to establish whether spectral measures contribute to the distinction of the three voicing categories and see whether the relative weight of the predictors is different if VOT is in the predictor list. It is important to remember that the dependent variable in an LDA analysis is categorical, which is in this case Voicing with three levels, i.e. voiced, voiceless unaspirated and voiceless aspirated.

5.11.1 Result of linear discriminant analysis at vowel onset

Table 52. Coefficients of linear discriminants at vowel onset for the model without VOT

Predictors	LD1	LD2
H1*-H2*	-0.2541561	-0.0182456
H2*-H4*	-0.1263951	-0.0399309
CPP	0.0832555	0.0636252
H1*-A2*	0.0302428	0.1393462
H1*-A3*	0.0216505	0.0184105
H1*-A1*	-0.0134025	-0.0006468
F0	0.0077599	0.0054577

Table 52 above shows the coefficients of linear discriminants for each predictor. The predictors under the first discriminant function (LD1) can be ordered in terms of how well they discriminate voiced, voiceless unaspirated and voiceless aspirated stops. The higher the values the better the variable predicts the voicing categories. As we can see, H1*-H2* (-0.254) appears to be the most highly weighted predictor, followed by H2*-H4* (-0.126), CPP (0.083), H1*-A2* (0.03), H1*-A3* (0.022), H1*-A1* (-0.013) and F0 (0.008). The proportion of trace, which is the proportion of between-group variance that is explained by discriminant functions, for LD1 is 0.86 and for LD2 is 0.14. The fact that the proportion of trace for LD2 is fairly high suggests that it also contributes some share in discriminating the three voicing categories. It appears that H1*-A2* (0.139) is the most weighted predictor.

We also checked the Wilk's lambda in order to find out the total proportion of unexplained variance, using the *manova* function. The result shows that there is a lot of variance that goes unexplained by the model, i.e. about 70 percent. However, the fact that the LD model that only includes spectral predictors can account for about 30 percent of the total variance suggests that they contribute something to distinguishing the categories. We also assessed the accuracy of the model's prediction. The result shows that the accuracy of the model in predicting the voicing types in Madurese is around 57 percent. The most likely explanation for why there is a lot of unexplained variation by this model is probably because VOT is not included in the model.

To figure out that this was the case, we carried out another LDA for vowel onset using the model, which includes both spectral measures and VOT. The coefficients of linear discriminants for all predictors are shown in Table 53 below.

Table 53. Coefficients of linear discriminants at vowel onset for the model with VOT

Predictors	LD1	LD2
H1*-H2*	0.0755533	-0.2394358
VOT	-0.0470321	-0.0053636
H1*-A2*	-0.0350812	0.0001014
CPP	-0.0335805	0.0667459
H2*-H4*	0.0281638	-0.1144307
H1*-A1*	0.0150574	-0.0114219
H1*-A3*	-0.0098657	0.0168825
F0	-0.0006369	0.0065929

As we can see in Table 53, there is a change in the weight of the predictors when VOT is included. In this instance, H1*-H2* (0.076) appears to be the most weighted predictor, followed by VOT (-0.047), H1*-A2* (0.035), CPP (-0.034), H2*-H4* (0.028), H1*-A1* (0.015), H1*-A3* (-0.0099) and F0 (-0.0006). The proportion of trace, which is the proportion of between-group variance that is explained by discriminant functions, for LD1 is 0.93 and for LD2 is 0.07. The fact that the proportion of trace for LD2 is relatively high suggests that it also contributes some share in discriminating the three voicing categories. It appears that H1*-H2* (0.239) is the most highly weighted predictor.

The Wilks' lambda also shows that there is relatively less variation unexplained by the model, i.e. about 16 percent. Thus, the LDA model with spectral measures and VOT can explain about 84 percent of the total variance while the model without VOT can only explain 30 percent. This suggests that VOT does most work in discriminating the three voicing types in Madurese. We also assessed the model's prediction accuracy and found that it reaches 82 percent accuracy, which is high.

5.11.2 Results of linear discriminant analysis at vowel midpoint

Table 54. Coefficients of linear discriminants at vowel midpoint for the model without VOT

Predictors	LD1	LD2
H1*-H2*	-0.2347047	-0.1402759
H2*-H4*	-0.1499029	-0.0668796
CPP	0.0849794	-0.0955603
H1*-A3*	0.0232962	0.0213405
H1*-A1*	-0.0207565	-0.0265127
H1*-A2*	0.0193718	0.1220839
F0	0.0043312	0.0113026

Table 54 above shows the coefficients of linear discriminants for each predictor. The predictors under the first discriminant function (LD1) can be ordered in terms of how well they can discriminate voiced, voiceless unaspirated and voiceless aspirated stops. In this case, the higher the values the better the variable predicts the voicing categories. Like at vowel onset, H1*-H2* (-0.235) also appears to be the most highly weighted predictor, followed by H2*-H4* (-0.149), CPP (0.085), H1*-A3* (0.023), H1*-A1* (-0.021), H1*-A2* (0.019) and F0 (0.004). The proportion of trace for LD1 and LD2 at vowel midpoint is 0.92 and 0.08 respectively.

Wilk's lambda was also checked in order to know the total proportion of unexplained variance, using the manova function. The result shows that there is a lot of variance unexplained by the model, i.e. about 75 percent. However, the fact that the LD model that only includes spectral predictors can account for about 25 percent of the variance is also not trivial. Furthermore, we also assessed the model's prediction accuracy. The result shows that the accuracy of the model in predicting the voicing categories in Madurese is around 54 percent. The fact that less variance is explained at midpoint as opposed to at vowel onset suggests that this is due to the consonantal effects as vowel onset is clearly closer to the preceding stops.

We also conducted another LDA for vowel midpoint using the model, which includes both spectral measures and VOT. This is to confirm what contribution VOT may give in discriminating the three voicing types. The coefficients of linear discriminants for all predictors are shown in Table 55 on the following page.

Table 55. Coefficients of linear discriminants at vowel midpoint for the model with VOT

Predictors	LD1	LD2
H1*-H2*	0.0731434	-0.2195714
VOT	-0.0496487	-0.0030945
H2*-H4*	0.0268125	-0.1431597
H1*-A2*	-0.0225298	0.0089123
H1*-A1*	0.0098631	-0.0181621
CPP	-0.0047308	0.0916925
H1*-A3*	-0.0027867	0.0215237
F0	-0.0001632	0.0034814

As we can see in Table 55, there is a change in the weight of the predictors when we include VOT as a predictor. In this instance, H1*-H2* (0.073) appears to be the most weighted predictor, followed by VOT (-0.049), H2*-H4* (0.027), H1*-A2* (0.023), H1*-A1* (0.009), CPP (-0.005), H1*-A3* (-0.0028) and F0 (-0.0002). The proportion of trace, which is the proportion of between-group variance that is explained by discriminant functions, for LD1 is 0.94 and for LD2 is 0.06. The fact that the proportion of trace for LD2 is high suggests that it also contributes some share in discriminating the three voicing categories. It appears that H1*-H2* (-0.219) is the most highly weighted predictor for LD2.

The Wilks' lambda also shows that there is relatively less variation unexplained by the model, about 15 percent. Thus, the LDA model with spectral measures and VOT can explain around 85 percent of the total variance while the one without VOT can only explain 25 percent. This indicates that VOT shares most work to the voicing discrimination in Madurese. In fact, the accuracy of the model in predicting the voicing types is 83 percent, which is very high.

5.11.3 Conclusion and implication

We have conducted a linear discriminant analysis in order to evaluate how well spectral measures, i.e. F0, H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* and CPP and VOT discriminate voicing categories in Madurese. For that purpose, we built two models of LDA, one that only includes spectral measures and the other that includes both spectral measures and VOT. Our results indicate that the model that includes both spectral measures and VOT can explain most of the total variance

(84%) as opposed to the one without VOT (30%). In addition, the accuracy of the model with VOT is 84 percent compared to the one without VOT (57%).

These results are interesting because they suggest that spectral properties are also important in distinguishing the three voicing categories in Madurese. It is true that their contribution is not as high as that of VOT. Moreover, the fact that H1*-H2*, which is a correlate of the open quotient, is the most robust predictor in all models suggests that it is the most salient acoustic correlate of the laryngeal contrast following VOT in Madurese.

5.12 First and Second Formant Frequencies (F1 and F2)

5.12.1 Descriptive statistics on F1 and F2

As it is well-known that there is a robust CV co-occurrence restriction in Madurese, it becomes crucial to examine not only the consonants but also the vowels. One way to do that is by examining the first and the second formant frequencies (F1 and F2) of Madurese vowels to see the phonetic realisations of the two formant frequencies particularly with respect to the preceding stop consonants. The results of these analyses will be used to later assess the proposal which suggests that high and non-high vowels in Madurese could also be described in terms of ATR distinction, whereby high vowels are proposed as [+ATR] while non-high vowels are proposed as [-ATR]. Recall that [+ATR] vowels predict that they have lower F1 values in comparison with their [-ATR] counterparts. Front vowels with [+ATR] also predict that they have higher F2 values compared to their [-ATR] counterparts while back vowels with [+ATR] have lower F2 values than those with [-ATR].

In addition, looking at the F1 and F2 values would also provide a more definitive description for each of the vowel pair of high and non-high vowels, i.e. how they look like in the vowel space. This is important since scholars have some disagreement about the phonetic and phonological status of certain Madurese vowels (see Anderson, 1991; Cohn, 1993a; Davies, 2010; Stevens, 1968).

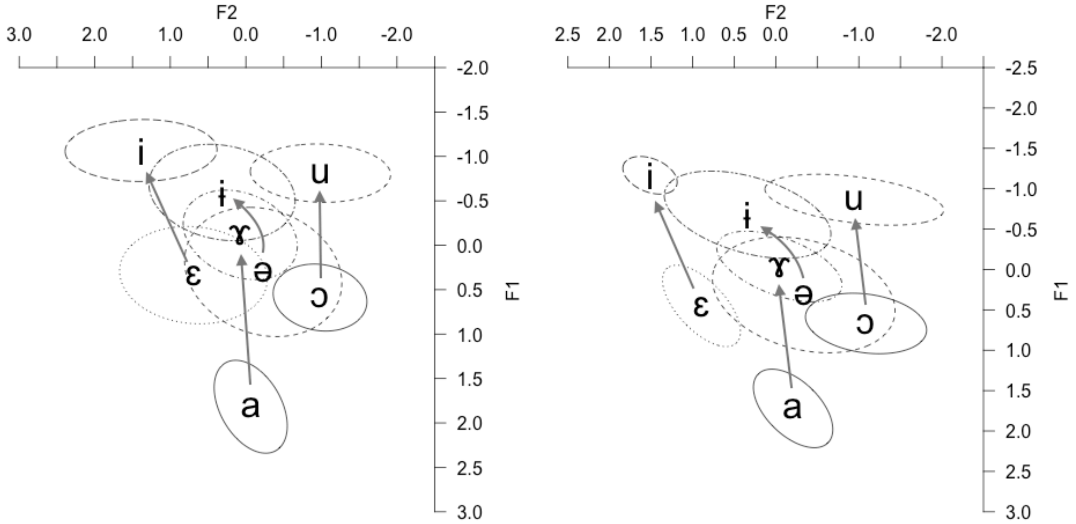


Figure 5–22. Distribution of vowels averaged over the vowel timecourse in a z-normalised F1 X F2 space with data from female on the left panel and male on the right panel. The arrows indicate the pair of non-high and high vowels.

Figure 5–22 shows the acoustic space of the eight surface vowels of Madurese and illustrates in particular the differences between the pairs of high and non-high vowels (i ~ ɛ, ɨ ~ ə, ɤ ~ a, u ~ ɔ) pooled across speakers, places of articulation and repetitions. The data come from female and male speakers plotted separately and F1 and F2 were sampled over the course of the vowels. The vertical axis stands for the first formant frequency while the horizontal axis represents the second formant frequency. All the values have been normalised using z-transformation. The ellipses indicate one standard deviation away from the mean and each ellipse contains approximately 68.27% of the data points.

As is shown in Figure 5–22, there are several instances of overlap in the F1 and F2 values in some vowels for both male and female speakers. For example, considerable overlapping F1 and F2 values can be seen in the central vowels [ə], [ɨ] and [ɤ] particularly for female speakers, and they considerably overlap in the vowels [ə] and [ɤ] for both genders. Furthermore, if we look at individual speakers, we will observe a lot of variations as well. For example, some of the ranges of variation can be seen in Figure 5–23 on the following page, displaying the vowel plots of two speakers (UH, a female speaker and KA, a male speaker). These two speakers behave quite differently in the way they produce their central vowels in particular. The central vowels for UH are all overlapping, but KA appears to keep the central vowels relatively quite separated.

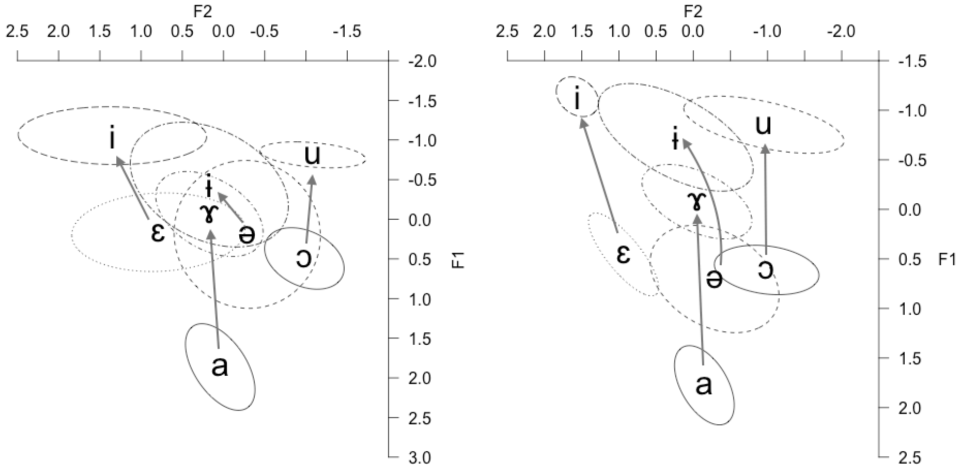


Figure 5–23. Distribution of vowels averaged over the vowel timecourse in a z-normalised F1 X F2 space with data from UH (female) on the left panel and KA (male) on the right panel. The arrows indicate the pair of non-high and high vowels.

With regard to high and non-high vowel pairs, F1 for the non-high member of each vowel pair is consistently higher than for the high member, although the difference in magnitude between [ə] and [ɨ] is less than for the other three vowel pairs. With respect to F2 values for high and non-high vowels, it appears there is also some variation. We can see that the F2 value for the vowel [i] looks higher than the vowel [ɛ] and the F2 value for the vowel [ɨ] is also higher than the vowel [ə], suggesting that the high vowels in these pairs are more fronted than the non-high vowels. However, this does not seem to be really the case for the other two vowel pairs in which case we see that the F2 values for the vowel pairs [ɤ ~ a] and [u ~ ɔ] look very similar. Thus, some variations are also observed in F2 values between the high and non-high vowels pairs, particularly between [i ~ ɛ] and [ɨ ~ ə]. However, such variations do not look to be as dramatic as those in F1 values.

5.12.2 Model comparison for F1 and F2

In order to estimate the differences in F1 and F2 values for high and non-high vowels in Madurese, we compared the following linear-mixed effects models:

f1a: zF1 ~ Vowel + (1 | Speaker) + (1 | Word)
f1b: zF1 ~ Vowel + (1 + Vowel | Speaker) + (1 | Word)

f1c: zF1 ~ Vowel + Place + (1 + Vowel | Speaker) + (1 | Word)
f1d: zF1 ~ Vowel + Place + (1 + Vowel + Place | Speaker) + (1 | Word)[14]

Table 56. Log-likelihood results for F1 model comparison

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
f1a	11	3755.9	3832.6	-1866.9			
f1b	46	2375.5	2696.4	-1141.8	1450.37	35	< 2.2e-16***
f1c	49	2280	2621.8	-1091	101.51	3	< 2.2e-16***
f1d	79	2000.2	2551.3	-921.1	339.79	30	< 2.2e-16***

The result of the log-likelihood ratio test in Table 56 shows that the model f1d was the maximal model justified by our data. This model includes Vowel and Place as fixed effects and as random effects it includes by-speaker and by-word random intercepts as well by-speaker random slopes for Vowel and Place. It is important to note that Place here means the place of articulation of the preceding consonants.

5.12.3 Inferential statistics on F1 and F2 at vowel onset

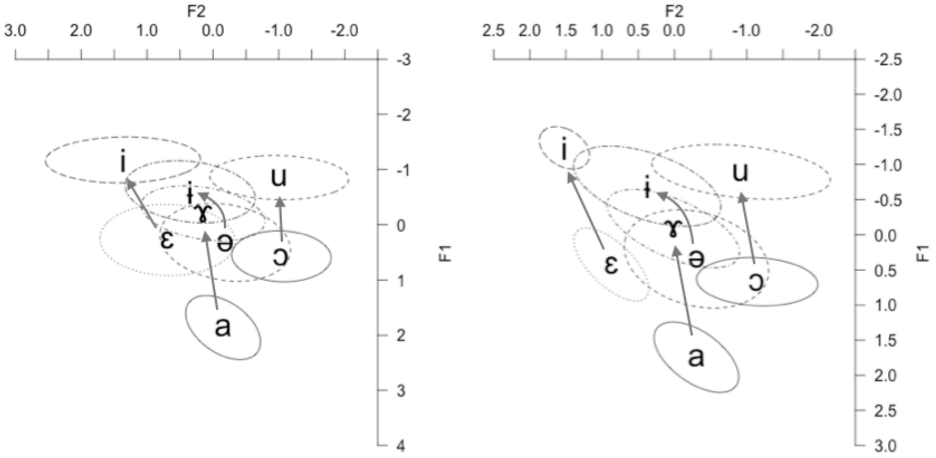


Figure 5–24. Distribution of vowels measured at onset in a z-normalised F1 X F2 space with data from female on the left panel and male on the right panel. The arrows indicate the pair of non-high and high vowels.

Figure 5–24 shows the vowel space of Madurese and demonstrates the differences between the pairs of high and non-high vowels (i ~ ɛ, ɨ ~ ə, ɤ ~ a, u ~ ɔ). F1 and F2

---

[14] The model for F2 has the same structure as the model for F1.

values were pooled across speakers and repetitions and were sampled at vowel onset by averaging timepoints 1-3.

Table 57. Mean frequencies (Hz) and standard deviations (in parentheses) for the first and second formant frequencies of vowels pooled across places of articulation, speakers and repetitions sampled at vowel onset.

	ə	ɨ	ɤ	a	ɛ	i	ɔ	u
				F1				
Female	609	452	527	900	610	360	673	410
	(97)	(84)	(81)	(100)	(87)	(62)	(74)	(75)
Male	552	406	490	740	565	335	595	382
	(74)	(51)	(54)	(70)	(52)	(35)	(45)	(39)
				F2				
Female	1711	1964	1867	1739	2138	2439	1283	1334
	(328)	(335)	(319)	(213)	(390)	(456)	(252)	(356)
Male	1459	1706	1582	1465	1890	2134	1157	1251
	(268)	(245)	(234)	(156)	(151)	(129)	(217)	(319)

Table 57 provides the averaged measurement results for the first and the second formant frequencies of vowels at vowel onset. The values were pooled across places of articulation, speakers and repetitions. To compare differences in vowel height, we conducted a series of post-hoc pairwise comparisons between vowels. First, we present the pairwise comparisons between the pair of high and non-high vowels. Table 58 reports a subset of those comparisons. As seen in Table 58, the results show that there was a significant difference in F1 values between all pairs tested.

Table 58. Pairwise comparison of high versus non-high vowels for F1 at onset. *P* values are adjusted for multiple comparisons based on the Tukey method for a family of 8 means.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
ɛ - i	1.1250	0.1190	9.50	9.451	**0.0001**
ɨ - ə	1.1441	0.1231	16.55	9.296	**< .0001**
ɤ - a	-1.6018	0.1355	25.08	-11.821	**< .0001**
ɔ - u	1.6782	0.1096	8.96	15.310	**< .0001**

The next question that needs to be addressed is whether high and non-high vowels also significantly differed in terms of their F2 values. To confirm this, the same model was used to model F2. As shown in Table 59, the only pair for which F2 showed a significant difference in F2 at onset was the pair [ɨ] and [ə] ($p < .0001$).

Table 59. Pairwise comparison of high versus non-high vowels for F2 at onset. *P* values are adjusted for multiple comparisons based on the Tukey method for a family of 8 means.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
ɛ - i	0.4423	0.1480	57.69	2.988	0.1092
ɨ - ə	-1.1959	0.1278	27.55	-9.354	< **.0001**
ɤ - a	-0.1697	0.0908	38.13	-1.869	0.8660
ɔ - u	-0.1751	0.1028	135.94	-1.703	0.9305

A further interesting question with regard to F1 and F2 values at vowel onset is whether the vowels [ɤ] and [ɨ] as well as [ɤ] and [ə] were also significantly different from one another. It is important to be borne in mind that these vowels do not belong to the pair of high and non-high vowels compared previously. The reason why it is also important to look at them here is because they are impressionistically very similar. This is also evident if we look at the vowel plots in Figure 5–24, in which both the F1 and F2 values of these vowels look overlapping. In order to assess them, we used the same linear mixed-effects model described earlier.

Table 60. Pairwise comparison of the central vowel pairs for F1 and F2 at onset. *P* values are adjusted for multiple comparisons based on the Tukey method for a family of 8 means.

	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
F1	ɤ - ɨ	0.6288	0.0725	4.99	8.674	**0.0095**
	ɤ - ə	1.7729	0.0995	3.13	17.816	**0.0084**
F2	ɤ - ɨ	-0.4122	0.0923	34.34	-4.467	**0.0023**
	ɤ - ə	-1.6081	0.1148	16.27	-14.004	< **.0001**

Table 60 reports the pairwise comparisons from the previous model for the central vowel pairs. As shown in Table 60 above, the differences in the F1 and F2 values for the vowels [ɤ] and [ɨ] and the vowels [ɤ] and [ə] were all significant at vowel onset.

5.12.4 Inferential statistics on F1 and F2 at vowel midpoint

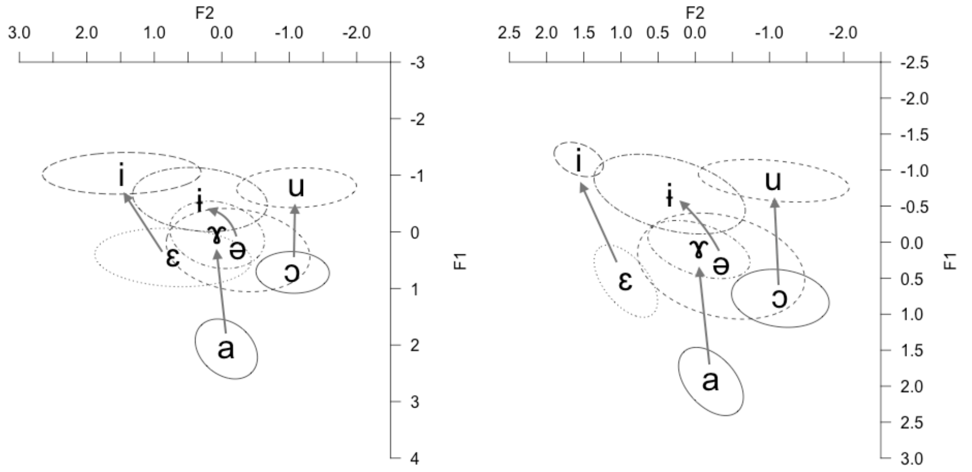


Figure 5–25. Distribution of vowels measured at midpoint in a z-normalised F1 X F2 space with data from female on the left panel and male on the right panel. The arrows indicate the pair of non-high and high vowels.

Figure 5–25 shows the acoustic realisations of the eight surface vowels in Madurese and displays the differences between the high and non-high vowel pairs (i ~ ɛ, ɨ ~ ə, ɤ ~ a, u ~ ɔ) at vowel midpoint. F1 and F2 values were also pooled across speakers, places of articulation and repetitions and sampled at vowel midpoint by averaging the middle four timepoints 5-7.

Table 61. Mean frequencies (Hz) and standard deviations (in parentheses) for the first and second formant frequencies of vowels pooled across places of articulation, speakers and repetitions at vowel midpoint.

	ə	ɨ	ɤ	a	ɛ	i	ɔ	u
				F1				
Female	612	457	571	934	640	381	693	424
	(102)	(85)	(92)	(89)	(71)	(71)	(64)	(77)
Male	551	411	516	770	581	347	611	388
	(77)	(48)	(41)	(70)	(52)	(29)	(48)	(34)
				F2				
Female	1696	1956	1837	1774	2153	2509	1283	1275
	(360)	(338)	(244)	(174)	(408)	(433)	(200)	(297)
Male	1450	1702	1560	1495	1913	2154	1151	1187
	(288)	(251)	(180)	(120)	(126)	(113)	(177)	(253)

Table 61 shows the averaged measurement results for the first and second formant frequencies of vowels measured at vowel midpoint. The values were pooled across places of articulation, speakers, and repetitions. In this regard, the same question that also needs to be addressed here is whether the high and non-high vowels have significantly different F1 values at vowel midpoint.

Table 62. Pairwise comparison of high versus non-high vowels for F1 at midpoint. *P* values are adjusted for multiple comparisons based on the Tukey method for a family of 8 means.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
ɛ - i	1.2481	0.1128	12.28	11.060	< **.0001**
ɨ - ə	1.2907	0.1206	24.09	10.705	< **.0001**
ɤ - a	-1.8328	0.1503	37.54	-12.193	< **.0001**
ɔ - u	1.7319	0.1042	10.48	16.626	< **.0001**

To find out whether there was a significant difference in F1 and F2 values between high and non-high vowels at vowel midpoint, we fitted models as described in Section 5.12.2 and conducted a similar series of between-vowel post-hoc tests. As seen in Table 62 above, all high and non-high vowel pairs had significantly different F1 values at vowel midpoint.

Table 63. Pairwise comparison of high versus non-high vowels for F2 at midpoint. *P* values are adjusted for multiple comparisons based on the Tukey method for a family of 8 means.

Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
ɛ - i	0.5105	0.1499	64.20	3.404	**0.0317**
ɨ - ə	-1.3814	0.1355	40.63	-10.193	< **.0001**
ɤ - a	-0.2368	0.0951	36.24	-2.491	0.3891
ɔ - u	-0.0327	0.1086	172.79	-0.302	1.0000

The next question that needs to be addressed is whether there was a significant difference in F2 values between high and non-high vowels at vowel midpoint. As shown in Table 63, only the F2 values for the pair [i] and [ɛ] and the pair [ɨ] and [ə] were significantly different at vowel midpoint.

Table 64. Pairwise comparison of the central vowel pairs for F1 and F2 at midpoint. *P* values are adjusted for multiple comparisons based on the Tukey method for a family of 8 means.

	Contrast	Estimate	Std.Error	d.f.	t.ratio	*p*-value
F1	ɤ - ɨ	0.3706	0.0725	4.54	5.114	**0.0417**
	ɤ - ə	1.6613	0.1083	3.05	15.341	**0.0034**
F2	ɤ - ɨ	-0.3387	0.1005	34.34	-3.369	**0.0481**
	ɤ - ə	-1.7202	0.1227	22.54	-14.015	< **.0001**

Like F1 and F2 values at vowel onset, the same question is whether the differences between the F1 and F2 values for the vowels [ɤ] and [ɨ] as well as [ɤ] and [ə] were also significantly different at vowel midpoint. As shown in Table 64, the differences between the F1 and F2 values for the vowels [ɤ] and [ɨ] and the vowels [ɤ] and [ə] were also all statistically significant at vowel midpoint.

5.12.5 F1 and F2 as a function of Vowel and Voicing

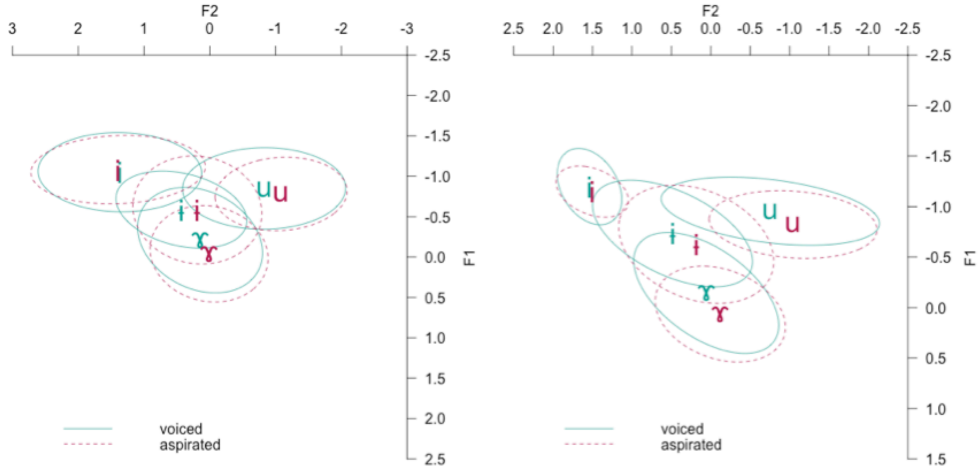


Figure 5–26. Distribution of high vowels following voiced and voiceless aspirated stops averaged over the course of the vowels in a z-normalised F1 X F2 space with data from female on the left panel and male on the right panel.

A number of studies (e.g. Fischer-Jørgensen, 1968; Shimizu, 1996, pp. 61–63) have shown that F1 values following voiceless stops are higher than those following voiced stops. Since voiced and voiceless aspirated stops in Madurese are both followed by high vowels, it is possible to examine these vowels as a function of voicing to see whether the two stop categories exert different effects on F1 and F2. This analysis relates to the research question on whether or not voiced and voiceless aspirated share acoustic features. That is, if F1 and F2 following voiced and voiceless aspirated stops are not significantly different, it suggests that they share the features.

Figure 5–26 shows mean F1 and F2 values for high vowels following voiced and voiceless aspirated stops. As we can see, the F1 values following voiced stops tend to be lower than the F1 values following voiceless aspirated stops. This particularly seems to be the case for the vowels [ɨ], [ɤ] and [u], but not for the vowel [i]. The F2 values for vowels following voiced stops look higher than those after voiceless aspirated stops. Again this is only apparent for the vowels [ɨ], [ɤ] and [u] while the vowel [i] shows no such a tendency. However, as expected based on the plots in

Figure 5–26, in which the F1 and the F2 values for voiced and voiceless aspirated stops overlap considerably, none of the terms reached statistical significance.

5.12.6 Summary and implication of results for F1 and F2

In this section we have examined the first and second formant frequencies of Madurese vowels at vowel onset and vowel midpoint by looking at whether the high and non-high vowel pairs show significant differences in their F1 and F2 values. Using linear mixed-effects models, we have established that all the high and non-high vowel pairs show significant differences in their F1 values at both vowel onset and midpoint. The results for F2 is, however, quite variable. At vowel onset, only the pair [ɨ ~ ə] turns out to show a significant difference in F2 values and at vowel midpoint the vowel pairs [i ~ ɛ] and [ɨ ~ ə] show significant differences. Furthermore, we have also looked at the vowels [ɤ] and [ɨ] and the vowels [ɤ] and [ə] to see whether they also differ in their F1 and F2 values. We have confirmed that at both vowel onset and midpoint the F1 and F2 values for the vowels [ɤ] and [ɨ] and the vowels [ɤ] and [ə] turn out to be significantly different.

In conclusion, the pairs of high and non-high vowels in Madurese consistently show significant differences in their F1 values. On the other hand, F2 values have been shown to vary with vowel pairs and vowel timepoints. What is also interesting here is the fact that the vowels [ɤ] and [ə], which are very similar impressionistically even though they do not constitute a pair of high and non-high vowels, demonstrate consistent differences in their F1 and F2 values at both measurement points. With regard to the question whether voiced and voiceless aspirated stops share acoustic properties to the exclusion of voiceless unaspirated stops, the results for F1 and F2 show positive answers. Recall that there are no differences in F1 and F2 values between the two stop categories indicated by considerable overlap in their values, suggesting that they share these properties.

The theoretical implication of these results in particular with respect to whether voiced and voiceless aspirated stops have the feature [ATR] or [LL] can be explained in the following way. Either [+ATR] or [+LL] vowels predict that they have lower F1 values in comparison with their [-ATR] or [-LL] counterparts. In this case, the results for F1 are in line with both of these predictions, indicating that the two sets of high and non-high vowels in Madurese are consistent with either ATR or LL phonetic prediction. In terms of F2 values, however, the two vowel sets are variable, suggesting that F2 may not be a reliable acoustic correlate for ATR contrast as is also the case in a number of ATR languages.

5.13 General Summary of Results

This chapter has presented the results of statistical analyses for VOT, closure duration, F0, H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4*, CPP, F1 and F2. An overview of the findings is presented in Table 65 below.

Table 65. Comparisons of acoustic measures following voiced, voiceless unaspirated and voiceless aspirated stops at vowel onset and vowel midpoint. 'Yes' indicates statistical significance at $p < 0.05$.

No.	Acoustic Measure	Contrast	Female		Male	
			Onset	Midpoint	Onset	Midpoint
1.	VOT	Voiced vs. Voiceless	Yes			
		Voiced vs. Aspirated	Yes			
		Voiceless vs. Aspirated	Yes			
2.	Closure Duration	Voiced vs. Voiceless	Yes			
		Voiced vs. Aspirated	Yes			
		Voiceless vs. Aspirated	No			
3.	F0	Voiced vs. Voiceless	Yes	No	No	Yes
		Voiced vs. Aspirated	Yes	Yes	Yes	No
		Voiceless vs. Aspirated	No	Yes	No	No
4.	H1*-A1*	Voiced vs. Voiceless	Yes			
		Voiced vs. Aspirated	No			
		Voiceless vs. Aspirated	Yes			
5.	H1*-A2*	Voiced vs. Voiceless	No			
		Voiced vs. Aspirated	Yes			
		Voiceless vs. Aspirated	Yes			
6.	H1*-A3*	Voiced vs. Voiceless	No	No	Yes	Yes
		Voiced vs. Aspirated	Yes	Yes	No	No
		Voiceless vs. Aspirated	No	No	Yes	Yes
7.	H1*-H2*	Voiced vs. Voiceless	Yes	Yes	Yes	No
		Voiced vs. Aspirated	No	No	Yes	Yes
		Voiceless vs. Aspirated	Yes	Yes	Yes	Yes
8.	H2*-H4*	Voiced vs. Voiceless	No	No	Yes	Yes
		Voiced vs. Aspirated	No	No	No	No
		Voiceless vs. Aspirated	No	No	Yes	Yes
9.	CPP	Voiced vs. Voiceless	No	No	Yes	Yes
		Voiced vs. Aspirated	No	Yes	Yes	No
		Voiceless vs. Aspirated	Yes	Yes	Yes	Yes

The main goal of measuring VOT was to find out whether voiced, voiceless unaspirated and voiceless aspirated stops in Madurese have different VOT values. The results show that the three stop series in Madurese can be divided into three types, comprising prevoiced stops, short lag stops and slightly long lag stops. Prevoiced stops are characterised by VOT that begins before the stop release while short lag and long lag stops are characterised by VOT that begins after the stop release and before the onset of the vowel.

The VOT values in Madurese stops are interesting given the fact that the VOT difference between voiceless unaspirated and voiceless aspirated stops is not so large although it is true that they are statistically significant. More important, however, is the fact that these two voiceless stop categories show a considerable overlap in their VOT distributions. This may raise the question about whether VOT is in fact the primary cue that can distinguish between these two voiceless categories in Madurese.

We have also looked at fundamental frequency of vowels following voiced, voiceless unaspirated and voiceless aspirated stops. We found that the F0 following voiced stops is consistently lower than that following voiceless unaspirated and voiceless aspirated stops. Specifically, the F0 for voiceless aspirated stops tends to be higher than the F0 for voiceless unaspirated stops. In this case, the results for F0 have a consistent pattern with those for VOT in the sense that there is a correlation between VOT and F0. That is, negative VOT correlates with lower F0 while positive VOT is associated with higher F0. Recall that just like the difference in VOT between voiceless unaspirated and voiceless aspirated stops is not large (i.e. 17 ms and 40 ms respectively for females, 15 ms and 32 ms respectively for males), the F0 difference between these two voiceless stops is not large either. Indeed, the F0 results in the vowels following each voicing category are F0 patterns that are also commonly found across languages (Hombert & Ladefoged, 1976; House & Fairbanks, 1953; Löfqvist *et al.*, 1989; Ohde, 1984).

We have also examined the voice quality in the vowels following voiced, voiceless unaspirated and voiceless aspirated stops by measuring a number of its acoustic correlates, namely H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. Our results demonstrate that in general H1*-A1* values for voiced and voiceless aspirated stops are greater than H1*-A1* value for voiceless unaspirated stops. More important is the fact that there is no significant difference between H1*-A1* values for voiced and voiceless aspirated stops, suggesting that they pattern together in this spectral measure to the exclusion of voiceless unaspirated stops. In contrast, the results for H1*-A2* are quite opposite to those for H1*-A1* given that it is voiceless unaspirated and voiced stops that pattern together in this measure. However, similar to the results for H1*-A1*, the H1*-A2* values for voiceless aspirated stops remain consistently greater than those for the other voicing categories.

Unlike the results for H1*-A1* and H1*-A2*, the results for H1*-A3* yield a rather mixed picture. This is probably due to the fact that gender variation is also quite prominent in this case. Specifically, females' H1*-A3* values for voiceless unaspirated and voiced stops pattern together to the exclusion of H1*-A3* for voiceless aspirated stops. On the contrary, consistent with the results for H1*-A1*, males' H1*-A3* values for voiced and voiceless aspirated stops pattern together to the exclusion of H1*-A3* for voiceless unaspirated stops.

With respect to H1*-H2*, our results demonstrate that the H1*-H2* values for voiced and voiceless aspirated stops are consistently greater than the H1*-H2* value for voiceless unaspirated stops. More importantly, there is no significant difference between the H1*-H2* values for voiced and voiceless aspirated stops, suggesting that they pattern together in this measure. There is also evidence that the H2*-H4* values for voiced and voiceless aspirated stops pattern together to the exclusion of the H2*-H4* value for voiceless unaspirated stops. However, similar to the results for H1*-A3*, the patterning together of voiced stops and voiceless aspirated stops in H2*-H4* values to the exclusion of voiceless unaspirated stops can only be observed in male speakers. That is, there are significant differences between males' H2*-H4* values for voiceless unaspirated and voiced stops and for voiceless aspirated and voiceless unaspirated stops. In contrast, there is no significant difference between males' H2*-H4* values for voiceless aspirated and voiced stops. Furthermore, there is a general trend for voiced and voiceless aspirated stops to have lower CPP values compared to voiceless unaspirated stops. There is also evidence that the CPP values for voiced and voiceless aspirated stops pattern together to the exclusion of the CPP value for voiceless unaspirated stops. However, they also vary with gender and vowel timecourse whereby the patterning in CPP values for females primarily occurs at vowel onset while it mostly occurs at vowel midpoint for males.

With regard to our results for F1 and F2, we found that in general the high and non-high vowel pairs can be distinguished by their F1 values. In contrast, F2 values only distinguish between the pair [i] and [ɛ] at vowel midpoint and between the pair [ɨ] and [ə] at vowel onset and midpoint. The other pairs, i.e. [ɔ ~ u] and [a ~ ɤ], appear to have similar F2 values in this case. We also examined the three similar sounding vowels [ɤ], [ə] and [ɨ] to see whether they have differences in F1 and F2. We found that the F1 and F2 values for [ɤ] and [ə] and for [ɤ] and [ɨ] are significantly different at vowel onset and midpoint. In addition, we also examined whether voiced and voiceless aspirated stops affect F1 and F2 differently. Our results confirm that there is no significant difference between F1 and F2 values for the two voicing categories.

6 Discussion

6.1 Introduction

This chapter is structured in the following way. Section 6.2 discusses acoustic properties of the three-way laryngeal contrast in Madurese and evaluates whether voiced and voiceless aspirated stops share acoustic property (or properties) to the exclusion of voiceless unaspirated stops. Section 6.3 discusses whether voice quality is independent of vowel quality. Based on phonetic evidence, we argue for the relative independence of voice quality of vowel quality. Section 6.4 addresses what plausible phonological feature may be responsible for triggering the CV co-occurrence restriction in Madurese by evaluating how the acoustic findings bear on the proposals of [ATR] and [LL] features. Some implication of the results for theories in which phonological features are expected to have transparent phonetic realisations will also be discussed. Section 6.5 discusses the possible origins of voiceless aspirated stops in Madurese by looking at some historical evidence and loanword phenomena. Section 6.6 addresses the vowel system of Madurese and argues for the proposal that Madurese can be best described as a language with a four-vowel system. This section also establishes in particular the debated status of the vowels [ə] and [ɨ] in Madurese based on the acoustic findings.

6.2 Acoustic Properties of the Voicing Contrast in Madurese

The acoustic study found evidence for three categories of stop consonants based on the distribution of VOT values, but voiceless unaspirated and voiceless aspirated stops were found to have relatively small differences in their VOT values. More importantly, the VOT distributions for the two voiceless stop categories overlap considerably. The results are consistent with previous findings by Cohn and Lockwood (1994) and Cohn and Ham (1998) who also identify that the VOT distinction between voiceless unaspirated and voiceless aspirated stops is not as robust as that between voiced stops on the one hand and voiceless unaspirated and aspirated stops on the other. See Table 66 on the following page for an overview of the acoustic correlates of voiced, voiceless unaspirated and voiceless aspirated stops.

Table 66. Summary of results for statistical analyses for acoustic correlates

Measures	Voicing Type		
	Voiced	Unaspirated	Aspirated
VOT	Negative	Positive, short-lag	Positive, slightly longer-lag
Closure Duration	Long	Short	Short
F0	Low	High	High
H1*-A1*	High	Low	High
H1*-A2*	Unpredictable	Unpredictable	Unpredictable
H1*-A3*	High (m)	Low (m)	High (m)
H1*-H2*	High	Low	High
H2*-H4*	High (m)	Low (m)	High (m)
CPP (f)	Low (onset)	High	Low (onset)
CPP (m)	Low (midpoint)	High	Low (midpoint)
F1	High	Low	High

In terms of F0, we found that voiced, voiceless unaspirated and voiceless aspirated stops display a relatively similar behaviour to what we have observed in their VOT results. Of particular interest here is the fact that the F0 following voiced stops is significantly lower than that following voiceless (unaspirated and voiceless aspirated) stops across genders. This result contradicts the findings of Cohn and Lockwood (1994) who observed that the F0 following voiced and voiceless aspirated stops is lower than that following voiceless unaspirated stops. The difference in the F0 finding from that of earlier researchers is probably due to the fact that the present study has a more representative sample. The present finding is, however, consistent with a general trend that F0 onset following voiced stops is lower than that following voiceless stops as has also been demonstrated in a number of studies involving tonal languages such as Thai and Yoruba (Hombert, 1978) as well as non-tonal languages such as French and English (Hombert, 1975; Mohr, 1971).

In addition, for female speakers the F0 following voiceless aspirated stops in Madurese is significantly higher than that following voiceless unaspirated stops. However, male speakers do not show a significant difference between their F0 values for voiceless unaspirated and voiceless aspirated stops. This variability has also been observed in previous studies. For example, the finding is in agreement with Lai *et al.* (2009), who found that in Taiwanese F0 following voiceless aspirated stops was higher than F0 following voiceless unaspirated stops and the raising effect of aspiration on F0 was particularly greater for female speakers. This latter effect may be related to physical differences between females and males as physiologically females have smaller vocal tracts than males. In contrast, the result is not in line with Xu and Xu (2003), who studied the effects of aspiration on Mandarin tones and found that the onset F0 of tone is significantly higher following voiceless unaspirated stops than following voiceless aspirated stops.

With respect to the question of whether voiced and voiceless aspirated stops share phonetic properties to the exclusion of voiceless unaspirated stops, our findings on VOT and F0 have provided a clearly negative answer. As a natural step, we have moved further to see whether voice quality would probably provide evidence for shared phonetic qualities between voiced and voiceless aspirated stops. For this purpose, we examined a number of voice quality correlates: H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. These acoustic measures have been explored in various studies (e.g. DiCanio, 2009; Esposito, 2010; Esposito & Khan, 2012; Garellek & Keating, 2011).

It is important to bear in mind that H1-H2 is the relative difference between the amplitudes of the first harmonic (H1) and the second harmonic (H2). H1-H2 is an acoustic correlate of the open quotient (OQ), indicating the percentage of the glottal cycle during which the glottis is open (Holmberg *et al*., 1995). The mechanism that may explain the correlation of OQ with H1-H2 is that the greater the open quotient (i.e. the longer the vocal folds are abducted), the greater the amplitude of the first harmonic relative to the amplitude of the second harmonic. In this case, H1-H2 for breathy vowels is expected to be greater than for modal vowels.

H1-A1 is the relative difference between the amplitudes of the first harmonic (H1) and the most prominent harmonic in the F1 region (A1). This acoustic parameter (A1) indicates F1 bandwidth. Formant bandwidths have been associated with some energy losses in the vocal tract due to such factors as the yielding walls' resistance of the vocal tract, conduction of heat and losses at the walls due to frictions (Stevens & Hanson, 1995). The airflow that goes through the open glottis triggers glottal resistance and this can subsequently contribute to the loss of energy adding up significantly to the F1 bandwidth (Stevens & Hanson, 1995).

There are two acoustic parameters that are commonly used for measuring spectral tilt or spectral balance, namely H1-A2 and H1-A3. Stevens (1977) suggests that the slope of the source spectrum has a correlation with the abruptness or gradualness of the vocal fold closure. That is to say, the vocal folds which come together in a gradual fashion primarily causes an excitation of the lower frequencies of the vocal tract, resulting in a spectrum with a steep slope whose energy is mostly concentrated in the region close to the fundamental frequency while very little energy is found at higher frequencies. On the other hand, the vocal folds which come together simultaneously may provide a sufficient excitation on a wider range of frequencies, resulting in a less steep spectrum whose higher frequency components are relatively stronger. Since breathy phonation is characterised by the vocal folds with a gradual closure, the fundamental frequency is expected to be much higher in amplitude than the higher harmonics.

Cepstral peak prominence (CPP), which is a measure of the signal strength over noise across the spectrum, is another measure of periodicity and has also been used to measure breathiness. A well-defined harmonic structure indicates a high periodicity of a signal, which results in a signal having a more prominent cepstral peak than a less periodic one (Hillenbrand *et al.,* 1994). In this case, higher values indicate more periodic signals while lower values indicate less periodic signals. Since breathy phonation has less distinct harmonics, they are expected to have lower CPP values than modal phonation.

In relation to the acoustic correlates of voice quality mentioned above, it is also important to bear in mind that different languages may have different results with regard to these voice quality measures. This is because phonation types in some languages may be more sensitive to some measures than others. In fact, there is also evidence that even speakers of the same language may show different results with respect to these spectral measures (DiCanio, 2009). There is also evidence of gender-related differences in the realisation of phonation types (see e.g. Esposito, 2010b; Wayland & Jongman, 2001).

Our results demonstrate that in general H1*-A1* values for voiced and voiceless aspirated stops are greater than H1*-A1* value for voiceless unaspirated stops. More important is the fact that there is no significant difference between H1*-A1* values for voiced and voiceless aspirated stops, suggesting that they pattern together in this spectral measure as opposed to voiceless unaspirated stops. In contrast, the results for H1*-A2* are quite opposite to those for H1*-A1* given that it is voiceless unaspirated and voiced stops that pattern together in this measure. However, similar to the results for H1*-A2*, the H1*-A2* values for voiceless aspirated stops remain consistently greater than those for the other voicing categories. In summary, unlike H1*-A1*, the results for H1*-A2* show a different pattern with respect to the shared phonetic properties of voiced and voiceless aspirated stops in particular. This is because it is voiceless unaspirated and voiced stops that pattern together.

The results also show that voiceless unaspirated and voiceless aspirated stops differ in their H1*-A3* values for both genders. However, we found that the results for H1*-A3* is to some extent also gender-specific. Specifically, H1*-A3* distinguishes between voiceless unaspirated and voiced stops for male speakers but it does not contrast between these categories for female speakers. H1*-A3* also distinguishes between voiced and voiceless aspirated stops for female speakers, but it does not for male speakers. Thus, at least for male speakers, voiced and voiceless aspirated stops share this acoustic property.

With respect to H1*-H2*, the results demonstrate that the H1*-H2* values for voiced and voiceless aspirated stops are consistently greater than the H1*-H2* value for voiceless unaspirated stops. Specifically, H1*-H2* consistently contrasts between voiced and voiceless unaspirated stops and also between voiceless unaspirated and voiceless aspirated stops. More importantly, the fact that there is no significant difference between the H1*-H2* values for voiced and voiceless aspirated stops suggests that they share this acoustic property as well.

There is also evidence that H2*-H4* values for voiced and voiceless aspirated stops pattern together to the exclusion of voiceless unaspirated stops. However, similar to H1*-A3*, this can be observed only in male speakers. Specifically, significant differences are observed between males' H2*-H4* values for voiceless unaspirated and voiced stops and for voiceless aspirated and voiceless unaspirated stops, but not between their H2*-H4* values for voiceless aspirated and voiced stops. On the contrary, H1*-H4* does not distinguish the three stop types for female speakers.

Furthermore, the results for CPP are particularly interesting in the context of Madurese given that this measure does not require F0 analysis and therefore theoretically it is not correlated with vowel height. Interestingly, there is a general trend for voiceless aspirated and voiced stops to have lower CPP values compared to voiceless unaspirated stops. There is also some evidence that the CPP values for voiced and voiceless aspirated stops pattern together to the exclusion of voiceless unaspirated stops. However, they also vary with gender and vowel timecourse. Specifically, the patterning can be observed at vowel onset for female speakers and at vowel midpoint for male speakers. This result is consistent with what was found in Gujarati where post-aspirated vowels have lower CPP values than modal (or breathy) vowels (Esposito & Khan, 2012).

Our results indicate that voiceless aspirated stops consistently have higher H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, and H2*-H4* compared to voiceless unaspirated stops. We may speculate that differences in these acoustic measures may also be used by these two voiceless stop categories for expressing or enhancing their contrast. This would make sense if we consider the fact that both VOT and F0 do not robustly distinguish between voiceless unaspirated and voiceless aspirated stops. Recall that their VOT values are significantly different, but the fact that their VOT distributions overlap considerably is also important to be taken into consideration. Also recall that the difference in F0 between voiceless unaspirated and voiceless aspirated stops is only significant for female speakers.

However, it is important to bear in mind that unless we conduct a perceptual experiment designed to examine which acoustic correlates are relevant for their distinction, we cannot determine whether these acoustic cues are also perceptually used by Madurese listeners to discriminate between, for example, voiceless unaspirated and aspirated stops. Thus, at this stage of study we only assume that the acoustic features examined in this study are relevant correlates that can distinguish one category from another. Whether or not they are also perceptually relevant cues or the listener attends to them for distinguishing segments becomes another matter.

Furthermore, with regard to the question of whether voiced and voiceless aspirated share phonetic qualities, our findings demonstrate that they indeed consistently exhibit similar H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP values. Thus, our findings on voice quality measures here are in contrast with our findings on VOT and F0 given that voiced and voiceless aspirated stops have very different VOT and F0 values. In other words, we observe that voiced and voiceless aspirated stops pattern together in some of their voice quality measures while they do not share similar VOT and F0. We may interpret these results as possible evidence that may throw phonetic light on the reason why voiced and voiceless aspirated stops phonologically pattern together in the CV co-occurrence restriction in Madurese.

6.3 Evidence for Voice Quality Independence of Vowel Height

Spectral slope is known to vary with vowel quality, especially vowel height. In order to try and make comparisons between the phonation types of vowels differing in height, it is important to remove or compensate for the effects of particular formant frequencies on the magnitudes of the harmonics. In this thesis, this was accomplished using the correction formula developed by Iseli and Alwan (2004). They argue that the formula they propose, which corrects for both formant frequencies and their bandwidths, would allow for comparison between vowels regardless of their qualities; therefore, it is not limited to comparison between certain (low) vowels.

While their correction compensates for the effects of formant frequencies on the speech spectrum, it does not remove any *source*-related differences that may be present. In their study of English speakers, Iseli *et al.* (2007) observed F1-related dependencies for both the corrected measures H1*-H2* and H1*-A3*. This raises the possibility that any spectral slope differences observed between vowels of different heights may be due to source-related differences between the vowels.

It is true that teasing apart voicing and vowel height is difficult in Madurese due to the CV co-occurrence restriction, i.e. voiced and voiceless aspirated stops only co-occur with high vowels while voiceless unaspirated stops only co-occur with non-high vowels. However, we can try to look for other possible evidence that voice quality in Madurese is to some extent independent of vowel height. This evidence is obtained by comparing high and non-high vowels whose height does not differ dramatically. For this purpose, we compare three vowels that are acoustically similar in height but phonologically different, namely the non-high vowels [ɛ], [ə] and the high vowel [ɤ]. If voice quality depends on vowel height, we expect the H1*-A1* or H1*-H2* values for the high vowel [ɤ] would be different from the non-high vowels [ɛ] and [ə], which pattern together. As seen in Figure 6–1 below and Figure 6–2 on the following page, despite the fact that they have similar height, we can still observe that the high vowel [ɤ] in general has higher H1*-A1* or H1*-H2* values, suggesting that voice quality is to some extent independent of vowel height.

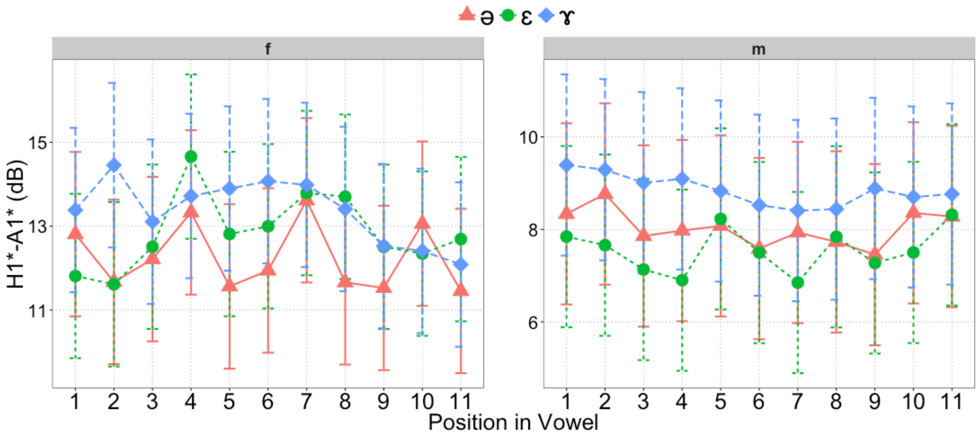


Figure 6–1. Mean H1*-A1* of vowels [ə], [ɛ], and [ɤ] measured at 11 equidistant timepoints; female on the left panel and male on the right. Error bars represent 95% confidence interval.

Post-hoc pairwise tests confirmed that there was a significant difference in H1*-A1* between the vowels [ɤ] and [ɛ] at vowel onset (*p* < .001), but the difference was not significant at vowel midpoint (*p* > 0.81). However, there was no significant difference in H1*-A1* between the vowels [ɤ] and [ə] and between the vowels [ɛ] and [ə] at vowel onset and midpoint (*p* > 0.07 in all cases). The results also showed that there was a significant difference in H1*-H2* between the vowels [ɤ] and [ɛ] at vowel onset and midpoint (*p* < .001 in both cases). Moreover, while the difference in H1*-H2* between the vowels [ɤ] and [ə] was significant at vowel onset (*p* < .001), the difference was not significant at vowel midpoint (*p* > 0.93). In contrast, while the difference in H1*-H2* between the vowels [ɛ] and [ə] was not significant at vowel onset (*p* > 0.39), it was significant at vowel midpoint (*p* < .0001). Overall, these results suggest that there is some evidence of the relative independence between voice quality and vowel height.

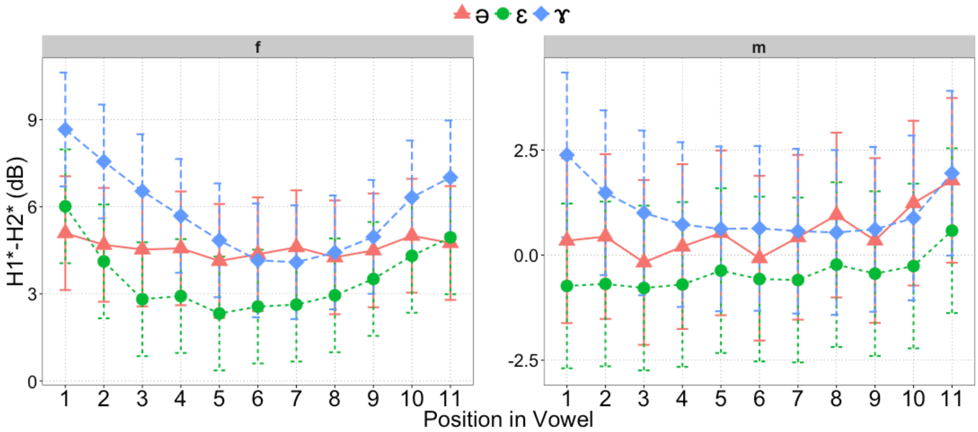


Figure 6–2. Mean H1*-H2* of vowels [ə], [ɛ], and [ɤ] measured at 11 equidistant timepoints; female on the left panel and male on the right. Error bars represent 95% confidence interval.

Thus, we would argue that the results for voice quality measures above are due to some property on the consonantal realisation, instead of just being a property of the vowel height. On the other hand, what connection, if any, exists between these spectral measures and the CV co-occurrence restriction, is a different issue, which is beyond the scope of this dissertation. In summary, we conclude that we did find phonetic evidence that voiced and voiceless aspirated stops in Madurese share some correlates of voice quality, namely H1*-A1* for both genders, H1*-H2* for females, H1*-A3* and H2*-H4* for males, and CPP for females at vowel onset and for males at vowel midpoint.

However, it is also worth noting that the patterning of voiceless aspirated and voiced stops in terms of voice quality measures discussed above may involve two different processes: (1) a vocal fold setting for voiced stops and (2) a purely acoustic effect for aspirated stops. That is, there is a difference between voiceless aspirated stops being ‘breathy’ in the sense of their being produced with a specific laryngeal configuration that also produces high H1-H2 like in Hindi (Dutta, 2007) and simply having a high H1-H2 that may result from general spectral degradation due to aspiration being temporally co-extensive with the onset of periodic voicing during the vowel. This evidence is consistent with our finding that the two voicing categories pattern together in their CPP values as well, a measure that does not require F0 analysis and is therefore arguably independent of vowel height.

Finally, we also examined whether voiced and voiceless aspirated stops affect F1 and F2 differently. Our results indicate that there are no significant differences in the quality of the high vowels following the two stop categories. This suggests that voiced and voiceless aspirated stops also pattern together in this case.

6.4 Possible Feature Accounting for the Patterning together of Voiced and Voiceless Aspirated Stops

Table 67. The features [+ATR] and [+LL] and their predicted acoustic correlates

Feature	Acoustic Cues				
	F0	F1	F2	Spectral Tilt	CPP
[+ATR]	?[15]	lower	lower	higher	lower
[+LL]	lower	lower	?	higher	lower

In this section we further assess the features [+ATR] or [+LL] that have been proposed as possible phonological features that may account for the patterning together of voiced and voiceless aspirated stops in Madurese (see Cohn, 1993a, 1993b; Trigo, 1991). If [ATR] or [LL] is the relevant feature, and if we assume a transparent phonetics-phonology mapping, we may expect to see a certain constellation of phonetic properties. We assess the proposals on the basis of some predicted acoustic correlates that may result from either advancing the tongue root or lowering the larynx. Each proposed feature is then evaluated on the basis of our phonetic findings, which are summarised in Table 67 and a proposal is made on how to best account for the findings relative to the proposed features.

---

[15] James Kirby (personal communication) points out to me that if consonants are [+ATR], this may exert pull on the hyoid bone and therefore potentially have higher F0. However, we put '?' here because there is not much evidence supporting an ATR/F0 correlation.

6.4.1 Is there phonetic evidence for an [ATR] feature?

As the name suggests, the feature [ATR] is a phonological feature associated with an articulatory gesture involving the tongue root advancement. In this respect, the value of [+ATR] is usually borne by high vowels because advancing tongue root lowers F1. A number of studies (e.g. Jacobson, 1978; Ladefoged, 1968; Lindau, 1979; Tiede, 1996) which look at languages that are said to have the feature [ATR] active in their phonologies have provided phonetic evidence associated with the feature. The feature [ATR] in those studies belongs to vowels and in some of the languages the feature [ATR] spreads between vowels triggering vowel harmony. On the other hand, if we assume that it is the prevocalic consonants that bear the feature [ATR] in Madurese, then the spread of this feature to the following vowels could be advanced to explain the CV harmony patterns in the sense of feature spreading as proposed by, for example, Stevens (1980) and Trigo (1991).

The question that needs to be addressed here is whether the acoustic findings support voiced and voiceless aspirated stops sharing the feature [+ATR] in Madurese. It is important to note that in languages where [ATR] has been claimed to be active, it is usually vowels that bear the feature (Casali, 2008). However, some authors (e.g. Trigo, 1991; Vaux, 1996) argue that the feature [ATR] can also belong to consonants and this feature spread rightward to the vowels that follow them. In fact, the relationship between the feature [ATR] with consonants can be traced back to the works of Chomsky and Halle (1968), Perkell (1969) and Westbury (1983), which all suggest that voiced stops (except bilabial ones) are produced with an advanced tongue root to help maintain voicing during the closure. This is because this mechanism also helps reduce air pressure so that subglottal pressure is kept higher than supraglottal pressure as one of the requirements for the vocal fold vibration. However, it is not at all clear why voiceless aspirated stops in Madurese should also be considered as [+ATR], as they are phonetically voiceless during the closure.

In order to test the hypothesis that voiced and voiceless aspirated stops in Madurese have the feature [+ATR], we looked at F1, F2 and some measures of voice quality, i.e. H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP. Recall that here we assume that the vowel alternations in Madurese are due to feature spreading from the consonants. That is why F1 and F2 are also relevant measures to look at. The two acoustic measures H1*-H2* and H1*-A3* in particular have been associated with acoustic correlates of ATR (Fulop, Karib, & Ladefoged, 1998; Guion, Post, & Payne, 2004; Halle & Stevens, 1969; Jacobson, 1980; Jacobson, 1978; Lindau, 1979; Remijsen, Ayoker, & Mills, 2011). If voiced and voiceless aspirated stops have the feature [+ATR], we might expect that these two stop categories would have lower F1 and less peripheral F2 in comparison with voiceless unaspirated stops at vowel onset. We might also expect that voiced and voiceless aspirated stops are produced with a breathy voice quality, which is indicated by greater values in H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H1*-A3* and lower CPP values.

As expected, our results show that there is a significant difference between F1 values for vowels following voiced and voiceless aspirated stops on the one hand and those following voiceless unaspirated stops on the other. Specifically, F1 values for vowels following voiced and voiceless aspirated stops are significantly lower than those following voiceless unaspirated stops. In contrast, there is no significant difference between F1 values for vowels following voiced and voiceless aspirated stops, suggesting that they have similar values. The result for F1 is thus consistent with the prediction if voiced and voiceless aspirated stops were to share an [+ATR] feature.

In terms of F2, the results indicate that not all F2 values of vowels following voiced and voiceless aspirated stops show a significant difference from F2 values for vowels following voiceless unaspirated stops. In this case, only the F2 values for [i] and [ɛ] and [ɨ] and [ə] turn out to be significantly different. However, the way they differ is in contrast with the phonetic prediction of the feature [+ATR]. Recall that [+ATR] vowels are expected to have lower F2 for front vowels while in this case the F2 values for the vowels following voiced and voiceless aspirated stops turn out to be higher, indicating that they are more peripheral than vowels following voiceless unaspirated stops. Therefore, the result for F2 does not support the proposal that voiced and voiceless stops share the feature [+ATR].

In relation to this, it is worth mentioning that F2 values may not be regarded as a reliable correlate for an ATR contrast. In fact, evidence suggests that even languages from the same family appear to vary with respect to this acoustic parameter. For example, Jacobson (1980) found that some vowel sets in the Nilotic languages Dho-Luo and Shilluk demonstrate such F2 effects, but he did not observe the same effects in all vowel pairs for Dinka. However, the finding in Madurese is in line with Fulop *et al.* (1998), who reported that the F2 values of the [+ATR] vowels in Degema were consistently more peripheral than their [-ATR] counterparts. To add to the inconsistency, Guion *et al.* (2004, p. 536) recently also reported that there was no significant difference in F2 values between the [+ATR] and [-ATR] vowels in Maa. They found that the ATR distinction in Maa can be distinguished by F1 and voice quality whereby [+ATR] vowels have lower F1 values and steeper spectral slopes than their [-ATR] counterparts.

Furthermore, the results also show that there are significant differences in H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP values between voiced and voiceless aspirated stops on the one hand and voiceless unaspirated stops on the other. Specifically, voiced and voiceless aspirated stops have higher values in H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and lower values in CPP in comparison with voiceless unaspirated stops. In contrast, no significant differences are observed in H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP values between voiced and voiceless aspirated stops, suggesting that they share these acoustic properties. Thus, like the result on F1, the results on voice quality measures are in line with the phonetic prediction that vowels following both voiced and voiceless aspirated stops may be produced with an advanced tongue root and could be both associated with the feature [+ATR]. Recall that the evidence that the consonants may bear this feature involves the fact that they seem to trigger vowel harmony throughout the word.

6.4.2 Is there phonetic evidence for an [LL] feature?

As the name also suggests, the feature [LL] is associated with the articulatory gesture of lowering the larynx. Unlike [ATR], which has been argued to be phonologically active in a number of languages, the feature [LL] has only been proposed for a limited number of languages. To my knowledge, the feature has been argued to be active in Buchan Scots (Paster, 2004; Youssef, 2010) and Madurese (Cohn, 1993b; Trigo, 1991) and has also been suggested to be responsible for the CV harmony in these languages. If voiced and voiceless aspirated stops share the feature [+LL], we should expect that the two stop categories would have a lowering effect on F0 and F1 of vowels following them. In addition, we should also expect that both voiced and voiceless aspirated stops are produced with a breathy voice quality and this could also be indicated by greater values in H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2* and H2*-H4*, and lower CPP values in the following vowels.

If voiced and voiceless aspirated stops in Madurese are both [+LL], we should expect that the F0 values following voiced and voiceless aspirated stops would pattern together by having lower F0 values than that following voiceless unaspirated stops. Our result indicates that this phonetic prediction is contradictory with our finding whereby F0 following voiceless unaspirated and voiceless aspirated stops are in fact higher than that following voiced stops. Specifically, there is a significant difference in F0 values between voiced and voiceless aspirated stops and the F0 difference between voiced and voiceless unaspirated stops also turns out to be significant. Thus, the result on F0 does not support the phonetic prediction if voiced and voiceless stops share the feature [+LL].

If voiced and voiceless aspirated stops are [+LL], we may expect that the F1 values of vowels following them would be lower than those following voiceless unaspirated stops. This is because lowering the larynx will expand the cavity size between the glottis and the place of maximum constriction between the tongue and the palate. Our result shows that F1 values following voiced and voiceless aspirated stops are significantly lower than those following voiceless unaspirated stops. Specifically, there are significant differences in F1 values between vowels following voiced and voiceless unaspirated stops as well as between voiceless unaspirated and voiceless aspirated stops. In contrast, there is no significant difference in F1 values between vowels following voiced and voiceless aspirated stops, suggesting that they have the same values. Thus unlike the result on F0, the result on F1 is consistent with the phonetic prediction that voiced and voiceless aspirated stops share the feature [+LL].

Another predicted phonetic correlate of lowering the larynx is breathy voice quality. This is because when the larynx is lowered, the degree of the contact between them decreases, which subsequently could facilitate a leakage of the air from the subglottal space through the glottis. If voiced stops and voiceless aspirated stops are produced with a lowered larynx, we might expect that they have greater values in H1*-A1*, H1*-A3*, H1*-H2* and H2*-H4*, and lower CPP values than voiceless unaspirated stops. Our results indicate that there are significant differences in these voice quality measures between voiced and voiceless unaspirated stops as well as between voiceless aspirated and voiceless unaspirated stops. In contrast, no significant differences are found between voiced and voiceless aspirated stops in those measures, indicating that they have similar values. Thus, in general voiced and voiceless aspirated stops show greater values in H1*-A1*, H1*-A3*, H1*-H2* and H2*-H4*, and lower CPP values than voiceless unaspirated stops and the results are consistent with the phonetic prediction if voiced stops in particular are [+LL], but the problem is that it is unlikely that voiceless aspirated stops would also be produced with a lowered larynx.

6.4.3 Which feature is more plausible: [ATR] or [LL]?

In sections 6.4.1 and 6.4.2 above we assess whether the feature [ATR] or [LL] is a plausible feature shared by voiced and voiceless aspirated stop and therefore might be responsible for triggering the CV harmony in Madurese. Our data show that from seven acoustic parameters, namely F1, F2, H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP we evaluate as predicted acoustic correlates for the feature [+ATR], only F2 is not consistent with the phonetic prediction if voiced and voiceless aspirated stops are both [+ATR]. However, recall that previous studies (Fulop *et al.*, 1998; Guion *et al.*, 2004; Jacobson, 1980) also differ with respect to the effect of ATR on F2, suggesting that F2 may not be a reliable correlate of ATR. Similarly, from seven acoustic parameters, namely F0, F1, H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and CPP we assess as acoustic correlates for the feature [+LL], only F0 is not in line the prediction if voiced and voiceless aspirated stops share the feature [+LL]. Thus, the results are somewhat problematic because not all of the acoustic properties fit nicely with either the phonological feature [+ATR] or the feature [+LL] for becoming the possible feature which might be shared by voiced and voiceless aspirated stops.

Furthermore, if we consider the fact that voiceless aspirated stops in Madurese are synchronically voiceless during the closure, this may also lead us to question whether they are produced with an advanced tongue root at all. In fact, there is evidence that only voiced stops are produced with a tongue root advancement (Perkell, 1969). If this were the case, the feature [+LL] would become a relatively stronger candidate. One issue that may arise with the feature [LL] is that it has not been widely proposed as an active phonological feature in many languages. And more importantly, the fact that the F0 following voiceless aspirated stops is higher than that following voiced and voiceless unaspirated stops goes against the expected acoustic prediction of lowering the larynx. As a comparison, the F0 following lax stops in Javanese which have been suggested to be produced with a lowered larynx is consistently lower than that following tense stops which are articulated with a raised larynx (Brunelle, 2010; Hayward, 1995; Thurgood, 2004).

The question is how we should go about determining which feature is more plausible given the proposed features of [ATR] and [LL] do not fit in well with all the phonetic boxes, assuming a concrete, transparent phonetics-phonology mapping. In other words, the results would be problematic for the school of thought that subscribes to the idea that phonological features have acoustic correlates associated with them or that phonetic properties are predictable from features themselves (Archangeli & Pulleyblank, 1994; Hayes, Kirchner, & Steriade, 2004).

However, following most phonologists and phoneticians who allow for a more flexible phonetics-phonology mapping, we would argue that the phonetics-phonology mapping should allow us to propose that other non-phonetic factors can also be used to explain phonological phenomena. For example, Hyman (2001) posits that all sound changes principally have phonetic motivation and phonology is the result of the grammaticalisation of such sound changes. He further explains that the reason why phonology tends to be phonetically grounded is because phonology has a phonetic substance. However, phonology may have no connection with phonetics and thus need not be subject to phonetic conditioning once sound changes have undergone some sort of grammaticalisation (Hyman, 2001).

A similar viewpoint is also adopted by Hayes (1999), who suggests that a phonological system which is phonetically unnatural may follow from a sequence of changes that are phonetically grounded and productive grammars are responsible for reproducing such unnatural systems so that the child acquiring the language can learn such unnatural systems. Ohala (2005, p. 34) is also sympathetic to this stance stating that abandoning 'the requirement that phonological grammars reflect the phonetic naturalness of the sound patterns in language' may be necessary. This is because there is indeed 'evidence that non-phonetic factors such as morphology and semantics also play a much more important role in speakers' conception and manipulation of sound patterns' (2005, p. 35).

Therefore, on the basis of the arguments above, it is possible to label the feature responsible for the CV harmony or the feature shared by both voiced and voiceless aspirated stops in Madurese as either [ATR] or [LL], as far as labelling is concerned. However, we may not expect that they would have acoustic manifestations that can be directly associated with such articulatory gestures although in many cases we certainly observe the phonologisation of phonetically natural patterns. This is because the phonetic evidence does not neatly support a phonetically based feature in Madurese. As discussed earlier, arguing that the relationship between features and their correlates should be more flexible would allow us to account for a phonological phenomenon on the grounds of other non-phonetic factors. As we will discuss in Section 6.5 below, there is indeed historical and loanword evidence which supports that the voiced and voiceless aspirated stops in Madurese may have developed from earlier voiced stops and may retain a 'historical' laryngeal ATR feature.

6.5 Appeals to the Historical Source and Loanword Evidence

As far as Madurese voiceless aspirated stops are concerned, there is evidence that they may have developed from Proto-Malayo-Polynesian (PMP) voiced stops (Stevens, 1966), as shown in Table 68. Similarly, Anderson (1991) suggests that voiced and voiceless aspirated stops in Madurese may come from earlier voiced stops, both of which preserve some historical secondary laryngeal feature such as

[ATR]. All examples shown below are based on Stevens (1966), which he himself bases the reconstructed forms on Dempwolff's (1938) reconstructions (volume 3).

Table 68. Words exemplifying the development of Madurese *p*^{h} from the proto-phoneme *b*

PMP	Madurese	Gloss
*bayi	p^{h}ɤji	'baby'
*dara	t^{h}ɤrɤ	'pigeon'
*daging	t^{h}ɤk^{h}iŋ	'meat'
*gatel	k^{h}ɤtəl	'itchy'
*gigi	k^{h}ik^{h}i	'tooth'

Stevens (1966) also suggests that voiced stops such as /d/ and /g/ are mostly borrowings in Madurese, for example, *gânggu* [gɤŋgu] from Malay *ganggu* [gaŋgu] 'annoy' and *dâkwa* [dɤkwa] from Arabic *daqwa* [dakwa] 'accuse'. He also mentions that a number of PMP voiced stops did not develop into Madurese voiceless aspirated stops. However, it is not clear why some voiced stops developed into voiceless aspirated stops and why some others did not. Some examples where PMP voiced stops are retained in Madurese are shown in Table 69.

Table 69. Words exemplifying the retention of proto-phoneme *b as Madurese *b*

PMP	Madurese	Gloss
*babaq	bɤbɤ	'under'
*bahu	bɤu	'smell'
*bates	bɤtəs	'border'
*berat	bɨr:ɤʔ	'heavy'
*buka	buk:aʔ	'open'
*kulambu	kalambu	'mosquito net'
*abuh	abu	'ash'
*baRat	bɤrɤʔ	'west'
*buntut	buntɔʔ	'tail'

In relation to this, Stevens (1966, p. 152) proposes two possible accounts of consonant correspondence between Javanese and Madurese. The first explanation is that there may be two PMP phonemes which might have been involved here: the proto-phoneme **b* became Javanese *w* and Madurese *b* whereas **B* became Javanese *b* and Madurese *p*^{h}. For cases where the proto-phoneme *b became Javanese *w* and Madurese *b*, Stevens provides examples, some of which can be seen in Table 70.

Table 70. Words exemplifying the proto-phoneme *b corresponding to Javanese *w* and Madurese *b*

	Javanese *w*	Madurese *b*	Gloss
*abuh	awu	abu	‘dust’
*balaŋ	walaŋ	bɤlɤŋ	‘grasshopper’
*bales	waləs	bɤlɨs	‘pay back’
*bariŋin	wariŋin	bɤriŋɛn	‘Ficus benj.’
*baruŋ	waruŋ	bɤruŋ	‘coffee shop’
*batuk	watuk	bɤtɔʔ	‘cough’
*belut	wəlut	bɨl:uʔ	‘eel’
*besi	wəsi	bɨs:ɛh	‘iron’
*bubuŋ	wuwung	bubuŋ	‘ridge-pole’
*buku	buku	bukɔh	‘joint’
*laban	lawan	labɤn	‘against’
*lubaŋ	luwang	lɔbɤŋ	‘hole’
*sabaq	sawah	sabɤ	‘wet rice field’

On the other hand, for cases where the proto-phoneme *B became Javanese *b* and Madurese *pʰ*, Stevens also provides examples, some of which are listed in Table 71.

Table 71. Words exemplifying the proto-phoneme **B* corresponding to Javanese *b* and Madurese *pʰ*

	Javanese *b*	Madurese *pʰ*	Gloss
*Bagus	bagus	pʰɤkʰus	‘good’
*Baraŋ	baraŋ	pʰɤrɤŋ	‘thing’
*Bantu	bantu	pʰɤntɔh	‘help’
*Bantal	bantal	pʰɤntal	‘pillow’
*Baris	baris	pʰɤris	‘line’
*Bawaŋ	bawaŋ	pʰɤbɤŋ	‘onion’
*Bener	bənər	pʰɤndɤr	‘true’
*Biru	biru	pʰiruh	‘green’
*BuDu(q)	bodo	pʰuʈʰuh	‘stupid’
*BuTak	buʈak	pʰuʈak	‘bald’
*iBu	ibu	ɛpʰuh	‘mother
*reBut	rəbut	rəpʰ:uʔ	‘struggle for’
*teBang	təbaŋ	təpʰ:ɤŋ	‘cut’
*teBu	təbu	təpʰ:uh	‘sugarcane’
*teBus	təbus	təpʰ:us	‘ransom’

The second explanation Stevens (1966) suggests is that there may be only one proto-phoneme involved in this case. It is likely that the regular reflexes of standard Javanese *w* and standard Madurese *b* have been obscured by borrowings that have taken place between these two languages (although the direction of borrowings is not always clear). Furthermore, recent borrowings from Malay or Indonesian for words containing voiced stops in both word-initial and word-medial positions clearly follow the pattern that voiceless aspirated stops in Madurese may have been introduced by borrowing the prevoiced and/or breathy stops of loanwords as voiceless aspirated stops. Stevens (1966, p. 152) suggests that this is particularly the case for borrowing Javanese *b* as Madurese *pʰ* since Madurese listeners may hear the Javanese lax [b̥][16] which is voiceless and breathy, as a voiceless aspirated stop.

Two questions can be raised with regard to the proposal. The first question is why the same tendency also applies to the words borrowed into Madurese from other languages such as Indonesian (see Table 72). We know that unlike Javanese, which distinguishes between tense and lax stops, Indonesian has a two-way contrast distinguishing between voiced and voiceless unaspirated stops. The second question is why Madurese appears to have the tendency to change only this particular voicing category. That is, why the tendency to change voicing categories only applies to voiced stops and not to other stop types such as voiceless unaspirated and voiceless aspirated stops. In addition, there are also examples in which the preference for voiced stops becoming voiceless aspirated stops can also be seen in a number of words which appear to be potentially cognates in Indonesian, Javanese and Madurese (see e.g. Adelaar, 2005a; Dempwolff, 1938; Stevens, 1966). Some examples are also shown in Table 72 below.

Table 72. Words exemplifying Madurese voiceless aspirated stops derived from loanwords

Indonesian	Javanese	Madurese	Gloss
babat	b̥ab̥at	pʰɤpʰɤt	‘tripe’
bakal	b̥akal	pʰɤkal	‘candidate’
cagak	cagak	cakʰɤʔ	‘support’
dadar	d̥ad̥ar	tʰɤtʰɤr	‘omelet’
galah	galah	kʰɤlɤ	‘pole’
garap	garap	kʰɤrɤp	‘cultivate’
gɔrɛŋ	gɔrɛŋ	kʰuriŋ	‘fry’
gusi	gusi	kʰusɛ	‘gum’
guru	guru	kʰuru	‘teacher’
tətaŋga	tətaŋga	tataŋkʰɤ	‘neighbour’

Table 72 shows that voiced stops in Indonesian and lax stops in Javanese correspond to voiceless aspirated stops in Madurese. It is also evident that Madurese does not simply change voiced or lax stops into voiceless aspirated stops; vowels that follow

---

[16] Following the symbol suggested by Maddieson and Ladefoged (1996).

them also change to conform to the CV co-occurrence restriction rule which in this case requires voiceless aspirated stops to be followed by high vowels.

Some possible examples of loanwords relating to Javanese would be clear if we look at a number of Madurese higher register vocabulary shown in Table 73 below. It has been suggested that Madurese may have borrowed most of the words of this type from Javanese since they appear to be very similar (Stevens, 1968, p. 1). However, it is important to note that although some Madurese words in the high speech level may have been borrowed from Javanese, they also have undergone phonological adjustment to fit in with the phonological systems of Madurese.

Table 73. Correspondence between Javanese and Madurese high speech levels

Javanese	Madurese	Gloss
b̥abar	pʰɤpʰɤr	'give birth'
d̥ahar	ʈʰɤʔɤr	'eat'
d̥aləm	tʰɤlɨm	'house'
manawi	manabi	'if'
panɟ̥ənəŋan	pancʰɤnəŋan	'you'
rawuh	rabu	'come'
sare	sarɛ	'sleep'
sampun	sampɔn	'already'
sirɔ	sɛra	'head'

Table 73 shows some examples of the similarity between Javanese and Madurese high speech levels and how phonological adjustment also takes place in Madurese in order to conform to its own system. Here we can see that lax stops such as /d̥/ and /ɟ̥/ in Javanese become voiceless aspirated stops /tʰ/ and /cʰ/ in Madurese. We also see that there is a change in vowel quality as well and such a change follows the consonant-vowel interaction condition. That is, all voiced and voiceless aspirated stops are only followed by high vowels. In addition, there is one interesting thing taking place with respect to Javanese /w/. This segment consistently becomes /b/ when borrowed into Madurese.

It is important to note that even if it is true that in general voiced or lax stops in borrowed words become voiceless aspirated stops in Madurese, it does not entirely mean that all words containing voiceless aspirated stops may come from loanwords. As discussed earlier, there is evidence that voiceless aspirated stops in Madurese may have developed from the proto-language. Therefore, it is probably a matter of a historic coincidence that Madurese has a preference for changing some PMP voiced stops into voiceless aspirated stops. This innovation is clearly unique to Madurese since it is not observed in any of its other related languages such as Sundanese, Javanese and Indonesian.

It is important to mention, however, that Maranao, another Austronesian language spoken on the island of Mindanao, the Philippines, has also been suggested to have a three-way contrast among its stop and a relatively similar CV co-occurrence restriction (Lobel, 2010; Lobel & Riwarung, 2009). However, the extent to which Madurese and Maranao have similarities in their phonologies with respect to this still requires further research. This is due to the fact that to my knowledge there has been no information about phonetic studies of the three-way contrast in Maranao and its vowels, making it difficult to compare Madurese and Maranao directly.

Other important information about Maranao to this date is concerned with the fact that voiced stops which pattern with the voiced consonants can co-occur with either high or non-high vowels (Lobel, 2010; Lobel & Riwarung, 2009). This suggests that the CV co-occurrence restriction for voiced stops is not obligatory in Maranao and is therefore different from Madurese, whereby voiced and voiceless aspirated stops only co-occur with high vowels and the other consonants with non-high vowels.

In this regard, Lobel (2010, pp. 278–279) suggests that the reason why voiced stops in Maranao only show an optional raising effect on the following vowel may be related to the fact that consonants in the two languages develop from quite different historical sources. Unlike Madurese voiceless aspirated stops that apparently develop from proto-voiced stops, Maranao ‘heavy’ stops, which obligatorily trigger the following vowel raising, derive from earlier consonant clusters *bp, *dt, *gk, and *ds while its voiced cognates develop from earlier singletons *b, *d, *g.

The fact that voiced stops in Maranao can be followed by either high or non-high vowels are also interesting to look at. However, as the data on Maranao has not yet covered acoustic information about those segments, it is difficult to decide whether the so-called voiced stops in the language are phonetically voiced. Since the voiced stops can be followed by both vowel types, it may be interesting to also take a look at the contexts in which they occur. In addition to the historical explanation mentioned above, there may be certain phonological contexts such as vowel type and syllable structure that may condition the voiced stops to select high vowels instead of non-high vowels or the other way around. It may be the case that such variations also depend on whether the words are native Maranao or borrowed from other languages in the area. Again, these possibilities and others require further research.

In relation to our discussion above, we could therefore hypothesise that a plausible pathway by which modern Madurese contrast (voiced, voiceless unaspirated and voiceless aspirated stops) may have developed from initially two *p and *b phonemes in PMP, similar to what Stevens (1966) suggests. In this scenario, *b became breathy, triggering raising of the following vowel and this proto-phoneme later developed into modern Madurese voiced cognates. Later borrowings from Javanese, where b̥ is crucially voiceless but breathy, were probably borrowed as a kind of voiceless breathy stops but with phonologisation in vowel raising. This may also explain why the so-called voiceless aspirated stops in Madurese are not really all that aspirated compared with other languages such as Thai.

6.6 Madurese Vowel System

There has been a disagreement with respect to the number of vowel phonemes in Madurese. The disagreement has arisen partly from the fact that some researchers identify and describe Madurese vowels on the basis of surface realisation rather than based on Madurese phonology. In this thesis, we argue that Madurese is more economically described as a language with an underlying four-vowel system consisting of /ɛ, ə, a, ɔ/. If we also consider the vowels i and u as phonemes, this would create problems for the account of the vowel harmony processes and analysis of the onsets, as has been discussed in Section 1.2. That is, it simplifies the analysis of the consonants but complicates that of the vowels. Moreover, it is not clear whether the way the words that contain the vowels are pronounced reflect Madurese or simply the language from which the words in question have been borrowed instead. In this case, it would be reasonable to assume that they are pronounced in the way Indonesian words are pronounced given that, as mentioned in Chapter 2, many Madurese people also speak Indonesian.

To my observation, Madurese people rarely change Indonesian words to make them conform to the CV interaction rule when they speak in Madurese. This is particularly the case for Indonesian words borrowed from foreign languages such as Dutch and English. This may be related to the fact that Indonesian is considered to be more prestigious compared to Madurese due to its status as the national language. Thus, if they pronounce Indonesian words in the way native Madurese words are normally pronounced, they may feel the risk of being considered as having ‘low education’.

This is obvious when we have a look again at words which show exceptions to the general rule of the CV co-occurrence restriction or vowel raising in (25) below.

(25)	[bal]	‘ball’	[mɔgɔʔ]	‘strike’
	[ban]	‘tyre’	[ɔbat][17]	‘medicine’
	[baŋ]	‘bank’	[pɛnsiun]	‘retired’
	[baŋku]	‘bench’	[piŋpɔŋ]	‘Ping-Pong’
	[bɛcaʔ]	‘trishaw’	[pɔlisi]	‘police’
	[bijasa]	‘usual’	[ranɟaŋ]	‘bed’
	[buku]	‘book’	[rɔmbɛŋ]	‘old clothes’
	[dasi]	‘tie’	[rɔmbɔŋan]	‘group’
	[dɔktər]	‘doctor’	[sandal]	‘sandal’
	[dɔmpɛt]	‘wallet’	[satrika]	‘iron’
	[əmba]	‘grandparent’	[susu]	‘milk’
	[gaŋ]	‘alley’	[tabraʔ]	‘hit’
	[gas]	‘gasoline’	[taksi]	‘taxi’
	[kiblat]	‘facing Mecca’	[tɔpi]	‘hat’
	[kɔpi]	‘coffee’	[udur]	‘hindrance’

It appears that none of these words are native Madurese. However, it is interesting that a small number of these words apparently also have native Madurese counterparts that do conform to the rule, for example [bukɔ] ‘joint’ vs. [buku] ‘book’, [bɤn] ‘and’ vs. [ban] ‘tyre’ and [sɔsɔ] ‘breast’ vs. [susu] ‘milk’. However, except the word for ‘breast’, they are not semantically related.

---

[17] Some speakers pronounce this word as [ɔpʰɤt], which follows the general rule.

As has been discussed in Section 2.3.2 of Chapter 2, the disagreement with regard to the number of vowel phonemes in Madurese partly arises from the fact that some authors also include vowels from loanwords into Madurese vowel inventory. For example, Davies (2010) argues that since [i] and [u] can also be found in word-initial position in a number of words such as [imigrasi] ‘immigration’ and [uɟiɤn] ‘exam’, these vowels need to be incorporated into Madurese phonemes as well. The question is whether it is necessary to include them as phonemes given that they can only be found in loanwords in that position.

Indeed, there would be a price to pay for including the vowels [i] and [u] as phonemes. This is because it would be difficult to explain the existence of the two vowels on the grounds of the vowel raising rule or the CV co-occurrence restriction, making the rule more complicated than it needs to be. Therefore, it would be more parsimonious if we simply put the words that contain [i] and [u] in word-initial position into exceptions due to loanwords rather than categorise them as separate phonemes. Again, this needs to be done in this way if we prefer maintaining the vowel raising rule across the board in Madurese.

With regard to the vowels [ɨ] and [ə], about which previous scholarship has also questioned, we can establish that these two vowels are acoustically distinct both in terms of their F1 and F2 values. The results provide further phonetic evidence of the existence of the high vowel [ɨ] along with its non-high counterpart [ə]. This suggests that the vowel [ɨ] does not simply exist for convenience in the sense that every non-high vowel must have its high counterpart due to vowel height alternation under the process of vowel raising and/or lowering.

Thus, unless we take the phonology of Madurese into account particularly on how consonants interact with vowels, we may be led to conclude that Madurese, for instance, can be categorised into a language with a relatively symmetric eight-vowel system. However, such a conclusion also makes sense given that all of the vowels are phonetically distinct in the sense that they relatively occupy their own vowel space. This is especially obvious if we look at the five peripheral vowels, i.e. [i, ɛ, a, ɔ, u] although the three central vowels [ə, ɤ, ɨ] appear to be clustered closely together. Finally, it is also interesting to observe that the magnitude of the vowel raising for each vowel pair is also variable. This may suggest that the effect of consonantal feature spreading, whatever the feature is, depends on individual vowels following the consonants. It appears that the highest degree in vowel raising occurs to the pairs [a ~ ɤ], [ɛ ~ i] and [ɔ ~ u] respectively while the lowest occurs to the pair [ə ~ ɨ].

There are some interesting things that we can observe about the vowel system in Madurese particularly if we relate the Madurese system to vowel dispersion theory proposed by Liljencrants and Lindblom (1972). That is, considering Madurese only has four underlying vowels, why the vowels are not dispersed as the theory predicts. Specifically, as we argue for a four-vowel system in Madurese, we should expect the vowels to include the predicted /i, ɛ, a, u/ (Liljencrants & Lindblom, 1972, p. 845). This is not the case for Madurese as its vowel system only consists four underlying vowels which are all non-high, i.e. /ɛ, a, ə, ɔ/. This Madurese system is not observed in any four-vowel systems because all languages that belong to the four-vowel system always include the vowel /i/ as one of their vowels (Becker-Kristal, 2010; Liljencrants & Lindblom, 1972). In addition, the clustering together of the three central vowels [ə, ɤ, ɨ] in a relatively crowded space seems to be inconsistent with one important principle of dispersion theory that vowels have to be maximally dispersed from one another (Liljencrants & Lindblom, 1972).

It may be that the three Madurese vowels do not need to be maximally dispersed for their contrast because, as we have discussed in Chapter 2, they have different syllable structure in the case of the vowels [ə, ɨ] versus [ɤ], i.e. the former are always followed by geminates while the latter is not. On the other hand, the vowel [ɨ] is always preceded by a voiced or voiceless aspirated stop while the vowel [ə] always goes together with voiceless unaspirated stops and the other consonants. Thus, we can speculate that these non-vocalic aspects may also function to maximise the perceptual differences between the three vowels.

7 Conclusion and Directions for Further Studies

7.1 Conclusion

The thesis has addressed three main related questions. The first question asks about what acoustic property (or properties) do voiced and voiceless aspirated stops share (if any) in comparison with voiceless unaspirated stops. The second question is, in relation to the proposed features [ATR] and [LL], which more plausibly phonological feature might be responsible for the patterning together of voiced and voiceless aspirated stops to the exclusion of voiceless unaspirated stops in the CV co-occurrence restriction. The third question is what the implications of the results of the acoustic study are for phonetics-phonology mappings. That is, whether the results imply a more concrete, transparent phonetics-phonology mapping or a flexible phonetics-phonology one. To answer the questions, we examined a number of acoustic measures, which include voice onset time (VOT), fundamental frequency (F0), closure duration, frequencies of the first two formants (F1 and F2) and spectral measures (i.e. H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP).

With regard to the first question, i.e. whether voiced and voiceless aspirated stops share acoustic features to the exclusion of voiceless unaspirated stops, our results indicate that they do to some extent. Specifically, voiced and voiceless aspirated stops in Madurese synchronically share some acoustic properties such as H1*-A1* for both genders, H1*-H2* for females, H1*-A3* and H2*-H4* for males, and CPP for females at vowel onset and for males at vowel midpoint. However, they do not share acoustic properties such as VOT, closure duration and F0. In contrast, voiceless unaspirated and voiceless aspirated stops in Madurese can be distinguished by VOT (Section 5.2), F0 (Section 5.4) and spectral measures, i.e. H1*-A1* (Section 5.5), H1*-A3* (Section 5.7), H1*-H2* (Section 5.8), H2*-H4* (Section 5.9) and CPP (Section 5.10). However, they show similarity in closure duration (Section 5.3).

The fact that we found a number of gender-related findings with regard to spectral measures in particular is interesting. In this case, it is not clear why there are some differences. However, cases such as these are not uncommon because a number of studies which deal with voice quality measures also find similar cases where certain voice quality differences can be distinguished by certain spectral measures for females, but not for males or the other way around (see e.g. Esposito, 2010b; Wayland & Jongman, 2001; Abramson *et al.* 2007). The explanations they suggest are related to either physiological factors or sociolinguistic factors. This can be an interesting area of future research on Madurese.

With regard to the second question, i.e. whether voiced and voiceless aspirated stops might share an articulatorily grounded feature to the exclusion of voiceless unaspirated stops, we also looked at the results of acoustic analyses. Our results indicate that at a first consideration, either ATR or LL appears to be a plausible phonological feature which may account for the consonant-vowel interactions or feature spreading in Madurese. This is because there is phonetic evidence consistent with both types of articulations. Acoustic evidence that lends support to the feature [+ATR] includes lower F1 and greater spectral tilt measures including H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and lower CPP values. Acoustic evidence that supports the feature [+LL] also includes lower F1 and greater spectral tilt measures including H1*-A1*, H1*-A3*, H1*-H2*, H2*-H4* and lower CPP values. However, the fact that voiceless aspirated stops are synchronically voiceless during closure presents a problem for the feature [+ATR], and the fact that F0 of vowels following voiceless aspirated stops is higher than following voiced stops also presents a problem for the feature [+LL]. This is because if voiceless aspirated stops are produced with a lowered larynx, we would expect they also have lower F0.

These findings challenge the idea that the relationship between phonology and phonetics is transparent in the sense that phonological rules are expected to have a clear phonetic basis (Archangeli & Pulleyblank, 1994; Hayes, 1999; Hayes, Kirchner, & Steriade, 2004). This is because not all acoustic correlates fit in nicely with either the feature. Following the view that not all phonological features may not be expected to be phonetically grounded, especially when they are related to possible sound changes, we argue that Madurese provides further evidence that the phonetics-phonology mapping should be, or can be flexible. That is, we allow for other non-phonetic factors to account for a phonological phenomenon such as the one we observe in Madurese. The fact that voiced and voiceless aspirated stops pattern together in triggering vowel raising and other CV harmony processes in Madurese suggests that they share some phonological feature, whatever the feature may be called. We have provided historical and loanword evidence which supports the idea that voiceless aspirated stops may have derived from earlier voiced stops, and that they may retain their historical laryngeal contrast through phonologisation.

As argued in Section 1.2, Madurese can be more parsimoniously described as a language with a three-way laryngeal contrast: voiced, voiceless unaspirated and voiceless aspirated stops. Thus, this rules out the possibility that Madurese has a two-way contrast in their stops although the occurrences of voiceless unaspirated and voiceless aspirated stops appear to be environment-dependent. It appears that a constellation of acoustic properties works in concert to distinguish between the three voicing categories. In order to establish if the spectral measures (i.e. F0, H1*-A1*, H1*-A2*, H1*-A3*, H1*-H2*, H2*-H4* and CPP) contribute something to the voicing distinction, we conducted a linear discriminant analysis. Our result shows that they contribute about 30 percent to the discrimination for voiced, voiceless unaspirated and voiceless aspirated stops. A further discriminant analysis which includes both the spectral measures and VOT shows that together they contribute around 84 percent to the voicing distinction in Madurese. This suggests that VOT does most of the work in distinguishing the three voicing categories (despite the considerable overlap in VOT for voiceless unaspirated and voiceless aspirated stops).

In terms of vowels, Madurese can be more parsimoniously described as a language with a four-vowel system, where all of the vowels constitute non-high vowels. This system is typologically unusual because no languages are known to have only non-high vowels in their inventories (Becker-Kristal, 2010). Due to the fact that the system accounts for the vast majority of the lexical items, we consider the system with a three-way laryngeal contrast and four underlying vowels as best describing the system in Madurese. Furthermore, we can also establish that the existence of eight surface vowels is robust, as indicated by the acoustic space each has in the vowel space. Thus, this also has settled the ambiguity surrounding their phonetic realisations particularly with regard to the three central vowels (ə, ɨ and ɤ).

7.2 Directions for Further Studies

There are a number of areas which require further research with regard to the results of the present study. Those areas can include both production and perception studies. In terms of production studies, we may only focus on a certain dialect of Madurese and include speakers of different age categories that may represent different generations, for example, young and old speakers. This type of research is crucial to understand whether young and old speakers, for example, have differences in the way they may realise phonetic correlates of phonological contrasts such as VOT, F0, H1-H2, H1-A3 and CPP. In addition, examining possible age-related differences in phonetic realisations of phonological features can also contribute to our understanding about whether sound change is also in progress in Madurese. Based on the results of the production study, a perception study can be conducted to find out which acoustic correlates are perceptually relevant cues that listeners use to distinguish phonological contrasts. This study should also include gender as a variable and involve speakers with a balanced number of female and male speakers.

Another area of phonetic research that can be further pursued may involve other instrumental techniques such as electroglottography and palatography. Glottography is particularly crucial for our understanding about the glottal states during the production of the three stop categories in Madurese. The results of this study, for example, can provide more definitive information with regard to the results of the acoustic study showing that voiceless aspirated stops have higher H1-H2, suggesting an more open vocal fold condition. Palatography is also important because it will further provide more accurate information about the place status of certain stops which have been variably described as dental, alveolar and retroflex, for instance.

Another important area of studies which can be pursued is some prosodic aspects of Madurese. This may begin by looking at word-level stress. This study is important because, as far as Madurese is concerned, no phonetic studies deal with this phenomenon. The results of the study can provide information about word stress and its acoustic correlates in Madurese. They can also provide typological information as Indonesian and Betawi Malay have been suggested as stress-free languages.

8 References

Abdelli-Beruh, N. B. (2004). The stop voicing contrast in French sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure duration. *Phonetica*, *61*, 201–219.

Abdelli-Beruh, N. B. (2009). Influence of place of articulation on some acoustic correlates of the stop voicing contrast in Parisian French. *Journal of Phonetics*, *37*, 66–78.

Abramson, A. S., & Lisker, L. (1971). Voice timing in Korean stops. Presented at *the Seventh International Congress of Phonetic Sciences*, Montreal.

Abramson, A. S., Nyea, P. W., & Luangthongkum, T. (2007). Voice register in Khmu' Experiments in production and perception. *Phonetica*, *64*, 80–104.

Adelaar, K. A. (2005a). Malayo-Sumbawan. *Oceanic Linguistics*, 357–388.

Adelaar, K. A. (2005b). The Austronesian languages of Asia and Madagascar: a historical perspective. In Adelaar, K. A. and Himmelmann, N. (eds.) *The Austronesian languages of Asia and Madagascar* (vol. 7) (pp. 1-42). New York: Routledge

Alghamdi, M. (1990). *Analysis, synthesis and perception of voicing in Arabic* (Doctoral Dissertation). University of Reading.

Ananta, A., Arifin, E. N., Has bullah, M. S., Handayani, N. B., & Pramono, A. (2015). *Demography of Indonesia's ethnicity*. Institute of Southeast Asian Studies.

Anderson, S. (1991). Vowel-consonant interaction in Madurese. In Ann, J. and Yoshimora, K. (eds.) *Coyote Papers, Arizona Phonology Conference* (vol. 4) (pp.1-15).

Anderson, S. R. (1981). Why phonology isn't "natural." *Linguistic Inquiry*, 493–539.

Andruski, J. E., & Ratliff, M. (2000). Phonation types in production of phonological tone: the case of Green Mong. *Journal of the International Phonetic Association*, *30*, 37–61.

Archangeli, D. B., & Pulleyblank, D. G. (1994). *Grounded phonology* (Vol. 25). MIT Press.

Avery, P., & Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In Hall, T. A. (Ed.), *Distinctive feature theory* (Vol. 2) (pp. 41-70). Berlin: Mouton de Gruyter.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bach, E., & Harms, R. T. (1972). How do languages get crazy rules? In Stockwell, R and Macaulay, R. (eds.), *Linguistic Change and Generative Theory* (pp. 1-21). Bloominton: Indiana University Press.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Dai, B. (2014). Package "lme4." *R Foundation for Statistical Computing,* Vienna.

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and Dispersion Theory: insights from a large cross-linguistic corpus* (PhD Dissertation). University of California, Los Angeles.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: evidence for phonological oversimplification. *Journal of Phonetics*, *39*, 39–49.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, 1–26.

Benguerel, A.-P., & Bhatia, T. K. (1980). Hindi stop consonants: an acoustic and fiberscopic study. *Phonetica*, *37*(3), 134–148.

Benki, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, *29*, 1–22.

Blankenship, B. (1997). *The time course of breathiness and laryngealization in vowels* (PhD Dissertation). UCLA.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, *30*, 163–191.

Blevins, J. (2004). *Evolutionary phonology: the emergence of sound patterns*. Cambridge University Press.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound (Vol. 17, pp. 97–110). Presented at *the Proceedings of the Institute of Phonetic Sciences*, Amsterdam.

Boersma, P., & Weenink, D. (2014). {P}raat: doing phonetics by computer. Retrieved from www.praat.org

Bortscheller, M. (2007). *Madurese phonetic facts*. Unpublished manuscript, University of Iowa.

Bradley, D. (1982). Register in Burmese. *Pacific Linguistics. Series A. Occasional Papers*, (62), 117.

Brunelle, M. (2010). The role of larynx height in the Javanese tense~lax stop contrast. *Austronesian and Theoretical Linguistics, Amsterdam: John Benjamins*, 7–24.

Casali, R. F. (2008). ATR harmony in African languages. *Language and Linguistics Compass*, *2*(3), 496–549.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Midland Books.

Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, *39*, 612–6125.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.

Cho, T., Ladefoged, P., & Jun, S.-A. (2002). Acoustic and aerodynamic correlates of Korean stops. *Journal of Phonetics*, *30*, 193–228.

Clements, G. N., & Hallé, P. (2010). Phonetic bases of distinctive features: introduction. *Journal of Phonetics*, *1*(38), 3–9.

Cohn, A. C. (1993b). Consonant-vowel interaction in Madurese: the feature lowered larynx. *Chicago Linguistics Society*, *29*, 105–119.

Cohn, A. C. (1993a). Voicing and vowel height in Madurese: a preliminary report. *Oceanic Linguistics Special Publication*, 107–121.

Cohn, A. C., & Ham, W. H. (1998). Temporal properties of Madurese consonants: a preliminary report. *Working Papers of the Cornell Phonetics Laboratory*, *12*, 27–51.

Cohn, A. C., & Lockwood, K. (1994). A phonetic description of Madurese and its phonological implications. *Working Papers of the Cornell Phonetics Laboratory*, *9*, 67–92.

Davies, W. D. (2010). *A grammar of Madurese* (Mouton grammar library, 50). Walter de Gruyter.

Dempwolff, O. (1938). *Vergleichende lautlehre des austronesischen Wortschatzes, Band 3: Austronesisches Worterverzeichnis (19. Beiheft zur Zeitschrift fur Eingeborenen-Sprachen)*. Berlin, Hamburgh: D. Reimer.

Denning, K. (1989). *The diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages* (PhD Dissertation). Stanford University, Ann Arbor, MI.

DiCanio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, *39*(2), 162–188.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95.

Dutta, I. (2007). Four-way stop contrasts in Hindi: an acoustic study of voicing, fundamental frequency and spectral tilt (PhD Dissertation). University of Illinois at Urbana-Champaign.

Dyen, I. (1965). *A lexicostatistical classification of the Austronesian languages* (Vol. Memoir 19). Baltimore: Waverly Press.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress Laryngoscopic case studies, *23*(2), 157–191.

Esposito, C. M. (2004). Santa Ana del Valle Zapotec Phonation. *UCLA Working Papers in Phonetics*, *103*, 71–105.

Esposito, C. M. (2010a). The effects of linguistic experience on the perception of phonation. *Journal of Phonetics*, *38*, 306–316.

Esposito, C. M. (2010b). Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association*, *40*(2), 181–-198.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*, *40*, 466–476.

Esposito, C. M., & Khan, S. ud D. (2012). Contrastive breathiness across consonants and vowels: a comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association*, *42*(2), 123–143.

Fagan, J. L. (1988). Javanese intervocalic stop phonemes: the light/heavy distinction. *Studies in Austronesian Linguistics. Ohio University Center for International Studies.*, 173–200.

Fischer-Jørgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics*, *28*, 71–139.

Fischer-Jørgensen, E. (1968). Voicing, tenseness and aspiration in stop consonants, with special reference to French and Danish. *Annual Report of the Institute of Phonetics, University of Copenhagen*, *3*, 63–114.

Fulop, S. A., Karib, E., & Ladefoged, P. (1998). An acoustic study of the tongue root contrast in Degema vowels. *Phonetica*, *55*, 80–98.

Gao, J. (2015). *Interdependence between tones, segments and phonation types in Shanghai Chinese* (PhD Dissertation). Universite Sorbonne Nouvelle-Paris 3.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*(02), 185–205.

Goldstein, U. G. (1980). An articulatory model for the vocal tracts of growing children.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406.

Gósy, M. (2001). The VOT of the Hungarian voiceless plosives in words and in spontaneous Speech. *International Journal of Speech Technology*, *4*(1), 75–85.

Gósy, M., & Ringen, C. (2009). Everything you always wanted to know about VOT in Hungarian. Presented at *the 19th International Conference on the Structure of Hungarian*, Debrecen, Hungary.

Guion, S. G., Post, M. W., & Payne, D. L. (2004). Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics*, *32*, 517–542.

Halle, M., & Stevens, K. (1969). On the feature “advanced tongue root.” *Quarterly Progress Report of the MIT Research Laboratory in Electronics*, *94*, 209–215.

Halle, M., & Stevens, K. N. (1971). *A note on laryngeal features* (Quarterly Progress Report of the Research Laboratory of Electronics 101) (pp. 198–213).

Han, M. S., & Weitzman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/, *Phonetica, 22(2)*, 112–128.

Hanson, H. M. (1997). Glottal characteristics of female speakers Acoustic correlates. *Journal of the Acoustical Society of America*, *101*(1), 466–481.

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, *106*(2), 1064–1077.

Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., & Slifka, J. (2001). Towards models of phonation. *Journal of Phonetics*, *29*, 451–480.

Harris, J. (1994). *English sound structure*. Blackwell Oxford.

Hayes, B., Kirchner, R., & Steriade, D. (2004). *Phonetically based phonology*. Pearson Education.

Hayes, B. P. (1999). Phonetically driven phonology. *Functionalism and Formalism in Linguistics*, *1*, 243–285.

Hayward, K. M. (1995). /p/ vs. /b/ in Javanese: the role of the vocal folds. *Working Papers in Linguistics and Phonetics*, (5), 1–11.

Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics*, *36*, 607–628.

Henderson, E. J. (1952). The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies*, *14*(01), 149–174.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, *37*, 769–778.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech, Language, and Hearing Research*, *38*(6), 1212–1223.

Hombert, J. M. (1975). *Towards a theory of tonogenesis: an empirical, physiologically and perceptually based account of the development of tonal contrasts in languages* (Doctoral Dissertation). University of California, Berkeley.

Hombert, J. -M. (1978). Consonant types, vowel quality, and tone. In Fromkin, V. A. (Ed.), *Tone: A Linguistic Survey* (pp. 77-112). New York: Academic Press.

Hombert, J. -M., & Ladefoged, P. (1976). Effect of aspiration on the fundamental frequency of the following vowel. *The Journal of the Acoustical Society of America*, *59*(S1), S72–S72.

Hombert, J. -M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 37–58.

Honda, K. (1983). Relationship between pitch control and vowel articulation. *Haskins Laboratories Status Report on Speech Research, SR*, *73*, 269–82.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech*, *42*, 404–411.

Honeybone, P. (2005). Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In van Oostendorp, M., & van de Weijer, J. M. (Eds.), *The internal organization of phonological segments* (Vol. 77) (pp. 319-354). Berlin: Mouton de Gruyter.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105–113.

House, A. S., & Stevens, K. N. (1958). Estimation of formant bandwidths from measurements of transient response of the vocal tract. *Journal of Speech, Language, and Hearing Research*, *1*(4), 309–315.

Hualde, J. I., & Nadeu, M. M. (2011). Lenition and phonemic overlap in Rome Italian. *Phonetica*, *68*(4), 215–242.

Huffman, F. (1978). Synchronic evidence for the history of Khmer vowels. Presented at *the Second International Conference on Austroasiatic Languages*, Mysore, India.

Huffman, F. E. (1976). The register problem in fifteen Mon-Khmer languages. *Austroasiatic Studies*, *Oceanic Linguistics Special Publication*(13), 575–589.

Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America*, *81*(2), 495–504.

Husson, L. (1997). Eight centuries of Madurese migration to east Java. *Asian and Pacific Migration Journal*, *6*(1), 77–102.

Hyman, L. M. (2001). The limits of phonetic determinism in phonology:* NC revisited. *The Role of Speech Perception in Phonology*, 141–185.

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on* (Vol. 1, pp. I–669–72 vol. 1). IEEE.

Iseli, M., Shue, Y.-L., & Alwand, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America*, *121*, 2283–2295.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, *12*(03), 369–396.

Jacobson, L. C. (1978). *DhoLuo vowel harmony: a phonetic investigation*. Phonetics Laboratory, Department of Linguistics, UCLA.

Jacobson, L. C. (1980). Voice-quality harmony in Western Nilotic languages. *Issues in Vowel Harmony (studies in Language Companion Series, Vol. 6)*, 183–200.

Jessen, M. (1996). *Phonetics and phonology of the tense and lax obstruents in German* (PhD Dissertation). Cornell University.

Jessen, M. (2001). Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. In Hall, T. A. (Ed.), *Distinctive feature theory* (Vol. 2) (pp. 237-292). Berlin: Mouton de Gruyter.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, *19*, 189–218.

Kagaya, R., & Hirose, H. (1975). Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, *9*, 27–46.

Kang, K.-H., & Guion, S. G. (2008). Clear speech production of Korean stops: changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, *124*(6), 3909–3917.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: a corpus study. *Journal of Phonetics*, (45), 76–90.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319.

Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, *5*(2), 275–292.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, *18*(3), 321–334.

Keating, P. A., & Esposito, C. (2006). Linguistic voice quality. Presented at *the Proceedings of the Eleventh Australasian International Conference on Speech Science and Technology*.

Khan, S. ud D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics*, *40*, 780–795.

Kiliaan, H. N. (1897). *Madoereesche Spraakkunst: I Inleiding en Klankleer and II Woordleer en Syntaxis*. Batavia: Landsdrukkerij.

Kim, M.-R., & Duanmu, S. (2004). “Tense” and “lax” stops in Korean. *Journal of East Asian Linguistics*, *13*(1), 59–104.

Kingston, J. (2007). The phonetics-phonology interface. *The Cambridge Handbook of Phonology*, 435–456.

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language*, *70*(3), 419–454.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: the [voice] contrast. *Journal of Phonetics*, *36*, 28–54.

Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., & Bartels, C. (1997). Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America*, *101*(3), 1696–1709.

Kirby, J., & Brunelle, M. (in press). Southeast Asian tone in areal perspective. In *The Cambridge Handbook of Areal Linguistics*. Cambridge University Press.

Kirk, P. L., Ladefoged, P., & Ladefoged, J. (1984). Using a spectrograph for measures of phonation types in a natural language. *UCLA Working Papers in Phonetics*, *59*, 102–113.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*(2), 820–857.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *The Journal of the Acoustical Society of America*, *90*(1), 83–96.

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, *41*, 150–174.

Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: the cases of Greek and Japanese. *Journal of Phonetics*, *40*, 725–744.

Kreiman, J., Gerratt, B. R., & Khan, S. ud D. (2010). Effects of native language on perception of voice quality. *Journal of Phonetics*, *38*, 588–593.

Kulikov, V. (2012). *Voicing and voice assimilation in Russian stops* (PhD Dissertation). University of Iowa, Ann Arbor, MI.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30. https://CRAN.R-project.org/package=lmerTest.

Ladefoged, P. (1968). *A phonetic study of West African languages: an auditory-instrumental survey*. Cambridge University Press.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. University of Chicago Press.

Ladefoged, P. (1983). The linguistic use of different phonation types. *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, 351–360.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, P., & Maddieson, I. (2001). Vowels of the world's languages. *Phonology: Critical Concepts in Linguistics*, 5–39.

Laeufer, C. (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics*, *20*(4), 411–440.

Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on F0 in Taiwanese. In *Proceedings of the 2nd International Conference on East Asian Linguistics, Simon Fraser University Working Papers in Linguistics*

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals (Cambridge, Mass.)*. Cambridge: MIT press.

Lenth, R. V. (2016). Least-Squares Means: the R Package lsmeans. Journal of Statistical Software, 69(1), 1-33.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 839–862.

Lindau, M. (1978). Vowel features. *Language*, 541–563.

Lindau, M. (1979). Feature expanded. *Journal of Phonetics*, *7*(2), 163–176.

Lindau, M., & Ladefoged, P. (1986). Variability of feature specifications. *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Lawrence Erlbaum, 464–78.

Lisker, L., & Abramson, A. S. (1964). A cross language study of voicing in initial stops: acoustical measurements. *Word*, *20*, 384–422.

Lobel, J. W. (2010). Maranao and Madurese revisited: a follow-up. *Oceanic Linguistics*, *49*(1), 278–281.

Lobel, J. W., & Riwarung, L. H. S. (2009). Maranao revisited: an overlooked consonant contrast and its implications for lexicography and grammar. *Oceanic Linguistics*, *48*(2), 403–438.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America*, *85*(3), 1314–1321.

Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization* (PhD Dissertation). University of Massachusetts at Amherst.

Lombardi, L. (1995). Laryngeal neutralization and syllable wellformedness. *Natural Language & Linguistic Theory*, *13*(1), 39–74.

Lynch, J., Ross, M., & Crowley, T. (2002). *The Oceanic Languages* (Vol. 1). Psychology Press.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.

Mester, R. A., & Itô, J. (1989). Feature predictability and underspecification: palatal prosody in Japanese mimetics. *Language*, 258–293.

Mielke, J. (2008). *The emergence of distinctive features*. Oxford University Press.

Mills, T. (2010). spectralTiltMaster.praat version 0.0.5

Misnadin, M., Kirby, J., & Remijsen, B. (2015). Temporal and spectral properties of Madurese stops (Vol. 789). Presented at *the Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK.

Mohr, B. (1971). Intrinsic variations in the speech signal. *Phonetica*, *23*(2), 65–93.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association*, *24*(1), 1–18.

Ohala, J. J. (2005). Phonetic explanations for sound patterns. *A Figure of Speech: a Festschrift for John Laver*, 23–38.

Ohala, J. J., & Riordan, C. J. (1979). Passive vocal tract enlargement during voiced stops. *The Journal of the Acoustical Society of America*, *65*(S1), S23–S23.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*, *75*(1), 224–230.

Paster, M. (2004). Vowel height harmony and blocking in Buchan Scots. *Phonology*, *21*, 359–407.

Perkell, J. S. (1969). Physiology of speech production: results and implications of a quantitative cineradiographic study. MIT research monograph, No 53.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics*, *18*(3), 375–394.

Pike, K. L. (1967). Tongue-root position in practical phonetics. *Phonetica*, *17*(3), 129–140.

Pind, J. (1999). The role of F1 in the perception of voice onset time and voice offset time. *The Journal of the Acoustical Society of America*, *106*(1), 434–437.

Port, R. F., & Rotunno, R. (1979). Relation between voice onset time and vowel duration. *Journal of the Acoustical Society of America*, *66*(3), 654–662.

Quené, H. (2014). Hqmisc: Miscellaneous convenience functions and dataset. R package version 0.1-1. https://CRAN.R-project.org/package=hqmisc

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Remijsen, B., Ayoker, O. G., & Mills, T. (2011). Shilluk. *Journal of the International Phonetic Association*, *41*(01), 111–125.

Ren, N. (1992). *Phonation types and stop consonant distinctions: Shanghai Chinese* (PhD Dissertation). The University of Connecticut.

Ringen, C., & Kulikov, V. (2012). Voicing in Russian stops: cross-linguistic implications. *Journal of Slavic Linguistics*, *20*(2), 269–286.

Shimizu, K. (1996). *A cross-language study of voicing contrasts of stop consonants in Asian languages*. Seibido.

Shue, Y.-L., Keating, P. A., Vicenik, C., & Yu, K. (2011). VoiceSauce: a program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hongkong.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(02), 287–308.

Silverman, D. (1997). Laryngeal complexity in Otomanguean vowels. *Phonology*, *14*(02), 235–261.

Soetoko, Soegianto, Soekarto, & Soetarto, A. (1986). *Pemetaan Bahasa Madura di Pulau Madura*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Soetoko, Soegianto, Surani, Sariyono, & Suyanto. (1998). *Geografi Dialek Bahasa Madura*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidkan dan Kebudayaan.

Stevens, A. M. (1965). Language levels in Madurese. *Language*, 294–302.

Stevens, A. M. (1966). The Madurese reflexes of Proto-Malayopolynesian. *Journal of the American Oriental Society*, *86*, 147–56.

Stevens, A. M. (1968). *Madurese Phonology and Morphology*. New Haven: American Oriental Society.

Stevens, A. M. (1980). Formative boundary in phonological rules. In *Junctures: a collection of original papers* (pp. 135–141). Saratoga, California: Amma Libri.

Stevens, A. M. (1991). Madurese reduplication revisited. *Royal Institute of Linguistics and Anthropology*.

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica*, *34*, 264–279.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Stevens, K. N., & Hanson, H. M. (1995). Classification of glottal vibration from acoustic measurements. *Vocal Fold Physiology: Voice Quality Control*, 147–170.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced‐voiceless distinction for stops. *The Journal of the Acoustical Society of America*, *55*(3), 653–659.

Stewart, J. M. (1967). A theory of the origin of Akan vowel harmony (pp. 7–13). Presented at the Proceedings of *the Sixth International Congress on Phonetic Sciences,* Prague.

Stewart, J. M. (1967). Tongue root position in Akan vowel harmony. *Phonetica*, *16*(4), 185–204.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, *62*(2).

Thurgood, E. (2004). Phonation types in Javanese. *Oceanic Linguistics*, *43*(2), 277–295.

Thurgood, G. (2007). Tonogenesis revisited: revising the model and the analysis. *Studies in Tai and Southeast Asian Linguistics*, 263–291.

Tiede, M. K. (1996). An MRI-based study of pharyngeal volume contrasts in Akan and English. *Journal of Phonetics*, *24*(4), 399–421.

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*, *107*(6), 3438–3451.

Trigo, L. (1991). On pharynx-larynx interactions. *Phonology*, *8*(1), 113–136.

Tsuchida, A. (1997). *Phonetics and phonology of Japanese vowel devoicing* (PhD Dissertation). Cornell University.

Tucker, A. N., & Mpaayei, J. T. O. (1955). *A Maasai grammar, with vocabulary*. Longmans, Green.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: a practical guide. *Methods in Empirical Prosody Research*, *3*, 1–28.

van Heuven, V. J., Roosman, L., & van Zanten, E. (2008). Betawi Malay word prosody. *Lingua*, *118*, 1271–1287.

van Zanten, E., & van Heuven, V. J. (2004). Word stress in Indonesian: fixed or free. *NUSA: Linguistic Studies on Indonesian and Other Languages in Indonesia*, *53*, 1–20.

van Zanten, E., & van Heuven, V. J. (1998). Word stress in Indonesian: its communicative relevance. *Bijdragen Tot de Taal-, Land-En Volkenkunde*, 129–149.

Vaux, B. (1996). The status of ATR in feature geometry. *Linguistic Inquiry*, *27*(1), 175–182.

Wayland, R., & Jongman, A. (2001). Chanthaburi Khmer vowels: phonetic and phonemic analyses. *Mon-Khmer Studies*, *31*, 65–82.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics*, *31*, 181–201.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, *73*(4), 1332–1336.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*(02), 165–181.

Youssef, I. (2010). Laryngeal assimilation in Buchan scots. *English Language and Linguistics*, *14*(3), 321–345.

9 Appendices

9.1 The Wordlist

No.	Madurese Orthography	IPA	Gloss
1.	bâbâ	bɤbɤ	under
2.	bâbih	bɤbih	pig
3.	bâbun	bɤbun	traditional payment
4.	bâdhâl	bɤtʰɤl	origin
5.	bâghih	bɤkʰih	a day's name
6.	bâgi	bɤgi	divide
7.	bâghi	bɤkʰi	give
8.	bekkas	bɨk:as	second-hand
9.	betta	bɨt:a	resilient
10.	bettès	bɨt:ɛs	shin
11.	bhâbhâr	pʰɤpʰɤr	give birth
12.	bhâbhât	pʰɤpʰɤt	beef stomach
13.	bhâdhân	pʰɤtʰɤn	body
14.	bhâgus	pʰɤkʰus	good
15.	bhâjâng	pʰɤɟɤŋ	prayer
16.	bhâji'	pʰɤɟiʔ	baby
17.	bhâtah	pʰɤtah	brick
18.	bhellis	pʰɨl:is	angry
19.	bhelluh	pʰɨl:uh	foal
20.	bhengnges	pʰɨŋ:ɛs	cruel
21.	bhibhit	pʰipʰit	seed
22.	bhighâl	pʰikʰɤl	robber
23.	bhikang	pʰikaŋ	cake
24.	bhubut	pʰupʰut	pregnant
25.	bhucor	pʰucɔr	leak
26.	bhuḍhâk	pʰuʈʰɤk	bamboo bucket
27.	bhuju'	pʰuɟuʔ	ancestor
28.	bhundhu'	pʰuntʰuʔ	pack
29.	bhuṭok	pʰuʈɔk	fertiliser
30.	bibir	bibir	lip
31.	bighih	bikʰih	seed
32.	bitong	bitɔŋ	count
33.	buduh	buduh	stale
34.	buja	buɟɤ	salt
35.	bukoh	bukɔh	joint
36.	cabis	cabis	visit
37.	caca	caca	talk
38.	caghâ'	cakʰɤʔ	propeller
39.	cakang	cakaŋ	diligent
40.	cècèl	cɛcɛl	pay by installment
41.	cegghâ'	cəkʰ:ɤʔ	broken off
42.	cekka'	cək:aʔ	sticky
43.	cèṭak	cɛʈak	head
44.	cèṭo'	cɛʈɔʔ	small spade

45.	ceṭṭèk	cəʈ:ɛk	little finger
46.	coco	cɔcɔ	stab
47.	cocok	cɔcɔk	suitable
48.	cokop	cɔkɔp	enough
49.	copè'	cɔpɛʔ	narrow
50.	dâdâp	dɤdɤp	stupid
51.	dâjâh	dɤɟɤh	north
52.	dâpor	dɤpɔr	kitchen
53.	dâteng	dɤtəŋ	come
54.	dengngen	dɨŋ:ən	oblivious
55.	deppah	dɨp:ah	fathom
56.	dessa'	dɨs:aʔ	urge
57.	dhâbâ'	tʰɤbɤʔ	young catfish
58.	ḍhâbu	ʈʰɤbu	talk
59.	dhâdhâr	tʰɤtʰɤr	omellete
60.	dhâghâng	tʰɤkʰɤŋ	trade
61.	dhândhân	tʰɤntʰɤn	put on make-up
62.	dheddhel	tʰɨtʰ:ɨl	press
63.	dhegghân	tʰɨkʰ:ɤn	young coconut
64.	dherres	tʰɨr:ɨs	heavy
65.	dhibi'	tʰibiʔ	self
66.	dhikah	tʰikah	you
67.	dhisah	tʰisah	village
68.	dhudhing	tʰutʰiŋ	point
69.	dhudhul	tʰutʰul	traditional cake
70.	dhujân	tʰuɟɤn	like
71.	dhukah	tʰukah	angry
72.	dhupah	tʰupah	incense
73.	didik	didik	educate
74.	dinar	dinar	dinar
75.	dipan	dipan	bed
76.	dukar	dukar	gig
77.	dupolo	dupɔlɔ	twenty
78.	durih	durih	thorn
79.	gâgâ'	gɤgɤʔ	bold
80.	gâgân	gɤgɤn	stupid
81.	gâji	gɤɟi	salary
82.	gellas	gɨl:as	glass
83.	genna	gɨn:a	appropriate
84.	gessa	gɨs:a	chat
85.	ghâbâl	kʰɤbɤl	startled
86.	ghâbhâk	kʰɤpʰɤk	ceiling
87.	ghâjhâ	kʰɤcʰɤ	elephant
88.	ghebbhuk	kʰɨpʰ:uk	hit
89.	ghessèt	kʰɨs:ɛt	agile
90.	ghetta	kʰɨt:a	sap
91.	ghibâng	kʰibɤŋ	earring
92.	ghidhing	kʰitʰiŋ	bring along
93.	ghighih	kʰikʰih	tooth
94.	ghubâng	kʰubɤŋ	break in
95.	ghubhâr	kʰupʰɤr	finish
96.	ghubhuk	kʰupʰuk	quicklime

97.	ghudhur	kʰutʰur	withered
98.	ghughul	kʰukʰul	village name
99.	gibâs	gibɤs	sheep
100.	gilâp	gilɤp	shiny
101.	giri	giri	a place name
102.	gudâng	gudɤŋ	warehouse
103.	gudir	gudir	jelly
104.	gule	gulɛ	Madurese curry
105.	jâgâh	ɟɤgɤh	oversee
106.	jâgâl	ɟɤgɤ	slaughter
107.	jâma'	ɟɤmaʔ	plural
108.	jedding	ɟɨd:iŋ	bathroom
109.	jeddut	ɟɨd:ut	fat
110.	jellas	ɟɨl:as	clear
111.	Jhâbâh	cʰɤbɤh	Java
112.	jhâghâh	cʰɤkʰɤh	wake up
113.	jhâghung	cʰɤkʰuŋ	corn
114.	jhâjhâl	cʰɤcʰɤl	try
115.	jhejjhâl	cʰɨcʰ:ɨl	insert
116.	jhejjhek	cʰɨcʰ:ɨk	established
117.	jhemmur	cʰɨm:ur	dry out
118.	jhijhir	cʰicʰir	stand in line
119.	jhilâ	cʰilɤ	tongue
120.	jhimat	cʰimat	fetish
121.	jhuba'	cʰubɤʔ	ugly
122.	jhujhur	cʰucʰur	honest
123.	jhuko'	cʰukɔʔ	fish
124.	jidur	ɟidur	traditional drum
125.	jikar	ɟikar	bull carriage
126.	jirân	ɟiran	neighbour
127.	juju'	ɟuɟuʔ	grand grandparent
128.	junan	ɟunan	I
129.	jutah	ɟutah	Million
130.	kabhâr	kapʰɤr	news
131.	kacèr	kacɛr	left
132.	kacong	kacɔŋ	kid
133.	kadhih	katʰih	such as
134.	kagit	kagit	water apple
135.	kajuh	kaɟuh	wood
136.	kapeng	kapɛŋ	pocket
137.	kapor	kapɔr	chalk
138.	kata'	kataʔ	frog
139.	kebbhâng	kəpʰ:ɤŋ	wide
140.	kebbut	kəb:ut	rush
141.	kèding	kɛdiŋ	hear
142.	kèjhing	kɛcʰiŋ	mussel
143.	kèkèl	kɛkɛl	beef feet
144.	kèkèr	kɛkɛr	saw sharpener
145.	kèpa'	kɛpaʔ	empty
146.	keppay	kəp:aj	fan
147.	kobhung	kɔpʰuŋ	traditional cottage
148.	kocèng	kɔcɛŋ	cat
149.	kodhuh	kɔtʰuh	must

150.	kopèh	kɔpɛh	bottle
151.	kopèng	kɔpɛŋ	ear
152.	pacal	pacal	hoe
153.	pacèh	pacɛh	noni
154.	padâh	padɤh	similar
155.	padih	padih	paddy
156.	paghâr	pakʰɤr	fence
157.	pajung	paɟuŋ	umbrella
158.	pakèl	pakɛl	young mango
159.	pakoh	pakɔh	nail
160.	patot	patɔt	appropriate
161.	pedḍhâk	pəʈʰ:ɤʔ	run after
162.	pegghâ'	pəkʰ:ɤk	disconnected
163.	pèghâ'	pɛkʰɤʔ	catch
164.	pejji	pəɟ:i	baby pigeon
165.	pèlèt	pɛlɛt	massage
166.	pètè'	pɛtɛʔ	chick
167.	pocèt	pɔcɛt	young fruit
168.	pojhur	pɔcʰur	lucky
169.	poka'	pɔkaʔ	traditional drink
170.	potong	pɔtɔŋ	cut
171.	tabâng	tabɤŋ	run after
172.	taghi	takʰi	claim
173.	tajhin	tacʰin	porridge
174.	tapah	tapah	asceticism
175.	tatah	tatah	arrange
176.	tebbhâk	təpʰ:ɤʔ	guess
177.	tebbus	teb:us	leak
178.	tegghu	təkʰ:u	strong
179.	tèkat	tɛkat	determined
180.	tèpès	tɛpɛs	thin
181.	tètèh	tɛtɛh	small bridge
182.	tètèr	tɛtɛr	spread
183.	tobi'	tɔbiʔ	pinch
184.	todi'	tɔdiʔ	knife
185.	todus	todus	shy
186.	tojhu'	tɔcʰu	be aimed at
187.	tokang	tɔkaŋ	handyman
188.	totop	tɔtɔp	close

9.2 Demographic and Language Questionnaire[18]

1. Age

What is your age?

---

[18] The distributed questionnaire was written in Indonesian.

2. Origin

Where are you originally from?

Have you been living in Madura most of your time? Or have you ever lived outside Madura for a certain period of time?

3. Sex

What is your sex?

a. Male
b. Female

4. Race/ethnicity

How do you describe yourself? (please check the one option that best describes you)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

5. How do you describe your father? (please check the one option that best describes you)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

6. How do you describe your mother? (please check the one option that best describes you)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

7. How do you describe your paternal grandfather? (please check the one option that best describes you)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

8. How do you describe your paternal grandmother? (please check the one option that best describes you)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

9. How do you describe your maternal grandfather? (please check the one option that best describes you)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

10. How do you describe your maternal grandmother? (please check the one option that best describes you)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

11. Where are they (your father, mother, grandfather, grandmother, spouse) originally from?

12. Marital status

Are you:

a. Married
b. Divorced
c. Widowed
d. Separated
e. Single

If you are married, how do you describe your spouse? (please check the one option that best describes your spouse)

a. Madurese
b. Javanese
c. Sundanese
d. Other (specify)

13. Employment status

Are you currently:

a. Employed for wages
b. Self-employed
c. A homemaker
d. A student
e. Retired

14. Education completed

What is the highest grade or year of school you completed?

a. Never attended school or only attended kindergarten
b. Grades 1 through 6 (Elementary)
c. Grades 7 through 9 (Junior high school)

d. Grade 12 (Senior high school graduate)
e. College 1 year to 3 years (Some college of technical school)
f. College 4 years (University graduate)
g. Graduate School (Advance Degree)

15. Languages

What language do you speak most of the time?

a. Madurese
b. Javanese
c. Sundanese
d. Indonesian

16. What language do your father and mother speak most of the time?

a. Madurese
b. Javanese
c. Sundanese
d. Indonesian

Do you speak English?

9.3 Speakers' Information

No.	Name (Initial)	Details[19]
1.	AF	a. Gender: Female b. Age: 19 years c. Job: Student d. Place and DOB: Pamekasan, 30 January 1994 e. Father's place of Birth: Pamekasan f. Mother's place of birth: Pamekasan g. Father: Madurese h. Mother: Madurese i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
2.	FZ	a. Gender: Male b. Age: 20 years c. Job: Student d. Place and DOB: Sumenep, 6 March 1993 e. Father's place of Birth: Sumenep f. Mother's place of birth: Sumenep g. Father: Madurese h. Mother: Madurese

---

[19] All speakers were asked verbally for their consents to participate in the study. Only if they approved to do so, the recording were then conducted. Each of the participant's age shown here was her/his age at the time of the recording.

		i. Mother’s mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
3.	HH	a. Gender: Female b. Age: 19 years c. Job: Student d. Place and DOB: Sumenep, 12 Dec 1994 e. Father’s place of Birth: Sumenep f. Mother’s place of birth: Sumenep g. Father: Madurese h. Mother: Madurese i. Mother’s mother tongue: Madurese j. Frequency of speaking Madurese: Always k. Other languages spoken: Indonesian l. Ever live outside of Madura: Yes m. Length of stay: One month
4.	DD	a. Gender: Male b. Age: 23 years c. Job: Student d. Place and DOB: Pamekasan, 1990 e. Father’s place of Birth: Sampang f. Mother’s place of birth: Sampang g. Father: Madurese h. Mother: Madurese i. Mother’s mother tongue: Madurese j. Frequency of speaking Madurese: Always k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
5.	HF	a. Gender: Male b. Age: 20 years c. Job: Student d. Place and DOB: Sumenep, 15 March 1993 e. Father’s place of Birth: Sumenep f. Mother’s place of birth: Sumenep g. Father: Madurese h. Mother: Madurese i. Mother’s mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
6.	KA	a. Gender: Male b. Age: 28 years c. Job: Student d. Place and DOB: Pamekasan, 4 May 1985 e. Father’s place of Birth: Jember f. Mother’s place of birth: Pamekasan g. Father: Not Madurese h. Mother: Madurese

		i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
7.	LH	a. Gender: Female b. Age: 20 years c. Job: Student d. Place and DOB: Sampang, 1 November 1993 e. Father's place of Birth: Sampang f. Mother's place of birth: Sampang g. Father: Madurese h. Mother: Madurese i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Always k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
8.	LJ	a. Gender: Male b. Age: 20 years c. Job: Student d. Place and DOB: Sumenep, 14 October 1993 e. Father's place of Birth: Jember f. Mother's place of birth: Sumenep g. Father: Not Madurese h. Mother: Madurese i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian l. Ever live outside of Madura: Yes m. Length of stay: Seven years
9.	MH	a. Gender: Male b. Age: 20 years c. Job: Student d. Place and DOB: Sumenep, 15 May 1993 e. Father's place of Birth: Sumenep f. Mother's place of birth: Sumenep g. Father: Madurese h. Mother: Madurese i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Always k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
10.	MS	a. Gender: Female b. Age: 21 years c. Job: Student d. Place and DOB: Bangkalan, 14 June 1992 e. Father's place of Birth: Bangkalan f. Mother's place of birth: Pamekasan g. Father: Madurese h. Mother: Madurese

		i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Always k. Other languages spoken: Indonesian l. Ever live outside of Madura: Yes m. Length of stay: Two months
11.	NA	a. Gender: Female b. Age: 21 years c. Job: Student d. Place and DOB: Sampang, 14 June 1992 e. Father's place of Birth: Sampang f. Mother's place of birth: Sampang g. Father: Madurese h. Mother: Madurese i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Always k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
12.	NK	a. Gender: Male b. Age:21 years c. Job: Student d. Place and DOB: Sampang, 17 September 1991 e. Father's place of Birth: Sampang f. Mother's place of birth: Sampang g. Father: Madurese h. Mother: Madurese i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian, Javanese, English l. Ever live outside of Madura: No m. Length of stay: NA
13.	OR	a. Gender: Female b. Age: 22 years c. Job: Student d. Place and DOB: Bangkalan, 14 August 1991 e. Father's place of Birth: Bangkalan f. Mother's place of birth: Bangkalan g. Father: Madurese h. Mother: Madurese i. Mother's mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian, English l. Ever live outside of Madura: Yes m. Length of stay: One year
14.	UH	a. Gender: Female b. Age: 19 years c. Job: Student d. Place and DOB: Sampang, 13 February 1994 e. Father's place of Birth: Sampang f. Mother's place of birth: Sampang g. Father: Madurese

		h. Mother: Madurese i. Mother’s mother tongue: Madurese j. Frequency of speaking Madurese: Often k. Other languages spoken: Indonesian l. Ever live outside of Madura: No m. Length of stay: NA
15.	WM	a. Gender: Female b. Age: 19 years c. Job: Student d. Place and DOB: Bangkalan, 27 July 1994 e. Father’s place of Birth: Bangkalan f. Mother’s place of birth: Bangkalan g. Father: Madurese h. Mother: Madurese i. Mother’s mother tongue: Madurese j. Frequency of speaking Madurese: Always k. Other languages spoken: Indonesian, English l. Ever live outside of Madura: No m. Length of stay: NA