Working Papers of the Cornell Phonetics Laboratory
2002, vol. 14, pp. 113-206

Guttural Vowels and Guttural Coarticulation in Ju|'hoansi*

Amanda Miller-Ockhuizen

Ju|'hoansi contains parallel glottalization, breathiness and epiglottalization on consonants and vowels, as well as uvularization on consonants. There are voice quality cues associated with epiglottalized vowels that parallel those found with breathy and glottalized vowels. There is also guttural coarticulation, whereby aspirated, glottalized, uvularized and epiglottalized consonants spread acoustic voice quality cues onto a following vowel. The laryngeal coarticulation exhibited by click consonants is shown to be largely analogous to coarticulation found with pulmonic consonants. Two levels of acoustic similarity involving gutturals are identified. Low HNR in the C-V transition unites all gutturals, and low spectral slope differentiates glottals and epiglottals from aspirates and uvularized sounds that are associated with high spectral slopes. The two natural classes identified by these cues are shown to be targeted in perceptually-motivated sound patterns. In addition to the four-way phonation type contrast on vowels, there are also diphthongs in breathiness and epiglottalization, leading to a three-way timing contrast in roots. Sufficient temporal separation of voice quality cues is not achieved between roots containing similar guttural consonants and guttural vowels. Phonotactic patterns assure that similar consonants and vowels do not co-occur within a single root, allowing Voice Onset Time to assure sufficient discriminibility of roots containing guttural consonants and roots containing guttural vowels. This suggests that sound inventories can not be evaluated for discriminability in isolation of existing co-occurrence patterns.

1. Introduction

Ju|'hoansi, typical of Khoesan languages, contains parallel non-modal phonation type contrasts on both consonants and vowels. Consonants with non-modal phonation types occurring in the C-V transition and parallel vocalic phonation types all identify a natural class targeted by an Obligatory Contour Principle (OCP) constraint (Leben 1973; Goldsmith 1976). Since all of the sounds in the natural class targeted by the OCP are articulated in the laryngeal or pharyngeal region of the vocal tract, the relevant natural class is [guttural], which is well-supported by its presence in Semitic and Salishan language phonologies. The primary hypothesis tested in the current study is that guttural consonants and vowels in Ju|'hoansi all display marked acoustic voice quality cues such

* I would like to thank Mary Beckman for her help in designing, implementing and analyzing the results of the experiments reported on here, which were originally included in my dissertation under her direction (Miller-Ockhuizen 2001). I would also like to thank Abby Cohn, Osamu Fujimura, Bonny Sands, Marisol del Teso-Craviotto, and members of my seminar on the acoustics of voice quality taught at Cornell in the spring of 2002 (Marc Brunelle, Rina Kreitman and Anastasia Riehl) for comments on a draft version of this paper.

as increased or decreased spectral slopes and a decreased harmonics-to-noise ratio (HNR) relative to modally voiced vowels and voiceless and voiced unaspirated consonants.

Ju|'hoansi is an excellent language for voice quality investigation of gutturals, given that it contains a full set of parallel phonation type contrasts on both consonants and vowels not attested in other languages of the world. This allows me to explicitly compare the timing and magnitude of cues associated with guttural vowels to those occurring on the vowel following parallel guttural consonants in the same language. Additionally, Ju|'hoansi, typical of Khoesan languages, has a large click consonant inventory, allowing me to compare laryngeal coarticulation involving click consonants that use the velaric airstream mechanism to laryngeal coarticulation involving pulmonic consonants. In addition to providing novel evidence regarding the acoustic cues associated with laryngeals and pharyngeals, this study provides evidence for the acoustic definition of the phonologically motivated natural class of guttural consonants and vowels.

The main hypothesis contains four sub-hypotheses. The first hypothesis investigated is that voice quality cues associated with laryngeally articulated vowels in other languages extend to Ju|'hoansi breathy and glottalized vowels. A related sub-hypothesis is that Ju|'hoansi epiglottalized vowels also bear similar voice quality cues. If all of these vowels contain voice quality cues, this will be a unifying property of all guttural vowels that pattern together phonologically in the language with respect to a guttural OCP constraint. Additionally, it will suggest that guttural vowels might all be marked by a laryngeal feature in the phonology. These hypotheses are tested in experiment 1, the results of which are reported in section 4. Results show that indeed Ju|'hoansi breathy, glottalized and epiglottalized vowels all display marked voice quality cues, providing evidence that the phonetic basis of the guttural natural class in Ju|'hoansi vowels can be defined in terms of voice quality cues.

The second hypothesis tested in this paper is that laryngeally articulated consonants spread marked voice quality cues to a following vowel through a process of laryngeal coarticulation. A related sub-hypothesis is that pharyngeally articulated consonants in the language, namely uvularized consonants and epiglottalized consonants, also spread voice quality cues to the following vowel. This hypothesis is tested in experiment 2, and reported in section 5. Results show that aspirated, glottalized, uvularized and epiglottalized consonants all involve marked voice quality cues on the following vowel, in a processed termed guttural coarticulation.

A third hypothesis, also tested in experiment 2, is that click consonants display similar laryngeal coarticulation to that found with aspirated and glottalized pulmonic plosives cross-linguistically (Löfqvist and McGowan 1992; Gobl and NíChasaide 1999; Cho, Jun and Ladefoged 2000). Results show that laryngeal release mechanisms for click and non-click consonants are highly similar, despite the differences in the primary airstream mechanism. Results also show VOT differences among click types that can be attributed to place of articulation differences just prior to release, either different locations of the posterior constriction in clicks as have been shown by Thomas-Vilakati (1999) to occur in Zulu clicks, or to slight differences in the amount of overlap between the releases of the anterior and posterior constrictions as has been suggested by Miller-Ockhuizen (2000) and Traill (1997: 115).

A fourth goal of the study is to investigate the temporal location of voice quality cues associated with the partially breathy and partially epiglottalized vowels in Ju|'hoansi, and to explicitly compare the timing of the cues associated with these vowels to those associated with fully breathy and epiglottalized vowels, and those associated with aspirated and epiglottalized consonants. In experiment 3, I test the hypothesis that the temporal location of voice quality cues associated with partially breathy and partially epiglottalized vowels is similar to the temporal location of cues associated with aspirated and epiglottalized consonants. Results, reported in section 6, show that the timing of voice quality cues is similar between roots containing initial aspirated consonants and roots containing partially breathy vowels, but that the roots containing consonants and vowels differ in terms of Voice Onset Time (VOT). The timing of the voice quality cues is then also a factor in determining the acoustic similarity of gutturals. Voiced clicks exhibit only slightly higher voice quality cues in the C-V transition than are associated with unaspirated clicks, which leads to a lack of similarity with other gutturals, and their classification of them as non-gutturals in the phonology.

The phonologically motivated natural class of [guttural] consonants found in Semitic (Hayward and Hayward 1989; McCarthy 1994), Afro-Asiatic (van der Hulst and Mous 1992; Rose 1996) and Salishan languages (Bessell 1992; Shahin 1997) has been difficult to define articulatorily. As noted by McCarthy (1994), there is no single articulator involved, and Goldstein (1994) and Lee (1995) have suggested that the class might be defined in terms of jaw movement. Zawaydeh (forthcoming) shows through fiberscopic experiments that laryngeals in Jordanian Arabic do not involve tongue root retraction, making an articulatory definition of [guttural] in terms of tongue root position intractable for that language. In Miller-Ockhuizen (In progress), I show that Ju|'hoansi epiglottalized consonants and vowels do not involve glottal constriction, but rather a more complex constriction involving the aryepiglottic sphincter as has been shown to exist in a related Khoesan language !Xóõ (Traill 1986) and in Arabic (El-Halees 1986; Esling 1996). The lack of glottal constriction in these sounds makes an analysis in terms of traditional laryngeal features (Halle and Stevens 1971; Ladefoged 1973) untenable.

On the other hand, acoustic similarity of gutturals is much more tractable. The results of the quantitative investigation of voice quality cues associated with laryngeal and pharyngeal consonants and vowels in Ju|'hoansi show that they all involve low HNR in the C-V transition. The voice quality cues associated with laryngeal and pharyngeal consonants and vowels in Ju|'hoansi are equally strong and consistently present in all vowel contexts. This can be contrasted to Zawaydeh's (forthcoming) more context-specific results, which show that guttural consonants in Jordanian Arabic all involve first formant frequency (F1) raising on the following vowel (consistent with the lower jaw position of Palestinian Arabic gutturals demonstrated by Lee 1995). Zawaydeh's results show that F1 is raised much more greatly in the [a] context than in the [i] context, and that the amount of raising is greater following pharyngeals than following laryngeals.

I hypothesize that F1 raising following pharyngeals is the phonetic basis of the behavior of uvular, uvularized and epiglottalized consonants and vowels in a separate process of vowel lowering and retraction known as the Back Vowel Constraint (Traill 1985; Miller-Ockhuizen 2000, 2001). This separate process of vowel lowering in Ju|'hoansi and other Khoesan languages is triggered by uvulars and epiglottals, but not laryngeals. F1 patterns in a related Khoesan language Khoekhoe, which maintains a contrast between front vowels and cross-place diphthongs following all click types, including glottalized, uvularized and epiglottalized clicks, show that there is a consistent phonetic lowering of front vowels following pharyngeals even in languages where a contrast between front vowels and cross-place diphthongs following them is maintained (Miller-Ockhuizen Submitted). On the other hand, F1 patterns following laryngeals are more transient and context-sensitive.

Given the marked strong acoustic similarity found in Ju|'hoansi guttural consonants and vowels in substance and timing, it is difficult to see how paradigmatic contrasts between consonants and vowels can be maintained. I suggest that the reason parallel contrasts in Ju|'hoansi consonantal and vocalic inventories are allowed to exist in the language is because there is a guttural OCP constraint active in the language that rules out the possibility of any two consonantal and vocalic phonation type contrasts occurring within the same root (Miller-Ockhuizen 2001). That is, the existence of a guttural OCP constraint targeting both consonants and vowels assures that voice onset time (VOT), which only applies to consonants, is always available to differentiate roots containing consonantal phonation type contrasts from those containing vocalic phonation type contrasts. In this way, consonantal cues assure the maintenance of the paradigmatic contrast between guttural consonants occurring in roots such as *!ˤàà ‘to dry out’* and vowels such as *‖ȁˤà ‘to hold’* and *!aaˤ ‘iron’*. Thus, while the temporal location of acoustic cues associated with parallel contrasts is extremely important for determining perceptibility of linguistic contrasts (Silverman 1997), perceptibility of a particular set of contrasts cannot be evaluated in isolation of the phonotactic constraints operative in the phonological system. The phonotactic constraints operative in a language also serve to assure that paradigmatic contrasts are recoverable by assuring availability of cues for a given contrast. That is, our understanding of what counts as sufficient discriminability of phonological contrasts in a language (Liljencrants and Lindblom 1972; Lindblom and Maddieson 1988; Lindblom 1990) can be improved greatly if we limit the assessment of sufficient discriminability to sounds that are allowed by the phonological system to co-occur within the same prosodic domain (or in some cases morphological domain, as shown by McDonough forthcoming). Perceptual experiments are planned to identify the cue weighting of voice quality cues and voice onset time in the perception of Ju|'hoansi guttural consonants and vowels in order to provide empirical evidence about the use of VOT in differentiating guttural consonants and vowels.

In section 2 of this paper, I provide a brief phonological sketch of Ju|'hoansi, outlining the guttural consonant and vowel inventories as well as the relevant phonotactic constraints affecting gutturals. Then in section 3, I describe and motivate the basic methodology used in this study, before turning to the three experiments focusing on the acoustics of guttural vowels and guttural consonants reported in sections 4 and 5 respectively. In section 6, I report the results of experiment 3, which explicitly compares the timing of voice quality cues associated with partially breathy and partially epiglottalized vowels to cues associated with aspirated and epiglottalized consonants. In section 7, I describe the synchronic and diachronic sound patterns that have their bases in the perceptual similarities revealed among guttural consonants and vowels in this study, and in section 8, I conclude with a broad discussion of the implications of this study for phonetic and phonological theories.

2. Ju|'hoansi guttural phonology

2.1 Guttural vowel inventory

The full inventory of Ju|'hoansi guttural vowels is provided in Table I. There are breathy, glottalized and epiglottalized vowels. These vowels contrast with modally voiced vowels that are considered to be unmarked. As can be seen in the inventory, there are no epiglottalized high or front vowels. This is discussed in section 2.4. Interestingly, all guttural vowel types also contrast in being oral and nasal as well. Ohala and Ohala (1993) and Fujimura and Lindqvist (1971) have discussed the acoustic similarity of nasality and breathiness, and Ohala and Ohala (1993) have discussed the relevance of the acoustic similarity to phonotactic constraints. Therefore, it is of considerable interest that the acoustic similarity in terms of voice quality cues which classifies guttural vowels together does not encompass nasality. That is, F1 bandwidth, which is assumed to be the principle acoustic correlate of nasality, can not be the acoustic cue that defines the natural class of gutturals since this has also been linked to breathiness (Huffman 1987).

2.2 Guttural consonant inventory

The full inventory of Ju|'hoansi guttural consonants is provided in Table II. Aspirated, glottalized and epiglottalized click consonants are parallel to the same phonation types found in the vowel inventory (with breathy vowels being articulatorily and acoustically equivalent to aspirated consonants, only differing in timing relative to the consonant release). However, there are also uvularized consonants that have no parallel in the vowel inventory. The epiglottalized clicks are those that have been transcribed as [!x’] by Snyman, that are rendered in Ju|’hoansi orthography as ‘!k’ by Dickens (1994), and are transcribed as [k!^{x}’] by Ladefoged and Maddieson (1996: 275-77), who refer to them as clicks followed by voiceless velar affricates, assuming a cluster analysis. Miller-Ockhuizen (2000) shows that the place of articulation of the fricated portion of these clicks have a constriction in the pharynx given the higher F1 found following these consonants when compared with plain unaspirated click consonants of the same type. These guttural consonant types contrast with unmarked voiceless unaspirated consonants that are unmarked, and voiced unaspirated consonants that have marked modal voice specifications.

Additionally, unlike the vowel inventory where all guttural vowel types also contrast in being oral or nasal, only aspiration occurs with nasalization in the consonant inventory. The lack of nasalization occurring with uvularized and an epiglottalized consonant is found across all known Khoesan languages. However, !Xóõ has two click consonants involving nasalization and glottalization (Ladefoged and Traill 1984), and Ju|'hoansi glottalized clicks themselves usually occur with nasalized vowels (Snyman 1975). There have been nasal airflow studies showing that glottalized clicks involve nasal airflow in Khoekhoe (the Nama variety) (Ladefoged and Traill 1984) Hadza (Sands et al. 1996) and Sandawe (Wright et al. 1995; Maddieson et al. 1999), but there have not been any nasal airflow studies of this particular click type in Ju|’hoansi. I have been unable to

	FRONT VOWELS						BACK VOWELS					
	BREATHY		GLOTTALIZED		EPIGLOTTALIZED		BREATHY		GLOTTALIZED		EPIGLOTTALIZED	
	ORAL	NASAL	ORAL	NASAL	ORAL	NASAL	ORAL	NASAL	ORAL	NASAL	ORAL	NASAL
High	$i^{ɦ}i^{ɦ}$	$i^{ɦn}i^{ɦn}$	$i^{ʔ}i$	$i^{ʔn}i^{n}$			$u^{ɦ}u^{ɦ}$	$u^{ɦn}u^{ɦn}$	$u^{ʔ}u$	$u^{ʔn}u^{n}$		
Mid	$e^{ɦ}e^{ɦ}$	$e^{ɦn}e^{ɦn}$	$e^{ʔ}e$	$e^{ʔn}e^{n}$			$o^{ɦ}o^{ɦ}$	$o^{ɦn}o^{ɦn}$	$o^{ʔ}o$	$o^{ʔn}o^{n}$	$ɔ^{ʕ}ɔ^{ʕ}$	$ɔ^{ʕn}$ $ɔ^{ʕn}$
Low							$a^{ɦ}a^{ɦ}$	$a^{ɦn}$ $a^{ɦn}$	a ^{ʔ}a	a $^{ʔn}a^{n}$	$ɑ^{ʕ}ɑ^{ʕ}$	$ɑ^{ʕn}$ $ɑ^{ʕn}$

Table I: Guttural Vowel Inventory

identify any Ju|’hoansi roots that contain oral vowels following glottalized clicks. Therefore, despite the exclusion of nasal vowels more generally in this study, the glottalized click words in this study all containing nasalized vowels. Further research is planned to determine the exact relationship between glottalization and nasalization in Ju|’hoansi and across Khoesan languages. It may be that the glottalized clicks should be placed in the nasal category, but in absence of empirical data as to their behavior as nasals, as well as to the presence of nasal airflow in Ju|’hoansi, I leave them in the oral column.

As can be seen in Table II, the full set of phonation type contrasts largely occur in parallel in both click and non-click consonant inventories. The pulmonic stop consonants contrast in labial, alveolar and velar places of articulation, and the velaric stop consonants contrast in having dental [|], palato-alveolar [ǂ], and post-alveolar [!], [‖] anterior places of articulation. The two post-alveolar clicks contrast in being central [!] vs. lateral [‖]. Miller-Ockhuizen (2001: 162) shows that there is a strong co-occurrence constraint blocking the occurrence of guttural release types on labial non-click obstruents. There are no instances of roots containing uvularized, epiglottalized or glottalized labial obstruents in Miller-Ockhuizen's Ju|'hoansi database, which is itself expanded from Dickens' (1994) dictionary. The labial aspirated plosives [pʰ] and [bʱ] are included in the table given that there are a few onomatopoeic words containing these consonants such as *pʰèèpʰèè 'glutton'* and *bʱèè 'to spit out'*, found in the database. However, aspirated labials are also extremely marginal. Traill (1985) suggests that labial consonants in !Xóõ only occur in loan-words, but such an analysis ignores the fact that labial consonants in Khoesan languages more generally, including !Xóõ, account for a large percentage of medial consonants. In Ju|’hoansi, about half of the medial consonants are labials. In Miller-Ockhuizen (2001: 131), I suggest that it is the lack of perceptual salience of labials generally that makes them less well-suited for co-occurrence with guttural release properties only found in initial position, as the noisy releases would tend to obscure the less salient burst cues of labials. Thus, the smaller inventory of labial consonants leads to

	NON-CLICK			**CLICK**			
	Oral		**Nasal**	**Oral**		**Nasal**	
Release	**Voiceless**	**Voiced**	**Voiceless / Voiced**	**Voiceless**	**Voiced**	**Voiceless**	**Voiced**
Aspirated	p^{h} t^{h} k^{h} ts^{h} tʃ^{h}	b^{ɦ} d^{ɦ} g^{ɦ} ds^{ɦ} dʃ^{ɦ}		ǀ^{h} ǂ^{h} !^{h} ǁ^{h}	gǀ^{ɦ} gǂ^{ɦ} g!^{ɦ}	ŋǀ^{h} ŋǂ^{h} ŋ!^{h}ŋǁ^{h}	ŋǀ^{ɦ} ŋǂ^{ɦ} ŋ!^{ɦ} ŋǁ^{ɦ}
Glottalized	ts’ tʃ’	ds’ dʃ’ dz’ dʒ’		ǀ^{ʔ} ǂ^{ʔ} !^{ʔ} ǁ^{ʔ}			
Uvularized	t^{χ} tʃ^{χ}	d^{χ} dz^{χ} dʒ^{χ}		ǀ^{χ} ǂ^{χ} !^{χ} ǁ^{χ}	gǀ^{χ} gǂ^{χ} g!^{χ} gǁ^{χ}		
Epiglottalized	t^{ʕ} k^{ʕ}			ǀ^{ʕ} ǂ^{ʕ} !^{ʕ} ǁ^{ʕ}	gǀ^{ʕ} gǂ^{ʕ} g!^{ʕ} gǁ^{ʕ}		

Table II. Ju|’hoansi Guttural Consonant Inventory

the low frequency of initial labials. Since gutturals do not occur medially, the frequency of occurrence of medial labials is similar to the frequency of medial coronals.

Non-parallelism between click and non-click consonant inventories is also seen when one contrasts the nasal vs. oral series, and their co-occurrence with guttural phonation types. There are both plain voiced nasal non-click and click consonants. On the other hand, while there are sets of both voiceless and voiced nasal aspirated clicks, there are no aspirated nasal non-click consonants occurring in roots.[1] The expansion of nasality in click inventories has lead Nakagawa (1998) to suggest that [nasal] may be the default velic position for click consonants, not [oral] as is widely assumed to be the default for non-click consonants.

2.3 Three-way timing contrasts in syllables

In addition to the four-way phonation type contrast present in the vocalic inventory, there is also a two-way timing contrast found among breathy and epiglottalized vowels. This leads to a three-way contrast in the timing of the opening of the glottis for breathy sounds in Ju|'hoansi monosyllabic roots, and a three-way contrast in the timing of the laryngeal sphincteric action present in epiglottalization (Traill 1986; Esling 1999; Miller-Ockhuizen, Namaseb, del Teso Craviotto and Iskarous, In progress) in monosyllabic roots. The Southeast Asian languages Suai and Wa are the only other languages that I am aware of that have both aspirated consonants and breathy vowels (Watkins 1999; Abramson and Luangthongkum 2001). Given the small number of languages that use the aryepiglottic sphincter mechanism, it is not surprising that there are no other known languages with timing contrasts in epiglottalization. The assumed prosodic organization of the three-way timing contrasts in monosyllabic roots is provided in Figure 1. Vowels with breathiness or epiglottalization associated to the first mora are also referred to as partially breathy vowels and partially epiglottalized vowels respectively throughout this study. In a similar fashion, vowels having breathiness or epiglottalization associated to both moras are referred to as fully breathy and fully epiglottalized vowels respectively throughout this study.

[1] The only morpheme containing a voiced nasal aspirate is the diminutive plural morpheme *m^ɦi* .

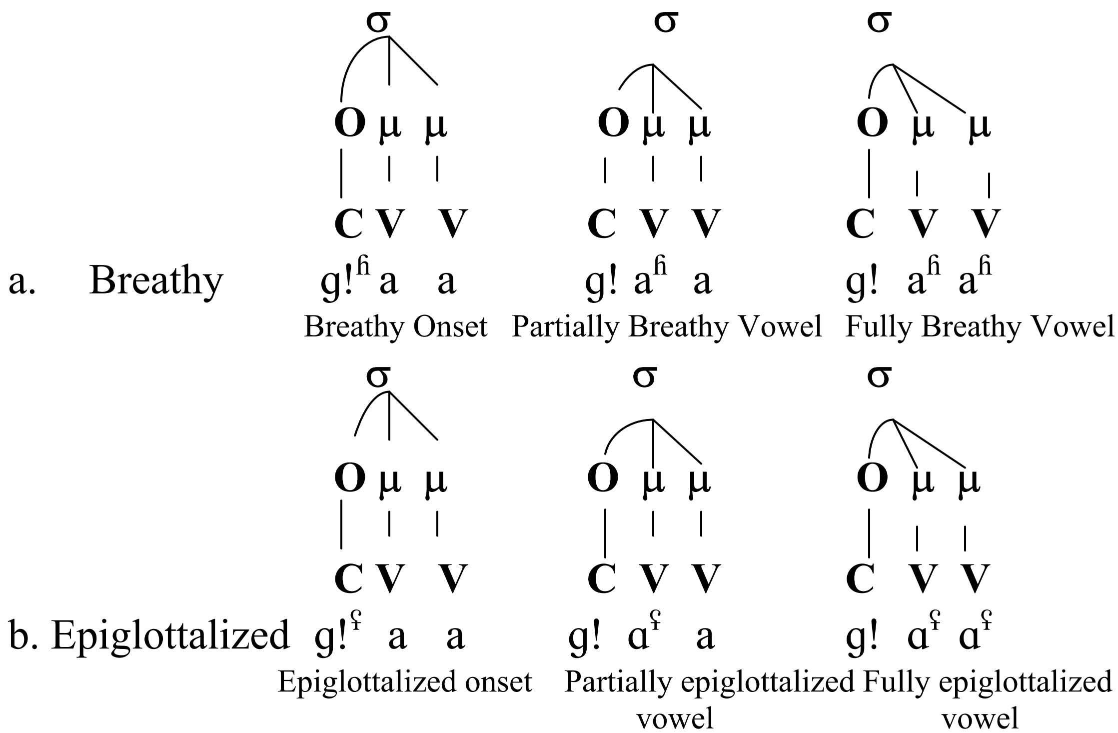


Figure 1. Three-way timing contrast in breathiness and epiglottalization in roots

As can be seen through comparison of Tables I and II, Ju|'hoansi consonant and vowel inventories have parallel contrasts in aspiration / breathiness, glottalization, and epiglottalization. The only difference between the consonant and vowel inventories is that Ju|'hoansi consonants can also be uvularized, while there are no uvularized vowels. A related language !Xóõ does have a contrast between strident vowels, which Traill (1985) deems to be both breathy and pharyngealized, and pharyngealized vowels, which are solely pharyngealized. Thus, it is possible to have completely parallel contrasts in aspiration, uvularization (assuming that strident vowels are parallel to uvularized consonants), glottalization and epiglottalization in consonant and vowel inventories, as shown by the system in !Xóõ.

2.4 Relevant guttural co-occurrence restrictions

As mentioned above, there is a Guttural OCP constraint in Ju|'hoansi that rules out the co-occurrence of any of the guttural consonants in table II with any of the guttural vowels in table I within the same root.[2] It is typical for languages with pharyngealized consonants in their inventories to have active co-occurrence constraints involving pharyngealization and vowel height and backness. Ju|'hoansi is no exception. As can be gleaned from Table I above, Ju|'hoansi epiglottalized vowels are always non-high back vowels. The Back Vowel Constraint (BVC) that accounts for the diphthongization of front vowels following uvular consonants (including the back clicks [!] and [||]), uvularized and epiglottalized consonants in the language described by Miller-Ockhuizen 2000) also accounts for the lack of high front epiglottalized vowels (Miller-Ockhuizen 2001). The BVC is then another active phonotactic constraint in the language that targets both consonants and vowels in a parallel fashion. However, this constraint only targets a subset of the guttural consonants and vowels, namely those that are uvularized or epiglottalized, but not laryngeally articulated consonants or vowels. Thus, the BVC allows us to identify a second natural class containing uvular and uvularized consonants, as well as epiglottalized consonants and vowels, which excludes laryngeals. Thus, we have evidence that laryngeals pattern with pharyngeals with respect to one constraint (the Guttural OCP constraint discussed above), and are excluded from a natural class containing pharyngeals in another (the BVC) in the same language. This is not predicted by the current representations of pharyngeals. Halle (1995) assumes that they are always specified for a laryngeal feature, McCarthy (1991, 1994) assumes that they are always specified for a place feature, and Rose (1996) assumes that the marking of pharyngeals is dependent on the presence of pharyngeals in the inventory.

In addition to the dependencies found between vowel height and a subset of the gutturals, there are also dependencies found between guttural sounds and lexical tone in

---

2 The domain of the constraint can also be stated in terms of the prosodic word, since Miller-Ockhuizen (2001) shows that the root is concomitant with the prosodic word and smaller morphemes that follow the root contain guttural vowels, in morphs such as the masculine marker *g!ɔˤ*, and guttural consonants occur in morphs such as the diminutive plural *mʱi*. Additionally, loan-word adaptation shows that obstruents with guttural release types are always parsed at the beginning of a prosodic word (Miller-Ockhuizen 2001: 103).

Ju|'hoansi (Miller-Ockhuizen 1998, 2001: 150-155). Breathy vowels and epiglottalized vowels are all lexically low toned, which is not surprising given the articulatory incompatibilities between high tone and pharyngeal narrowing described in Elgendy (2001) and Honda (1995). In addition, glottalized vowels, partially breathy vowels and partially epiglottalized vowels are always rising in tone, having either an L-H or SL-L root pattern. Glottalized vowels have a slight glottal constriction a quarter of the way throughout the vowel's duration, and therefore might be more precisely termed “interrupted” as suggested by Snyman (1977).

It is necessary to understand all of these phonotactic constraints in order to assure that sufficient phonetic control is maintained for reliable measures of voice quality. The constraints are also important to our understanding of the natural classes of sounds that are active in Ju|'hoansi phonology. I argue in section 7 below that the two different groupings identified through voice quality cues are both relevant to sound patterns attested across the Khoesan languages. For example, the acoustic similarity of all guttural consonants and vowels measured here via HNR, is the basis of the guttural OCP constraint found in Ju|'hoansi. However, voice quality cues are not the bases of the vowel lowering or retraction caused by Ju|'hoansi uvulars, epiglottalized and uvularized consonants known as the BVC. Instead, this pattern likely has its basis in the more extreme raising of F1 that occurs with uvulars and epiglottals, but not with laryngeals. Voice quality cues also do not account for the tonal co-occurrence restrictions found in Ju|'hoansi. Rather, these are based purely on the interaction between guttural consonants and vowels and fundamental frequency. Research is in progress to investigate F1 (Miller-Ockhuizen Submitted) and F0 patterns (Miller-Ockhuizen To appear) involving gutturals in Khoesan languages.

3. Methods

Acoustic studies of voice quality have lagged behind other types of analysis in part because of the difficulty of effectively separating acoustic properties of source from acoustic properties of filter (e.g. those that are caused by the configuration of the supralaryngeal cavity) in the guttural region. For laryngeal consonants and laryngealized vowels, there is only one type of modulation that varies for each phonation type – namely laryngeal configuration. That is, there is no additional supralaryngeal constriction involved in laryngeal consonants. However, recent work has shown that the articulatory to acoustic mapping involving laryngeally articulated consonants is not straight-forward, given the complex set of muscles and ligaments that comprise the larynx. For example, sounds like the glottal /t/ allophone in English are produced with a constriction of the false vocal folds rather than a constriction at the glottis (Fujimura and Lindqvist 1971). Still, comparison of sounds with modal voice (e.g. those where the vocal folds are close together enough to result in periodic vibration), and those with non-modal voice (e.g. where the vocal folds are either too far apart or too close together to vibrate normally) is relatively straightforward given that the filter configuration (the shape of the vocal tract above the glottis) is relatively stable in both types of sounds.

Pharyngeal and epiglottal sounds, on the other hand, contain both marked laryngeal configurations (e.g. spread glottis in uvularized consonants, and raised larynx and a constriction between the false vocal folds and the pharynx in epiglottalized consonants) as well as different supralaryngeal configurations (pharyngeal constriction) from modally voiced consonants and vowels, leading to difficulty in determining which of the acoustic properties of the signal can be attributed to differences in laryngeal configuration and which can be attributed to the supralaryngeal constriction. Given the fact that a constriction in the pharyngeal region results in substantial raising of the first formant frequency (F1) when compared with constrictions at other places in the vocal tract (Alwan 1989; Zawaydeh forthcoming), and sometimes also changes in the second and third formant frequencies (Obrecht 1968; Card 1983; Bessell 1998), it is difficult to use spectral slope measures to assess voice quality when comparing pharyngealized or epiglottalized vowels with vowels having either modal voice or other marked phonation types. For example, the F1 differences associated with Ju|'hoansi epiglottalized vowels compared with modal vowels are at least 100 Hz, and up to 400 Hz for some speakers (Miller-Ockhuizen 2001: 67), and the F1 differences found between vowels following unaspirated click consonants and uvularized click consonants in Ju|'hoansi are about 75 Hz for dental and palatal clicks, and about 150 Hz for central and lateral post-alveolar clicks (Miller-Ockhuizen 2000: 329). Thus, since it is impossible to maintain similar supralaryngeal configurations for laryngeal and pharyngeal consonants or vowels, in order to investigate spectral properties of guttural consonants and vowels in Ju|'hoansi, it is crucial to identify measures that are not dependent on F1 frequency.

Fundamental frequency (F0) lowering that occurs with epiglottalization and breathiness on both consonants and vowels leads to similar difficulties in controlling data. Lowering leads to different F0 values associated with roots containing different guttural consonants and vowels but that bear the same lexical tones. Thus, despite the phonological control that is obtained by using words containing lexical low tones in the study, it is still impossible to maintain enough phonetic control of fundamental frequency. The lack of phonetic control of F0 makes it thereby impossible to use any measure of voice quality which is dependent on the fundamental frequency. Despite the difficulty of maintaining perfect phonetic control, Ju|’hoansi is an excellent language for comparing laryngeal and pharyngeal consonants and vowels due to the large parallel set of phonation types present in consonant and vocalic inventories.

3.1 Data collection

Recordings were made using a SONY PCM-M1 Digital Audio Tape (DAT) recorder sampled at 44100 Hz, and an AKG C451E condenser microphone in order to assure good low frequency response necessary for spectral investigation and inverse filtering used in the automatic computation of the locations of the pitch epochs (which correspond to the time of vocal fold closure). The wordlists used in experiments 1-3 are provided in Appendices A-C respectively. All wordlists consists of almost entirely click-initial words, since the posterior release of clicks provides a robust acoustic landmark for the beginning of the release phase, and a goal of this study is to investigate the similarity of laryngeal coarticulation found between clicks and non-clicks studied in other languages. It is impossible to explicitly compare guttural coarticulation involving click and non-click consonants in this language, given that words containing non-click consonants have low lexical frequency, which leads to difficulty in finding minimal pairs across all consonant and vowel phonation type contrasts. A difficulty associated with the use of words containing clicks is that recording at a low enough volume to avoid clipping the high intensity part of the click consonants, results in low recorded intensity in the vocalic region. The posterior release of the click consonant is taken to be the beginning of the vowel in clicks with no unmarked release types, since it always occurs somewhat after the anterior release of the click.

In all three experiments, two female subjects (Sp1_f and Sp2_f) and two male subjects (Sp3_m and Sp4_m) were involved. For experiments 1 and 3, fifteen productions of each word in Appendix A and Appendix C respectively, which are subdivided by two vowel contexts, /a/ and /o/, were recorded, resulting in 30 tokens of each vocalic phonation type for each subject. For experiment 2, fifteen repetitions of each word type in Appendix B, containing all consonantal phonation types in three different vowel contexts: /a/, /o/ and /ao/, were recorded, yielding 45 tokens of each consonantal phonation type for each subject. In each experiment, the wordlist was recorded in full before moving on to the next repetition. Since I am fluent in the language, and since my consultants were illiterate, I produced the words myself and discussed the meaning with the subjects prior to each repetition in order to ensure that the subjects were producing the intended words.

3.2 Labeling

Following data collection, the data were digitized and downsampled to 22050 Hz. The data were then labeled using the ESPS XLABEL utility associated with XWAVES. The labels used are provided in Table III below:

VOICE :	beginning of pre-voicing
AR:	anterior release of the click
PR:	posterior release of the click
TR:	transient release of click
EP	end of the periodic portion of the vowel
V:	end of the vowel

Table III: Labels used in study

The transient release ("TR" label) is marked at the end of the initial transient flow as compressed air is released at the end of the posterior release of a click with an unmarked release type. In cases where there is a marked release type, the transient is marked at the end of the longer peak in flow as the vocal-tract walls displace back to rest position and resume position where trans-glottal pressure increases above a threshold to result in voicing, as described in Stevens (1998: 492). Thus, the time interval between the "PR" and "TR" labels includes the silent interval resulting from glottal constriction, aspiration noise resulting from glottal opening found in both aspirated and uvularized clicks, and the combination of frication noise and the subsequent silent interval occurring in epiglottalized consonants. That is, the time between the "PR" and "TR" labels reflects the entire release phase of the click consonant, and corresponds to the VOT (voice onset time). A sample labeled token is provided in Figure 2.

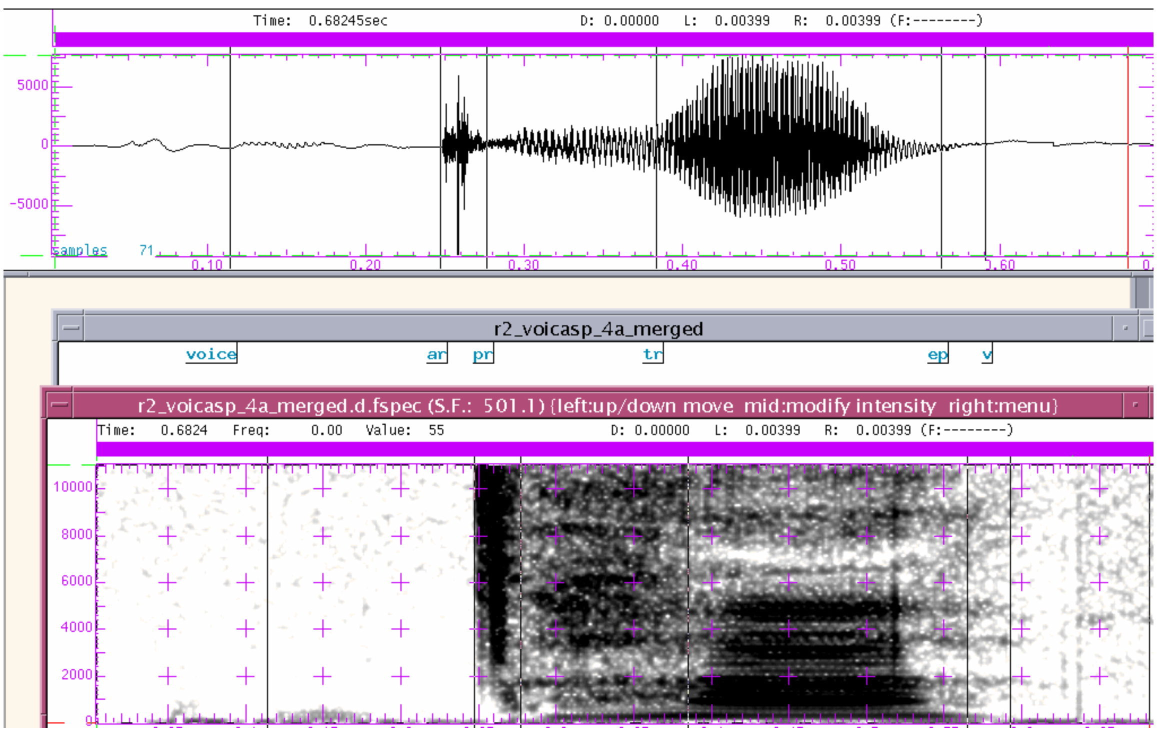


Figure 2: Labeled Token of a single production of the word *g|^{ɦ}àà 'place for drying meat'* produced by female subject Sp1_f

3.3 Pitch Period Determination

Pitch period durations were used to estimate the fundamental frequency in this study, and one of the measures of voice quality used, jitter, looks at the variation in the voice fundamental (pitch) periods within a given window, and thus uses pitch period durations as the primary data. This makes the accuracy of the measure heavily dependent on the accurate placement of the epoch marks at beginnings and endings of each pitch period. In this study, pitch periods were determined as the distance between two adjacent epoch marks. Epochs are the times within each glottal period corresponding to vocal fold closure, and thus correspond to the amplitude maxima which are generated from vocal fold contact as shown by the expanded portion of waveform of a modal vowel produced by subject Sp4_m provided in Figure 3 below. Each vertical line represents an epoch location. The time between each pair of two epoch labels corresponds to the duration of a single pitch period.

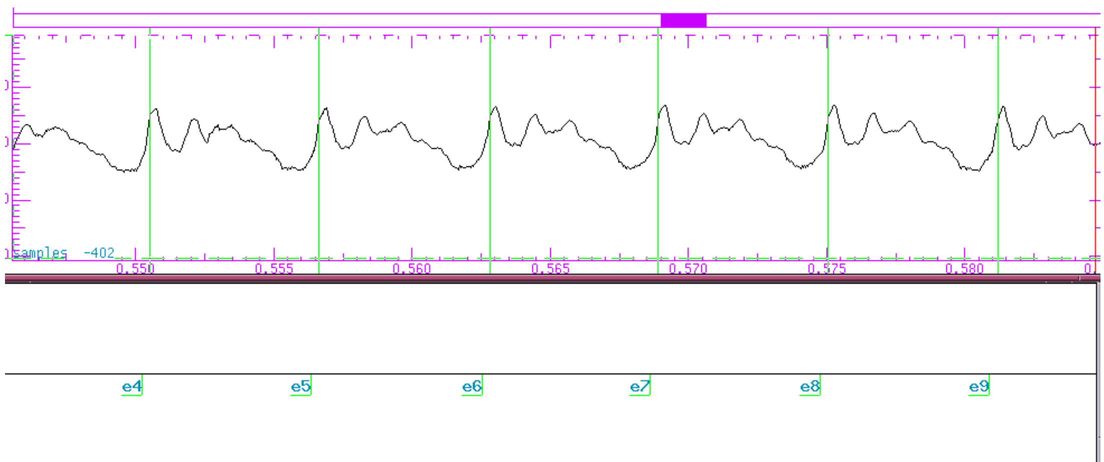


Figure 3: Epoch locations for expanded portion of waveform of a modal vowel in the root ‖*áá ’to warm hands’* produced by subject Sp4_m

Epochs were determined using the ESPS program *Epochs*, which processes the residual signal obtained through inverse filtering (Talkin 1989). Voiceless portions of the signal were gated out following Davis (1976) and as suggested by Talkin (1989). This was done by first computing the reflection coefficients of the vowel over a Hanning window of 20 ms in size, with a step size of 5 ms, and 24 linear predictive coefficients. Probability of voicing, obtained from the results of the ESPS utility *get_f0*, was used to gate out unvoiced portions of the signal. The probability of voicing was computed with a step size of 5 ms between the posterior constriction of the click (or the transient release if one exists) ("PR" or "TR" labels), and the end of the vowel ("V" label). The residual signal, which approximates the second time derivative of glottal flow, was computed using the reflection coefficients, and then the unvoiced portions of the signal (as defined by having a probability of voicing value of 0) were masked, before computing the locations of the pitch epochs. Standard epochs parameters provided in the program were used, except that the allowed jitter level was decreased from 0.1 to 0.05. Systematically varying the parameter, and re-running the program on the epiglottalized vowels that exhibited diplophonia, in order to optimize the identification of epochs, determined the needed decrease allowed in jitter. Given the typically high F0 values found in Ju|'hoansi subjects' speech, and the diplophonic signals found in Ju|'hoansi epiglottalized vowels, the level of allowed jitter needed to be constrained more than it typically is with signal processing in European languages that do not show either of these properties.

After the pitch epochs were computed automatically, I visually inspected each token, and added or deleted epoch markers where necessary. Epoch markers were never added before the automatically placed first epoch label, except in cases where the first epoch label was so late in the signal that the program was obviously missing periodic portions of the signal. In cases where the first automatically detected epoch mark occurred before my subjectively marked end of the transient release ("TR" label), I moved the "TR" mark back to match the location of the first epoch. However, in the majority of tokens, the first epoch was identified computationally only slightly after my more subjective "TR" label (the human eye does not place as stringent requirements on what counts as the same pattern as the epoch program parameters do). The first epoch mark identified after the "TR" label was used rather than the "TR" label itself, as the former provides a more objective measure of the VOT, especially in voiced aspirated clicks, where the end of aperiodicity after the aspiration is difficult to judge. The first epoch within the entire signal was not usable across the board due to the presence of epochs found during the pre-voicing of click consonants, and spurious epochs occasionally found during the releases of epiglottalized clicks.

Given the fact that breathy vowels do not ever reach complete closure, and that the period-to-period shape of the waveform varies greatly in epiglottalized vowels, the automatic computation of epochs within the signal was difficult. Since epoch marks were checked by hand for all of the data, the resultant F0 analysis is more reliable than standard methods of computing F0 for vowels with non-modal voice qualities.

3.4 Spectral slope (H1-H2)

A general measure of spectral slope, defined as the dB difference between the amplitude of the first harmonic (H1) and the amplitude of the second harmonic (H2), (H1-H2), has been claimed by Cho, Jun and Ladefoged (2000) to correspond to the open quotient of the vocal folds. That is, when the vocal folds are open during most of the glottal cycle, the spectrum is dominated by the energy at the fundamental frequency (H1), with the second harmonic (H2) being much lower in amplitude. When the vocal folds are closed most of the time, as they are in glottalized vowels, the spectrum has more energy in the higher harmonics than it does at the fundamental frequency. The more nearly sinusoidal the waveform is, the more predominant the fundamental component is. Higher harmonics are generated mostly by the discontinuous change of waveform slope at the glottal closure. Modally voiced vowels exhibit a pronounced spectral roll-off (generally estimated at 6 dB per octave when the radiation effect is included), but the roll-off is not as sharp as that found with breathy vowels. Breathy vowels in Gujerati and !Xóõ were shown by Bickley (1982) to have steeper downward spectral tilt in the region between 0 and 1000 Hz for breathy vowels than for modal vowels. Jackson et al (1985a, b) claim that H1-H2 is a good, general measure of spectral slope, because H2 generally is a good reflection of the entire spectral slope above H1.

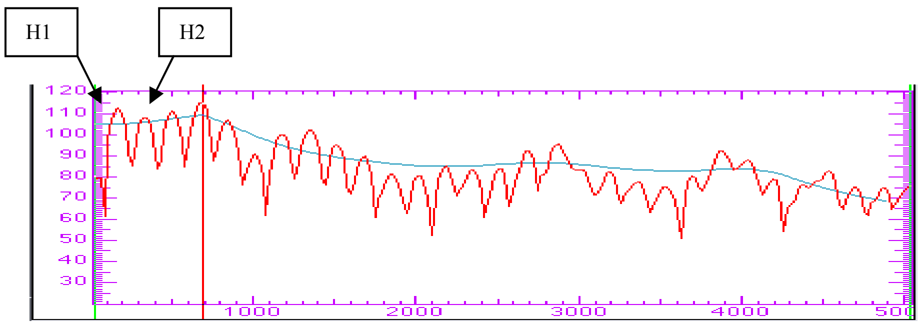
Ladefoged et al. (1988) and Kirk et al. (1993) showed similar results by looking at the difference between the amplitude of the fundamental and that of the harmonic with the highest amplitude in the first formant (H1-A1). Although the H1-A1 measure might at first appear to avoid effects of differences in first formant frequency across different phonation types with different F1 frequencies, this is not the case since the amplitude of F1 is affected by its frequency (NíChasaide and Gobl 1997: 443). Thus, differences in the amplitude of the highest peak within the first formant (A1) would reflect not only phonation type differences, but also differences in F1 frequency.

Based on acoustic results found for other languages, coupled with articulatory postures associated with Ju|’hoansi guttural consonants and vowels, predicted H1-H2 values are provided in Table IV. Breathy vowels in a related language !Xóõ have been shown to have higher H1-H2 values than modal vowels, and thus I hypothesize that Ju|’hoansi breathy vowels, and vowels following aspirated clicks would also have greater H1-H2 values, while vowels following glottalized clicks and glottalized vowels should display lower H1-H2 values. Since uvularized consonants have extremely high airflow, and my own productions of these consonants contain extremely wide-open glottal postures as seen in my high-speed fiberscopic investigation, they are predicted to have high H1-H2 values.

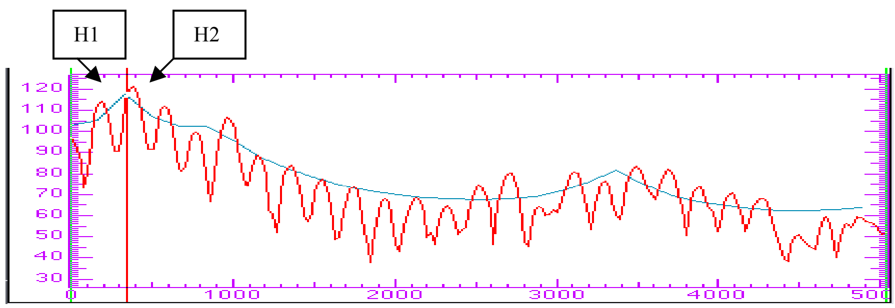
Phonation type	**Predicted H1-H2 values**
Modal	Slightly positive
Breathiness	Highly positive
Glottalization	Negative
Epiglottalization	Negative

Table IV. H1-H2 values predicted to occur with various phonation types

In order to measure spectral slope, it is important to use tokens where H1 and H2 are both well below the range of F1 in order to avoid boosting of either harmonic from proximity to F1. Figure 4(a) shows an LPC spectrum overlaid on an FFT spectrum associated with a lexically low-toned root ||*àà ’to warm hands by fire’*. As can be seen, H1 and H2 are well below F1, marked by the vertical line. In contrast to this, Figure 4(b) shows an LPC spectrum overlaid on an FFT spectrum associated with the lexically high-toned root !*óó ’older brother’* produced by the same speaker. As can be seen, the H2 peak is in close proximity to the peak of F1, and thus may be boosted due to its frequency. In cases like 4(b), H1-H2 would not be a good indicator of glottal posture. Of course boosting of H1 could also occur in words with high vowels where the F1 frequency would be even lower. Roots with high vowels were therefore avoided completely in this study. In order to gain a realistic assessment of glottal posture, H1-H2 was used only with lexically low-toned roots containing low vowels, in order to keep the distance between H1 and H2 peaks and the F1 peak great enough to avoid boosting of either harmonic that is associated with their proximity to F1.



(a) H1 and H2 well below F1



(b) H2 in proximity to F1

Figure 4: Proximity of H1 and H2 to F1

Spectra were computed using Fast Fourier transform (FFT) with 1024 points resulting in a 46.5 ms window, and 10 coefficients appropriate for a sampling rate of 22050 Hz. The windows were computed so that the first window started flush with the posterior release of the click in modally released clicks ("PR" label), or at the end of the transient release in clicks with guttural release properties ("TR" label). A step size of 10 ms was used, until there was less than 46.5 ms left in the vowel to compute a spectrum. All tokens were time normalized by subtracting the starting time of the file from the "PR" or "TR" label. Thus, lags seen between the beginnings of spectral slope values for different consonant types reflect Voice Onset Time (VOT).

The waveform for a lexically low-toned root with an initial unaspirated consonant and a modal vowel is provided in the top portion of Figure 5, with the window for the calculation of a spectrum centered over the midpoint of the vowel marked. A sample spectrum calculated over the window is provided in the bottom left portion of the figure. Estimates of the H1 and H2 frequencies were estimated as the average duration of the pitch periods contained within the same window that was used for the calculation of the spectrum. H1 frequency was estimated as the inverse of the average duration of the periods within the window, and the H2 frequency was estimated as twice that of the estimated H1 frequency. The bottom right portion of Figure 5 shows the first 100 frequency bins in the spectrum, with the windows used around the estimated frequency bins for H1 and H2, and the peak frequencies chosen by the peak-picking algorithm used within the GNU R statistics package marked. The R statistics package was used so that each token could be visually inspected while the peak amplitude values were being identified, to assure accuracy. The algorithm also computed the difference between the H1 and H2 peak amplitudes.

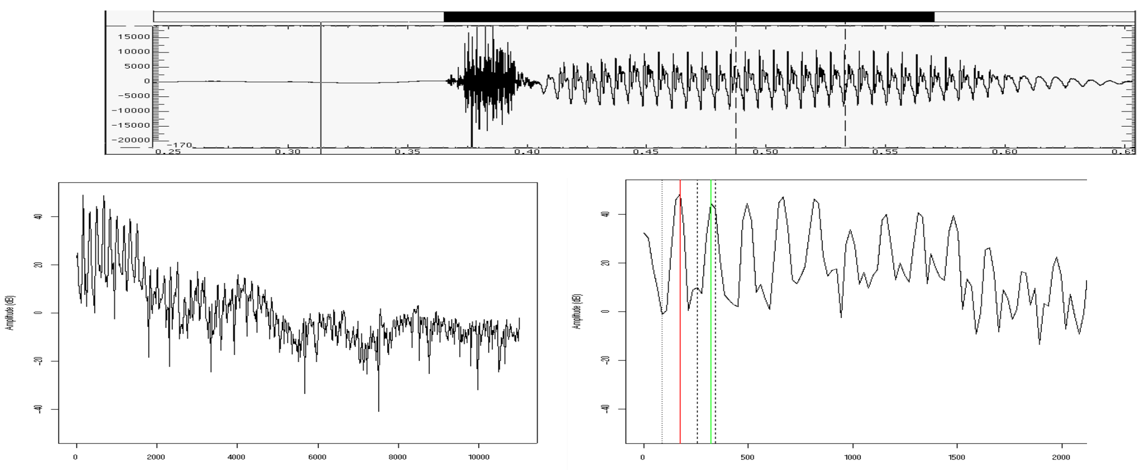


Figure 5: Waveform with the window for the calculation of the spectrum in ǁ*áá 'to warm hands by fire'* (top), spectrum at the midpoint of the vowel (bottom left), and spectrum showing the method of determining frequency bins for H1 and H2 (bottom right) (H1 window - dashed lines, H2 windows-dotted lines, H1 and H2 peaks - solid lines) (Subject Sp4_m)

3.5 Harmonics-to-noise ratio (HNR)

There have been multiple time and frequency based measures of harmonics-to-noise ratio used in the study of pathological voice. Within the time domain, HNR is measured by comparing the shape of successive glottal pulses. In all such studies, the underlying assumption is that the parts of the signal that match from glottal pulse to glottal pulse are harmonic components, and parts that do not match constitute noise in the signal. Studies in this vein have looked for ways to factor out jitter (random variation in the duration of successive glottal periods) from the signal, in order to assess purely variation in the shape of the glottal pulses. Yumoto, Gould and Baer (1982) used zero padding to normalize for differences in period duration. As noted by Qi (1992), this introduces some inaccuracy, since even pulses with similar shapes that display some degree of jitter will display lower HNR values. Qi (1992) uses dynamic time warping to stretch and minimize successive glottal periods, and Qi et al. (1995) added additional zero phase transformations, resulting in an improved model of HNR.

Frequency domain measures of HNR are easier to accomplish since no time normalization is necessary. That is, frequency analysis can be done over any chunk of signal, using windows that do not match with the beginning and end of a glottal cycle. Qi and Hillman (1997) note also that it is easier to estimate the fundamental frequency of a window in the frequency domain by looking at the frequency of the first harmonic. They also note that certain measures of HNR in the frequency domain are less sensitive to low frequency amplitude modulations, which make them more applicable to continuous discourse. These properties also make frequency domain measures of HNR more applicable to contexts where F1 and F0 cannot be completely controlled for, such as allophonic differences associated with phonation type contrasts discussed in section 2.4. This property made frequency based measures of HNR an obvious choice for the current study.

The measure of HNR used in this study is the gamnitude (an analog of the magnitude) of the first rahmonic (an analog of the harmonic) peak (R1) in the cepstrum (an analog of the spectrum). The cepstrum is calculated by taking the Fourier transform of the FFT spectrum, resulting in an analog of the FFT spectrum that is transformed back to the time domain. Since all harmonics within an FFT spectrum are, by definition, multiples of the fundamental (H1), the only prominent rahmonic quefrency (an analog of frequency) corresponds to the fundamental frequency, which in the time domain is the duration of a single glottal pulse.

The first use of the cepstrum was Noll (1964), who used the inverse of the distance between the onset of the cepstrum (power spectrum of the logarithm of the spectrum) and the first rahmonic peak to identify the location of the fundamental frequency. Since quefrency is in the time domain, the duration distance between the onset of the cepstrum and the first rahmonic is identical to the duration of a single glottal pulse, which is the inverse of the fundamental frequency. Noll notes that one of the advantages of cepstral techniques of pitch period identification is that it separates the effects of the source and the filter. The separation of source and filter makes the extension of cepstral analysis to harmonics-to-noise ratio (HNR) (deKrom 1993, 1995) immune to differences in the fundamental frequency and differences in formant frequencies that often accompany different voice qualities. This is crucial for the investigation of Ju|’hoansi phonation types, given the extreme differences in all three formant frequencies between epiglottalized vowels and modal vowels, making it impossible to control for first formant (F1) frequency and fundamental frequency (F0) differences across different phonation types even when maintaining the same phonemic vowel /a/. This methodology thus also allows me to investigate the acoustic attributes associated with different guttural consonant and vowel types across different vowel contexts.

As noted above, the R1 measure will be affected by both variability in the amplitude of individual harmonics in the spectrum and variability in the duration of individual glottal pulses (jitter). Research is in progress to determine the precise relationship between jitter, H1-H2 and HNR for the Ju|'hoansi data presented in this paper. What is important to note here is that the cepstral measure of HNR is completely independent of the H1 and F1 frequencies. While the amplitude of a harmonic is related to its frequency, the amplitude of any given harmonic does not affect the gamnitude of the first rahmonic in the cepstrum any more than any other harmonic. This is because the gamnitude of the first rahmonic (R1) is determined by the consistency of the amplitudes of the harmonics. In this way, R1 effectively captures any noise in the spectrum, independent of the frequency range where the noise is present. The detection of noise in the spectrum across all frequency ranges is what makes R1 an effective measure of the acoustic similarity of different voice qualities. However, in isolation it is incapable of signaling where in the spectrum the noise is located, and identification of the frequency range where noise is located must be left to other measures, such as H1-H2.

I used the ESPS utility *fftcep* that is associated with XWAVES to create the cepstra used in this study. The program first takes the fast fourier transform (FFT) of the waveform, producing a log-magnitude spectrum, then takes the FFT of the log-magnitude spectrum to produce the cepstrum. The method used here differs from Hillenbrand et al.'s (1994) method that computes the cepstrum from the un-normalized spectrum, and follows deKrom's (1993) original method of computing the cepstrum from a log-magnitude spectrum. Figure 6 shows a sample cepstrum. Notice that the cepstra used here are centered at zero, similar to deKrom's (1993) cepstra. The use of the log magnitude spectrum removes the need for normalization applied by Hillenbrand et al. (1994).

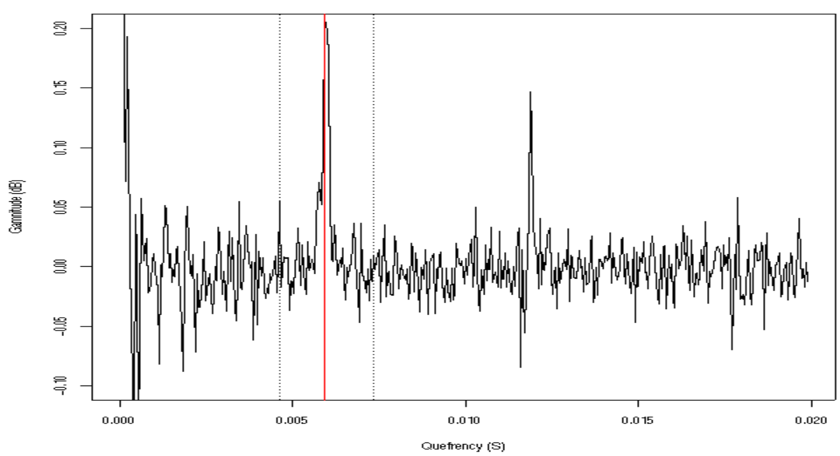


Figure 6: Cepstrum computed over log magnitude spectrum at the midpoint of the vowel in a production of the root word ‖*áá* *'to warm hands by fire'* with the window surrounding the estimated quefrency of R1 (dotted lines) and the chosen R1 peak (solid line) marked (Subject Sp4_m) (corresponding spectrum shown in figure 5)

The spectra and cepstra were computed with 10 coefficients, and 1024 points, appropriate for files sampled at 22050 Hz, using a Hamming window. This resulted in a 46.5 ms window size. Only the real part of the cepstrum was saved. A series of cepstra were computed at a step size of 10 ms from the beginning of the vowel (either the "PR" or "TR" label), until the end of the periodic portion of the vowel ("EP" label). This methodology ensures that the window over which the first cepstrum is computed will be exactly aligned with the beginning of the vowel. Since non-modal phonation types all occur at the C-V transition in Ju|'hoansi, it is critical to not lose any information right at the beginning of the vowel. Recall that the overall low maximum gamnitude of the cepstra in these data are a result of the rather low recording level, which was necessitated by the high frequency energy found in click bursts, in order to avoid clipping.

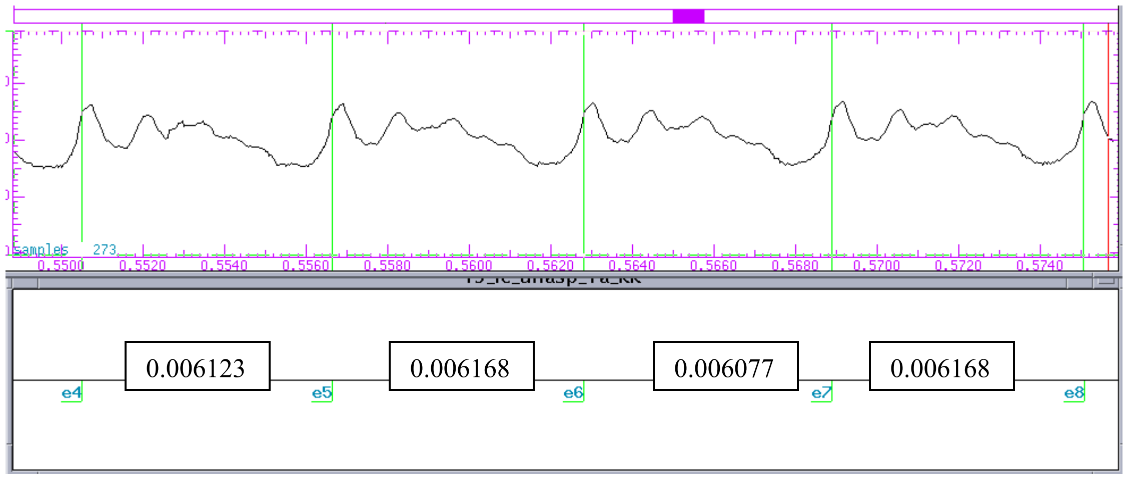
The pitch period durations computed over the interval between pitch epochs were used to estimate the quefrency of the first rahmonic in the cepstrum over each window. Since the quefrency is an analog of frequency, the average pitch period should correspond exactly to the quefrency of the first rahmonic peak. The average pitch period was computed over the same windows used to calculate the cepstra throughout the vowel, including all periods that fall completely within the window. These estimates were then fed into a peak-picking program within the GNU R statistics plotting package, and the peak gamnitude found within a 30 frame window around the estimated first rahmonic was determined to be the exact first rahmonic peak for each window. Figure 6 above shows the window around the estimated first rahmonic, and the actual peak gamnitude that was identified as the first rahmonic. Peak picking was checked visually. Given the initial hand checking of the epochs from which the first rahmonic values were estimated, there were very few cases where errors occurred. Errors were corrected by recalculating the estimate for the first rahmonic by fixing problems with the epoch locations. In this way, robust first rahmonics were identified in the cepstra every 10 ms throughout the duration of the vowel, and the gamnitude values were measured. The amount of checking is extremely important, given the tendency for there to be very low fundamental frequency values, and thus low first rahmonic values, associated with guttural vowels, and vowels following guttural consonants.

The real time of the first identified epoch is the beginning of the window used to calculate the first cepstrum and to estimate the averaged pitch period. These begin times were normalized by subtracting the time of the posterior release of the click (time of the “pr” label) from the window start time, so as to be able to compare across different files that have different start times. Thus, differences in begin times for different consonant types in the results shown here are associated solely with real VOT differences, where

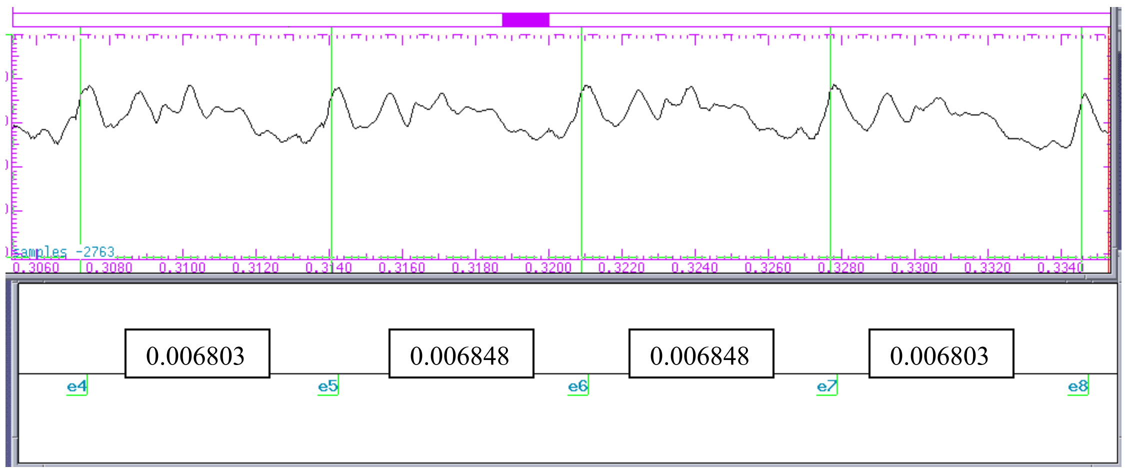
VOT is defined objectively as the first point in the vowel which is voiced (such that it will not be gated out by the ESPS program *get_f0* estimate of voicing), and where two successive periods are similar enough for an epoch to be identified.

3.6 Jitter (PPQ)

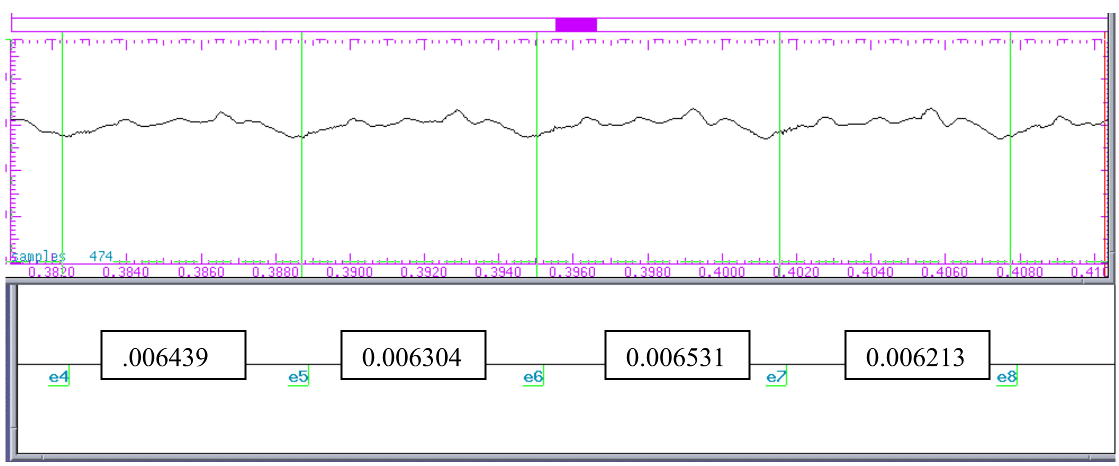
Jitter, the measure of variability of the duration of glottal pulses over a specific domain, is completely independent of F1 frequency. However, it is highly dependent on changes in fundamental frequency, which are found in roots containing all guttural consonants and vowels in Ju|'hoansi (Miller-Ockhuizen 2001). In figure 7(a) below, a waveform associated with a modal lexically high-toned vowel in the word *//áá 'to warm hands'* is shown, along with epoch marks delimiting four individual glottal periods. The durations of each individual period are provided in boxes. As can be seen, there are small differences in the overall duration of each glottal period, which is jitter that is found in all forms of speech (as noted by Klatt and Klatt 1990, if there is no jitter present in speech, it is perceived as odd). In figure 7(b), I provide a waveform associated with the breathy super-low toned root *!ȁʱȁʱ 'red crested korhaan'*. As can be seen, there is some random variation in the duration of each glottal period. Unlike with the jitter associated with the modal vowel, there seems to be an alternating pattern to glottal period durations in the breathy vowel. In figure 7(c), a waveform associated with a glottalized vowel in the rising super-low – low toned root *!ȁˀa` 'dry season'* is provided. As can be seen, there is a steady increase in the duration of each glottal period, which like the breathy voiced vowel also alternates from period to period. Although it is difficult to see with just four glottal periods, there is a steady increase associated with the rising fundamental frequency shape associated with glottalized vowels as well as a more random variation in the shapes of periods. Thus, while jitter is fully independent of vowel height and backness, it is highly dependent on F0 values, since it is difficult to differentiate random variation



(a) Modal monotone vowel (minimal jitter associated with modal voice)



(b) Breathy monotone vowel (high degree of jitter)



(c) Glottalized rising toned vowel (high degree of jitter)

Figure 7: Dependence of jitter on F0

in pitch period duration from more controlled steady increases or decreases in pitch period duration associated with lexically rising and falling tones.

Pinto and Titze (1990) showed that Pitch Perturbation Quotient (PPQ) (Davis 1976), a mean rectified measure of jitter, is more independent of F0 changes than other measures of jitter. Given the differences in F0 associated with all Ju|'hoansi guttural consonants and vowels, PPQ was chosen as the best possible measure of jitter. First, the duration of each of the periods defined as the duration between each two subsequent epoch marks found within the window were calculated. These pitch period durations were used as the raw data for computing PPQ. The PPQ formula is provided in Figure 8.

$$PPQ = \frac{\frac{1}{N-4}\sum_{I=1}^{N-4}\left|\frac{1}{5}\sum_{r=0}^{4} \mathrm{T}o^{(I+r)} - \mathrm{T}o^{(I+2)}\right|}{\frac{1}{N}\sum_{I=1}^{N} \mathrm{T}o^{(I)}}$$

Figure 8: Formula for PPQ (Pitch Perturbation Quotient)

To stands for the duration of the extracted pitch periods, and **N** stands for the number of pitch periods within the window. The relative independence of F0 from jitter in PPQ comes from the smoothing over five periods present in the formula. This measure is referred to as a measure of mean rectified jitter by Pinto and Titze (1990).

However, regression analysis of the results show that even PPQ is dependent on changes in F0, with 5 to 20% of the PPQ found in Ju|’hoansi vowels predicted from the change in F0 associated with the part of the vowel measured. The problem is that PPQ only factors out rises in F0 that occur at the same magnitude throughout the vowel. The patterns of changes in F0 associated with Ju|’hoansi guttural consonants and vowels are much more subtle, and the rate and slope of the curves are different for every type of consonant and vowel observed in this study. These differences make it difficult to factor out changes in F0 without modeling F0 pattern shapes explicitly for each type of guttural consonant and vowel. Given the often large changes in F0 associated with different phonation types in Ju|’hoansi specifically, and more commonly in other languages, jitter is not a good measure of linguistic phonation type differences. The variability in the strength of the relationship between F0 and the PPQ measure of jitter for different speakers suggests that cue trading may also be present among them as found by Abramson and Luangthongkum (2001).

In this study, PPQ was calculated as a measure of jitter over both (a) the first half of the vowel, and (b) the entire vowel, taking either the posterior release of the click ("PR" label), or the transient release associated with the click ("TR" label) if there is one, as the beginning of the vowel. The end of the window for PPQ calculated over the first half of the vowel is the midpoint of the vowel (calculated between the begin and end points of the periodic portion of the vowel), and the end of the window for the entire vowel PPQ measure is the "EP" label itself. It was discovered that controlling for lexical tone, and using only level and not rising lexical tones decreases the amount of F0 change throughout the vowel. However, there is still a large fall in F0 found at the end of the vowel in all tokens (see e.g., Miller-Ockhuizen 1998), and this fall accounts for most of the F0 change found within the level toned roots used in this study. Therefore, only PPQ calculations over the first half of the vowel are meaningful, in that only in these cases can we minimize the changes in F0 over the window being evaluated sufficiently. However, substantial changes in F0 due to differences in initial consonant release type still persist.

While most consonantal effects are local (10-20 ms), voiced aspirated consonants exhibit very large F0 excurses over the first half of the vowel. Since changes in F0 over the first half of the vowel associated with guttural release types cannot be completely controlled for, PPQ is plotted against the total change in F0 found over the first half of the vowel. The change in F0 throughout the window was calculated using the pitch period duration data. First, the initial F0, peak F0 and F0 trough within the window were calculated as the inverse of the duration of the first pitch period, the smallest pitch period, and the largest pitch period contained completely within the window. Then the total change in F0 over the entire vowel was calculated as the sum of the difference between the initial and peak F0 values, and the difference between the peak and minimum F0 values within the window. Since the minimum F0 values were determined by visual inspection to always occur at the end of the vowel for the root types used in this study, the total change in F0 over the first half of the vowel was calculated as the change in F0 between the begin and peak F0 points within the first half of the vowel, summed with the difference between the peak F0 and final F0 values in the window.

4. Experiment 1: Spectral Properties of Ju|’hoansi Guttural Vowels

4.1 Introduction

Experiment 1 tests the hypothesis that all guttural vowels in Ju|’hoansi, including breathy, glottalized and epiglottalized vowels, all contain voice quality cues such as low HNR, either a relatively high or relatively low spectra slope in the low frequency range, and increased jitter. There are several studies that have investigated the voice quality cues associated with vocalic phonation types, however previous studies have never investigated pharyngealized or epiglottalized vowels. Earlier studies focused on the investigation of the spectral properties of breathy vowels found in Gujarati (Fischer-Jorgensen 1967; Bickley 1982), Hmong (Huffman 1987), Mazatec (Ladefoged et al. 1988; Kirk et al. 1993; Silverman et al. 1995; Blankenship 2002) and a related Khoesan language !Xóõ (Bickely 1982; Ladefoged et al. 1988), but Javkin et al. (1987) has investigated the spectral properties of glottalized vowels in Burmese and Blankenship (2002) has studied Mazatec glottalized vowels (Blankenship 2002).

Results reported here show that epiglottalized vowels in Ju|’hoansi display low HNR values, similar to glottalized and breathy vowels, making HNR an ideal measure to capture the acoustic similarity of guttural vowels targeted by the Guttural OCP constraint. Epiglottalized vowels display relatively low values of spectral slope in the low frequency range (H1-H2) when compared with modal vowels, similar to values associated with glottalized vowels, suggesting that there is a constriction somewhere at or near the glottis affecting the vocal fold vibration mode. Ju|’hoansi breathy vowels show the typical greater spectral slope (H1-H2) values relative to modal vowels found in other languages.

4.2 Results

4.2.1 Spectral Slope (H1-H2)

In order to assess the accuracy of the H1-H2 measure, it is necessary to be sure that neither H1 nor H2 is in close enough proximity to F1 to be artificially boosted by the formant in the particular productions used in the experiment. Table V provides the median F1 frequency taken at the midpoint of the vowel and the median peak H1 frequency values for the vowels in the study in the /a/ context. For subjects Sp1_f, Sp2_f and Sp4_m, there is a good separation between H1, H2 and F1 frequencies in all contexts, ruling out the possibility that H1-H2 values will be affected by boosting of either harmonic by F1. Thus, we can be assured that the H1-H2 values reported here are accurately representing spectral slope. However, subject Sp1_m's productions do show close proximity of H2 to F1 for modal vowels, and therefore H1-H2 values for modal vowels should be treated as suspect for this subject. The [oa] and [o] vowel contexts were not considered reliable because of the proximity of H2 to F1, and the collapsing effect seen in the H1-H2 results for these contexts. Therefore, only /a/ context results are reported for the H1-H2 measure in this paper. Peak H1 frequencies reported here were computed as the inverse of the smallest distance between two epoch marks in the vowel.

Vowel type		Modal	Breathy	Epiglottalized	Glottalized
Sp1_f	**F1 (Hz)**	866	382	1066	923
	H1 (Hz)	314	288	287	331
Sp2_f	**F1 (Hz)**	733	734	921	838
	H1 (Hz)	243	216	211	256
Sp3_m	**F1 (Hz)**	392	769	769	579
	H1 (Hz)	200	180	176	198
Sp4_m	**F1 (Hz)**	525	564	707	600
	H1 (Hz)	165	149	144	176

Table V. Median F1 and peak H1 frequency values at the midpoint of the vowel with different vocalic phonation types
(Shaded cells have H1 or H2 values within 50 Hz of F1)

The spectral slope values over the entire duration of the vowel are provided for each subject in Figure 9. As can be seen, all four subjects' productions display much greater H1-H2 values for breathy vowels relative to modal vowels. Glottalized vowels display lower H1-H2 values for both female subjects, as well as for male subject Sp3_m, but subject Sp4_m's productions display similar values for glottalized vowels and modal vowels, with slightly higher values for glottalized vowels than modal vowels. Epiglottalized vowels display relatively lower H1-H2 values for three of the four subjects, with only subject Sp1_f displaying slightly higher H1-H2 values for epiglottalized vowels than modal vowels.

The small separation between modal, glottalized and epiglottalized vowels on this measure for subject Sp3_m can be attributed to the boosting of H2 in the modal vowels due to the proximity of H1 to F1 for these vowels (refer to Table V above). Subject Sp4_m exhibits the smallest degree of separation on this measure, which suggests that he does not use this cue dimension much to signal contrasts. Still, the breathy vowels and modal vowels are differentiated in the same direction found with the other subjects. That is, breathy vowels exhibit greater spectral slope values relative to modal vowels, and glottalized vowels exhibit lower spectral slope values.

Subjects Sp3_m and Sp4_m's productions show an increase of spectral slope at the end of the root for all vowel types, which suggests that they may tend to lapse into breathiness at the end of utterances. Since the roots here are produced in isolation they constitute an utterance. Subjects Sp1_f and Sp2_f show a similar pattern for modal, glottalized and epiglottalized vowels, but not for breathy vowels, which tend to become less breathy at the end of the root (and utterance).

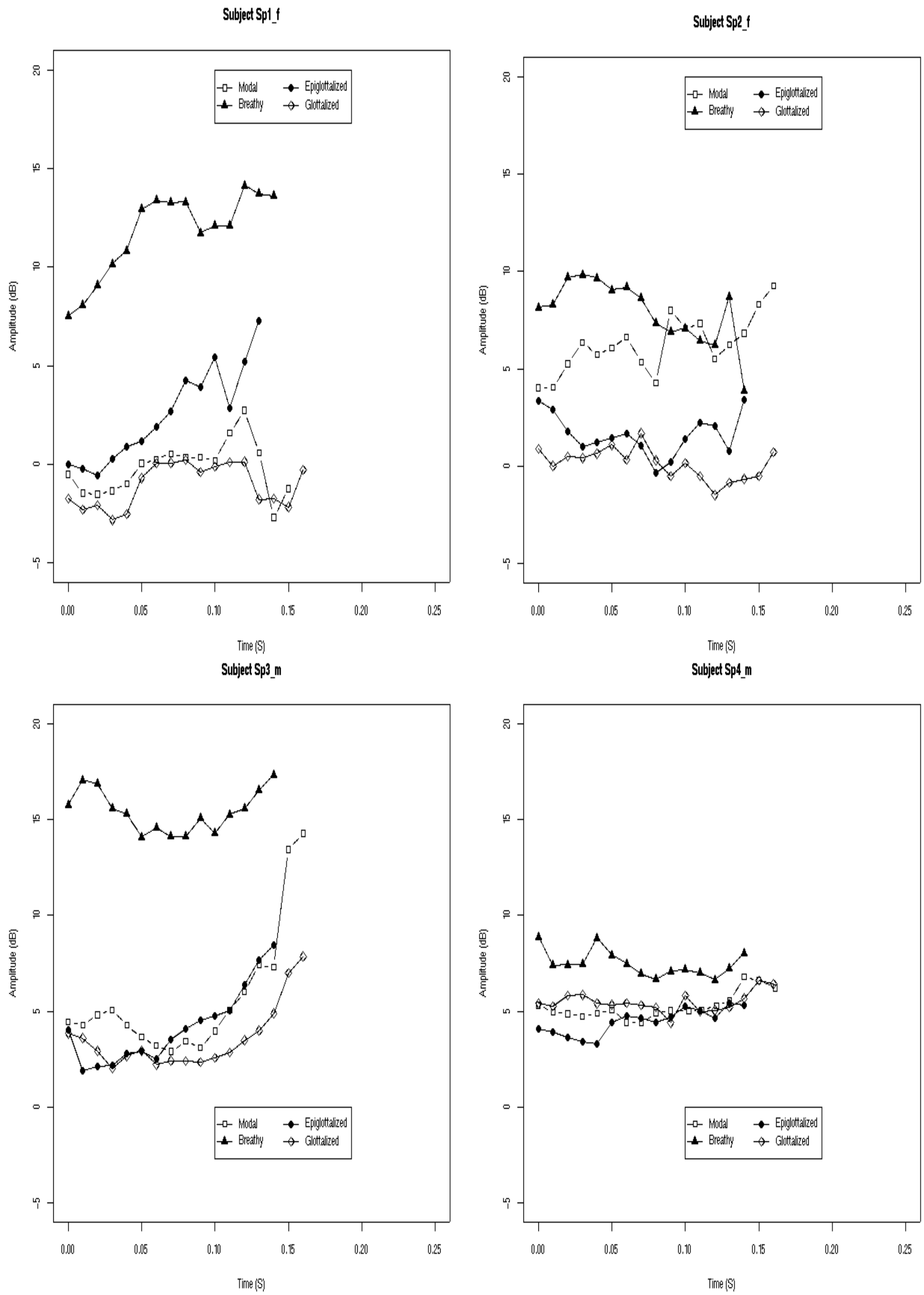


Figure 9: Median spectral slope values associated with guttural vowel types in the [a] context

4.2.2 HNR (R1)

Guttural vowels all exhibit relatively lower HNR values, as measured by the gamnitude of the first rahmonic peak (R1), than modal vowels. Figure 10 provides a plot of HNR values on the vertical axis against time on the horizontal axis for the guttural vowel contrasts found in Ju|'hoansi produced by all four subjects. Notice that the two female subjects, Sp1_f and Sp2_f's breathy and epiglottalized vowels have relatively lower R1 values throughout their duration than the modal vowels. The glottalized vowels are relatively lower in gamnitude over the first half of the vowel, but they rise in gamnitude by the end of the vowel. In fact, they overshoot the HNR values found with modal vowels by the end of the root. This is in contrast to the H1-H2 values found for glottalized vowels, which as we saw in section 4.2.1, remain low throughout the vowel. The discrepancy between the two measures is surprising given that the R1 measure of HNR is sensitive to noise in any frequency range in the spectrum.

The range of R1 values employed by the male subjects is much smaller than the range utilized by the female subjects, making it more difficult to see the clear separation between the different guttural vowel types. This may be attributed to the generally rough quality of subject Sp3_m's voice. Notice that all of his vowel types are very low in gamnitude compared with the other subjects, even though the recording level and the distance from microphone to mouth was the same for all subjects. Despite the smaller acoustic space employed by subject Sp3_m, he still maintains the same relationship between R1 and vowel type found in the other subjects' productions. Subject Sp4_m shows a similar pattern. For Sp4_m, all of the roots containing guttural vowels cluster together nicely, showing that a low R1 is a nice unifying characteristic of guttural vowels.

In addition to the general relative HNR values which correspond to vowel phonation type, notice that all vowel types fall in HNR at the end of the root. Again, this

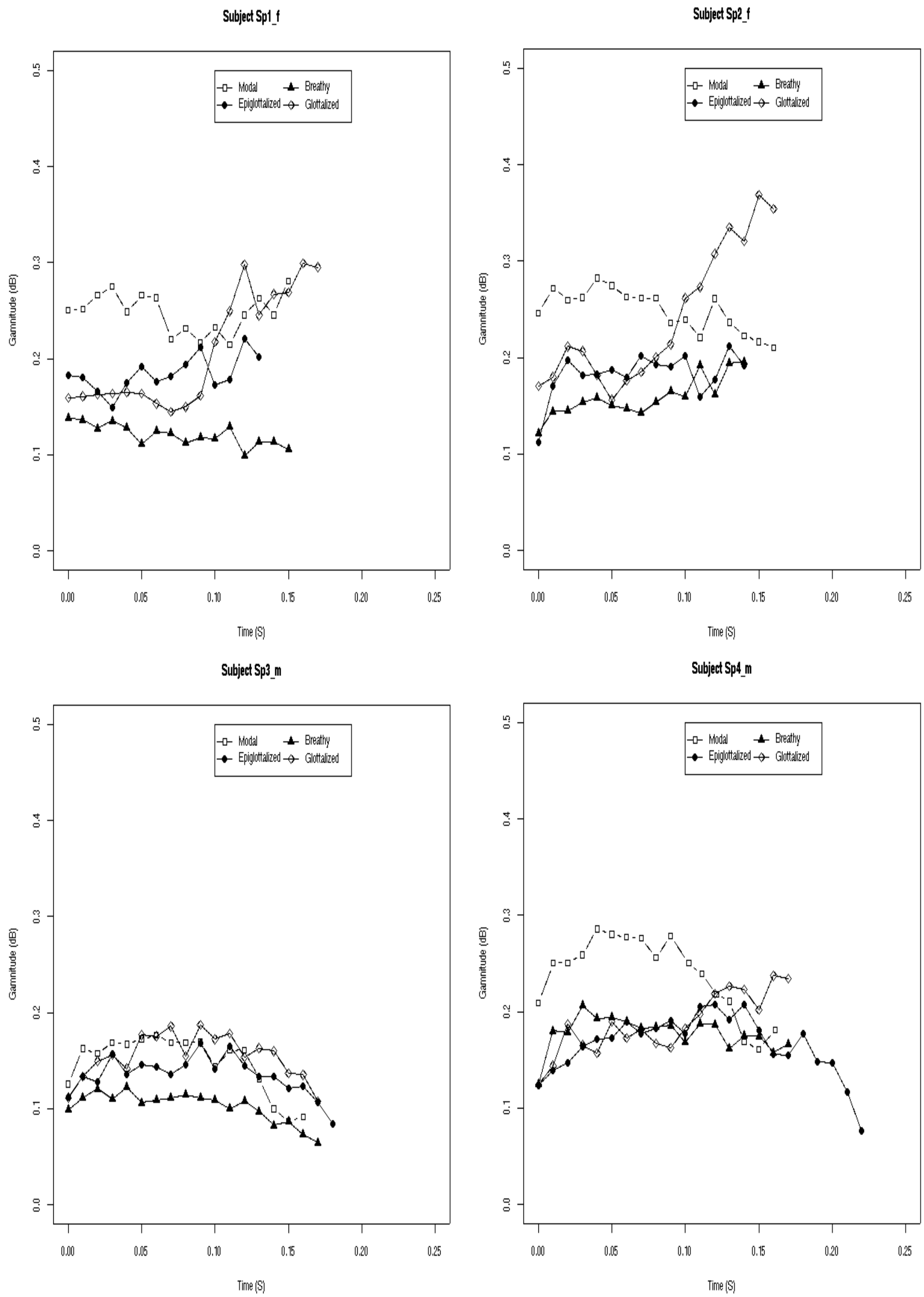


Figure 10: Median gamnitude of the first rahmonic peak (R1) in the cepstrum associated with guttural vowel types in the [a] context

can be attributed to prosodically conditioned breathiness, given that these roots are uttered in isolation and thus constitute an utterance.

It is interesting to note that subject Sp1_f shows the best separation among the four subjects on this measure. Subject Sp3_m on the other hand shows very limited separation on this measure, and subject Sp4_m shows a large separation. Comparing figures 9 and 10, we can see that these patterns are opposite the patterns found in the H1-H2 measure, where Sp3_m shows a larger separation than Subject Sp4_m. Since subject Sp4_m shows a good separation between H2 and F1 in all vowel contexts, and subject Sp3_m only shows proximity of H2 to F1 in the modal vowel context (see Table V below), the differences on the two measures can not be attributed to unrealistic spectral slope values. Thus, it is clear that the two measures are capturing different things, although there may still be some overlap.

4.2.3 Jitter

Figure 11 displays PPQ plotted against change in F0 over the first half of the vowel for all four guttural vowel types for all four subjects. Regression lines were calculated using PPQ as the independent variable, and the change in F0 over the first half of the vowel as the dependent variable. Notice that there is a linear relationship between the amount of F0 change and the amount of jitter found in the vowel. Only one regression line is provided for all vowel types, as separate regression lines did not show improved predictive power. As seen in the graphs, the jitter is less than expected just given F0 change throughout the window being evaluated. This is a result of the fact that the change in F0 over the first half of the vowel is too great to allow an objective measure of the amount of jitter present. Interestingly, three of the subjects show a large amount of variation in the amount of F0 change found in the vowel. For example, the speech of both of the two female subjects, Sp1_f and Sp2_f, displays F0 changes ranging from 0 to 200 Hz within the vowel, although subject Sp1_f's productions show much more variation than Sp2_f's productions.

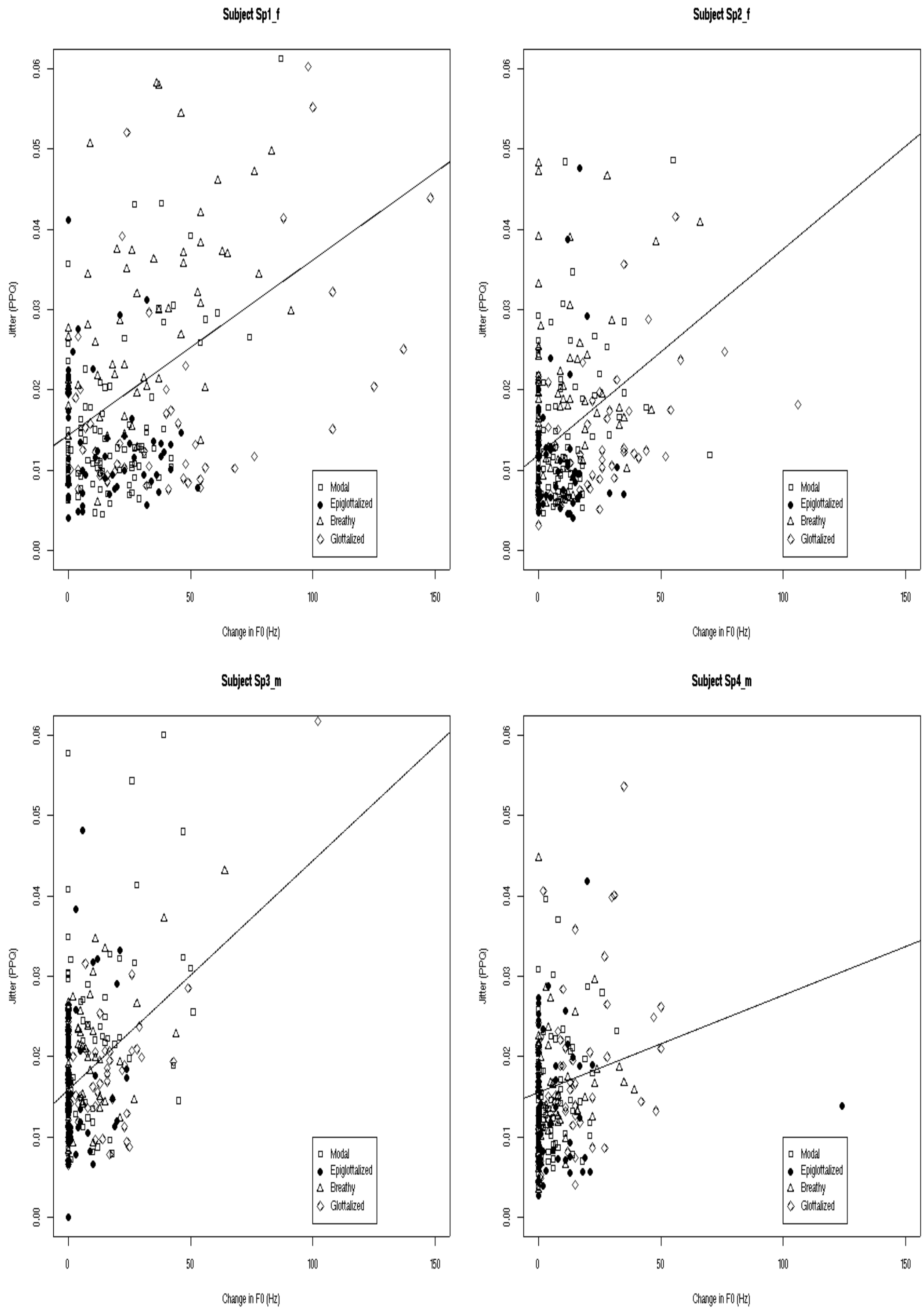


Figure 11: Change in F0 plotted against PPQ over the first half of the vowel associated with guttural vowel types in [a] and [o] contexts

Subject Sp3_m seems to have a slightly smaller pitch range than the two female subjects. For subjects Sp1_f and Sp2_f, the glottalized vowels clearly have the largest amount of F0 change, but there is not a correspondingly larger amount of PPQ present in the signal. For the two male subjects, Sp3_m and Sp4_m, the glottalized vowels do not display a larger change in F0 than is found in the other vowel types. Although for these two male subjects, the amount of jitter does appear to be greater than is found in the other vowels. Subject Sp4_m's data is of interest, as he appears to have the smallest pitch range, with F0 change values ranging between 0-100 Hz, with a few outlying tokens displaying up to 150 Hz change in F0. However, the PPQ values seem to vary randomly, and are not correlated well with vocalic phonation type. Thus, it appears that jitter, random variation in period length, is not a cue signaling guttural vowel contrasts.

I calculated two regressions for each subject in order to assess the dependence of PPQ on the amount of change in F0 and phonation type of the vowel. In the first regression, PPQ is the dependent variable, and the change in F0 over the first half of the vowel is the single independent variable. In the second regression, the dependent variable is PPQ, and the independent variables were change in F0, as well as a set of binary indicator variables specifying whether the vowel was breathy, epiglottalized, or glottalized. The R-squared values for each regression, as well as the values for Pr > /t/ values are provided in Table VI for the vocalic phonation types investigated. For all subjects, the change in F0 was a significant predictor of PPQ, meaning that the PPQ measure, hailed as the most independent measure of irregular changes in pitch period duration described as jitter, is dependent on the change in F0. For one of the subjects (Sp1_f), the change in F0 accounts for more than 15% of the variance, as noted by the R^2 value for the first regression. For the other three subjects, (Sp2_f, Sp3_m and Sp4_m), the change in F0 is a less strong predictor of the variance in the change in F0 over the first half of the vowel, accounting for less than 7% of the variance. For the two female subjects, Sp1_f and Sp2_f, there was also a contribution of vocalic phonation type in predicting PPQ. For subject Sp1_f, only the categorical variable of breathy vs. non-breathy contributes to the prediction of PPQ. For subject Sp2_f, only glottalization and epiglottalization are predictors of PPQ. In the second regression, vocalic phonation type only predicts an additional 9% of the variance above change in F0 for subject Sp1_f, and an additional 3% of the variance for subject Sp2_f. Vocalic phonation type does not significantly predict any additional variance above the change in F0 for either male subject. Glottalization, which is expected to have the largest effect on jitter given earlier studies such as Ladefoged et al. (1998) is only a useful predictor of jitter for one of the subjects in this study (Subject Sp2_f). This is likely due to the localization of the glottalization gesture to a small interval at the center of the vowel, causing the effect on periodicity to be overshadowed by the smoothing factor used in PPQ and the relatively large window used in the calculation. The bottom line of Table VI reports the results of an ANOVA, testing whether there is a significant difference between the two regression models. As can be seen, vocalic phonation type only contributes significantly to the prediction of PPQ for the two female subjects.

VOCALIC PHONATION TYPES Pr > /t/				
Subject	Sp1_f	Sp2_f	Sp3_m	Sp4_m
Change in F0	<.001	<.001	<.001	<.01
R^2 (model A)	0.16	0.06	0.05	0.04
Change in F0	<.001	<.01	<.01	<.015
Breathiness	<.001			
Epiglottalization		<.05		
Glottalization		<.05		
R^2 (model B)	0.25	0.09	0.06	0.05
Pr (> F)	<.001	<.05		

Table VI. Regression results for vocalic phonation type (Models A and B)

The change in F0 is itself correlated with the vocalic phonation type, as is shown by the results of a third regression, reported in Table VII. The results of a fourth regression, assessing the independent predictability of vocalic phonation type on PPQ, are reported in Table VIII.

Subject	Sp1_f	Sp2_f	Sp3_m	Sp4_m
Breathy		<.01		<.01
Epiglottalization		<.001		<.001
Glottalization			<.001	
R^2 (model C)	.03	.13	0.09	0.13

Table VII. Change in F0 ~ vocalic phonation type

Subject	Sp1_f	Sp2_f	Sp3_m	Sp4_m
Breathy	<.01			
Epiglottalization		<.001		
Glottalization				
R^2 (model D)	.10	.05	.02	.03

Table VIII. PPQ ~ vocalic phonation type

Given the results of the regressions in Tables VII and VIII, I also calculated a partial regression to ascertain whether vocalic phonation type predicts a substantial proportion of the jitter once the mutual correlation with the F0 change is taken into account. I calculated this partial regression by correlating the residuals between model C and model D. The correlation between the residuals of the two models is substantial for all subjects, as shown by the results provided in Table IX. Interestingly, the correlation is the largest for subject Sp1_f, the same subject for whom breathiness is a significant predictor of jitter over and above change in F0 as shown by the increase in R^2 between models C and D.

	Sp1_f	Sp2_f	Sp3_m	Sp4_m
Correlation	0.40	0.21	0.20	0.15

Table IX. Correlation of residuals between model C (change in F0 ~vocalic phonation type) and model D (PPQ ~ vocalic phonation type)

It is interesting to note that breathiness is the vocalic phonation type that coincides with the smallest change in F0 over the first half of the vowel when compared with modal vowels. Glottalized vowels have the largest change in F0 over the first half of the vowel. Thus, the fact that glottalization on vowels is the strongest predictor of change in F0, but the smallest predictor of jitter probably means that the change in F0 overshadows any jitter effect found in this phonation type. Although breathiness probably has a very small association with jitter (as shown by previous studies such as Ladefoged et al. 1988), it is picked up by the measure of PPQ, because the change in F0 over the vowel is small enough to allow the random fluctuation in period duration to be identified.

4.3 Discussion

Spectral slope, measured by the difference between the amplitude of the first and second harmonics in the FFT spectrum, shows the greatest separation between glottalized and epiglottalized vowels on the one hand, and breathy vowels on the other hand, for subjects Sp1_f and Sp3_m, with differences being smaller for subjects Sp2_f and Sp4_m. It is not clear what might predict the differences in these subjects' productions, as Sp1_f and Sp4_m have the clearest voice qualities, and Sp1_f and Sp2_f are female subjects. Thus, neither gender nor overall voice quality account for the differences found. The differences appear to be individual differences reflecting individual cue trading relationships.

HNR, measured via R1 gamnitude, nicely captures the similarity of all guttural vowels, though low R1 values can be due to both low amplitude higher harmonics, and the presence of jitter in the time domain (deKrom 1993). Results show that guttural vowels are all much noisier than non-guttural (modal) vowels. All four subjects in this study show similar patterns, with modal vowels displaying the highest R1 values, followed by glottalized vowels, epiglottalized vowels and breathy vowels. Thus, HNR unifies all guttural vowels, and may be the relevant factor that is picked out by the phonology in determining their patterning together in the Guttural OCP constraint and the tonal depression constraint, both of which were described above in section 2.4.

The results of this study show that it is difficult to assess the amount of jitter present in a signal using current measures when there are any somewhat large, more controlled changes in F0 in the window being evaluated. Given the difficulties associated with jitter calculations (the need for a large window with steady-state F0), PPQ and other measures of jitter may not be independently useful as measures of linguistic voice quality. However, PPQ is useful in providing evidence to differentiate between possible interpretations of results of less pitch-dependent, but also less specific measures such as the cepstral HNR measure used here. That is, the assessment of PPQ along with R1 and H1-H2 measures allows a better understanding of the contribution of different types of aperiodicity to HNR. The results of this experiment show that Ju|'hoansi vocalic phonation types all have a low degree of PPQ associated with them. Guttural vowels show an increase in PPQ, but most of the variation in PPQ is due to larger changes in F0 over the first half of the vowel, which is the window used for the computation of PPQ in this study. Glottalized and epiglottalized vowels have the largest changes in F0 over the first half of the vowel, and thus these vocalic phonation types contribute the most to the prediction of the variation in change in F0. More random variation in the glottal period duration over the first half of the vowel is only found with breathy vowels, which is contrary to our expectation that glottalized vowels should display more jitter. It is likely that more random variation associated with glottalized vowels and epiglottalized vowels are over-shadowed by the larger more controlled changes in F0 over the first half of the vowel.

The strength of the relationship between change in F0 and PPQ for some speakers suggests that F0 may be used as a cue for guttural vowels in Ju|'hoansi by some speakers. However, the variability of the correlation for different speakers suggests that there may be cue trading found between jitter and more controlled changes in F0.

5. Guttural Coarticulation

5.1 Introduction

A second aim of this study is to investigate the hypothesis that guttural consonants in Ju|'hoansi (aspirated, uvularized, glottalized and epiglottalized consonants) display guttural coarticulation, resulting in voice quality cues being realized on the following vowel. A related sub hypothesis is to investigate whether click consonants articulated with the velaric airstream mechanism employ the same type of laryngeal coarticulation that has been shown to occur with pulmonic consonants.

There is a growing body of literature investigating acoustic voice quality cues associated with vowels following laryngeally articulated consonants associated with laryngeal coarticulation (Löfqvist and McGowan 1992; Gobl and NíChasaide 1999; Watkins 1999; Cho, Jun and Ladefoged 2000; Blankenship 2002). Jessen and Roux (2002) have also looked specifically at laryngeal coarticulation following so-called voiced click consonants in Xhosa (a member of the Nguni cluster of the Bantu family), which appear to be articulated with an open glottal configuration. Jessen and Roux (2002) showed that these clicks display breathy voice on the following vowel in Xhosa through laryngeal coarticulation.

The current study expands the investigation of laryngeal coarticulation by looking at coarticulation involved with aspirated, uvularized, epiglottalized, and glottalized click consonants. Given the parallelism of voice quality cues found following laryngeal and pharyngeal consonants, I term the spread of voice quality cues from this class of consonants to following vowels guttural coarticulation. Experiment 2 thus extends the investigation of laryngeal coarticulation in two ways: first, by adding uvularized and epiglottalized consonants to the set of consonants that spread voice quality cues to the following vowel, and second, by comparing coarticulation of aspirated and glottalized click consonants to aspirated and glottalized plosive consonants.

5.2 Results

5.2.1 Spectral Slope (H1-H2)

Given the dependence of the amplitude of H1 and H2 on the proximity of the frequency of H1 and H2 to the first formant (F1) (Ladefoged et al. 1988; NíChasaide and Gobl 1997), it is important to investigate the possibility of H1 and H2 frequencies being within the range of F1. As seen by the median F1 and median peak H1 frequencies taken at the midpoint of the vowel following each click type for each subject provided in Table

X below, using the /a/ vowel allows enough separation between H1 and F1 for most consonant types and most subjects to be confident that there is no boosting of either H1 or H2 from proximity to F1. However, the F1 and H1 values in vowels following unaspirated, uvularized, and glottalized clicks do show that H1-H2 values should be treated as suspect in these contexts.

Consonant Type		**Voiceless Unaspirated**	**Voiced Unaspirated**	**Voiceless Aspirated**	**Voiced Aspirated**	**Voiceless Uvularized**	**Voiced Uvularized**	**Epiglottalized**	**Glottalized**
Sp1_f	**F1 (Hz)**	924	756	899	825	783	606	781	623
	H1 (Hz)	319	313	340	314	341	284	333	363
Sp2_f	**F1 (Hz)**	817	691	777	743	781	733	798	732
	H1 (Hz)	244	237	313	235	253	218	252	263
Sp3_m	**F1 (Hz)**	378	551	568	704	450	656	527	401
	H1 (Hz)	206	180	205	176	208	186	226	223
Sp4_m	**F1 (Hz)**	522	540	614	595	559	597	629	513
	H1 (Hz)	170	157	179	157	177	155	181	178

Table X. Median peak F1 and H1 frequency values at the midpoint of vowels following consonants with different phonation types (Shaded cells have H1 or H2 values within 50 Hz of F1)

Figure 12 shows the spectral slope values of vowels following root-initial voiceless guttural consonants for all four subjects. First, notice that vowels following voiceless aspirated clicks display higher H1-H2 values than are found in vowels following the voiceless unaspirated clicks for all four subjects. For subjects Sp1_f, Sp3_m and Sp4_m, this difference is maintained throughout the vowel, while for subject Sp2_f, the difference is very transient and is only found at the very beginning of the vowel, and then falls quite a bit lower than the modal vowel over the remainder of the vowel. Vowels following initial glottalized and epiglottalized clicks display lower H1-H2 values than are found following unaspirated click consonants throughout the vowel for subjects Sp1_f, Sp2_f, and Sp3_m. However, for subject Sp4_m's productions, all vowels following guttural consonants begin with greater spectral slope values than are found in vowels following unaspirated clicks, but fall to lower spectral slope values throughout the remainder of the vowel. That is, laryngeal coarticulation following glottalized clicks is much more transient for subject Sp4_m.

We expect vowels following uvularized consonants to display greater spectral slope values following the fricated portion of the click release, given the abducted glottal posture presumed to be necessary in order to produce enough pressure to create frication noise. While vowels following uvularized clicks produced by subjects Sp1_f and Sp3_m do display relatively greater spectral slope values just after the onset of voicing, the spectral slope of these vowels falls over the entirety of the vowel ending with quite low spectral slope values. Subject Sp2_f also displays relatively greater spectral slope at the beginning of the vowel following uvularized consonants, which falls throughout the vowel, but the value is below that of unaspirated consonants at the same time point. Subject Sp4_m's productions do not show much separation between vowels following uvularized, glottalized and epiglottalized consonants.

Note that subjects Sp2_f and Sp3_m show the widest separation on the spectral slope measure, while subject Sp4_m shows the smallest separation. As we will see in section 5.2.2, this is opposite of the separation patterns found for these two subjects in the HNR results. For subject Sp3_m, the larger separation on this dimension might help cue the contrasts that are not marked very well in HNR given his overall rough voice quality.

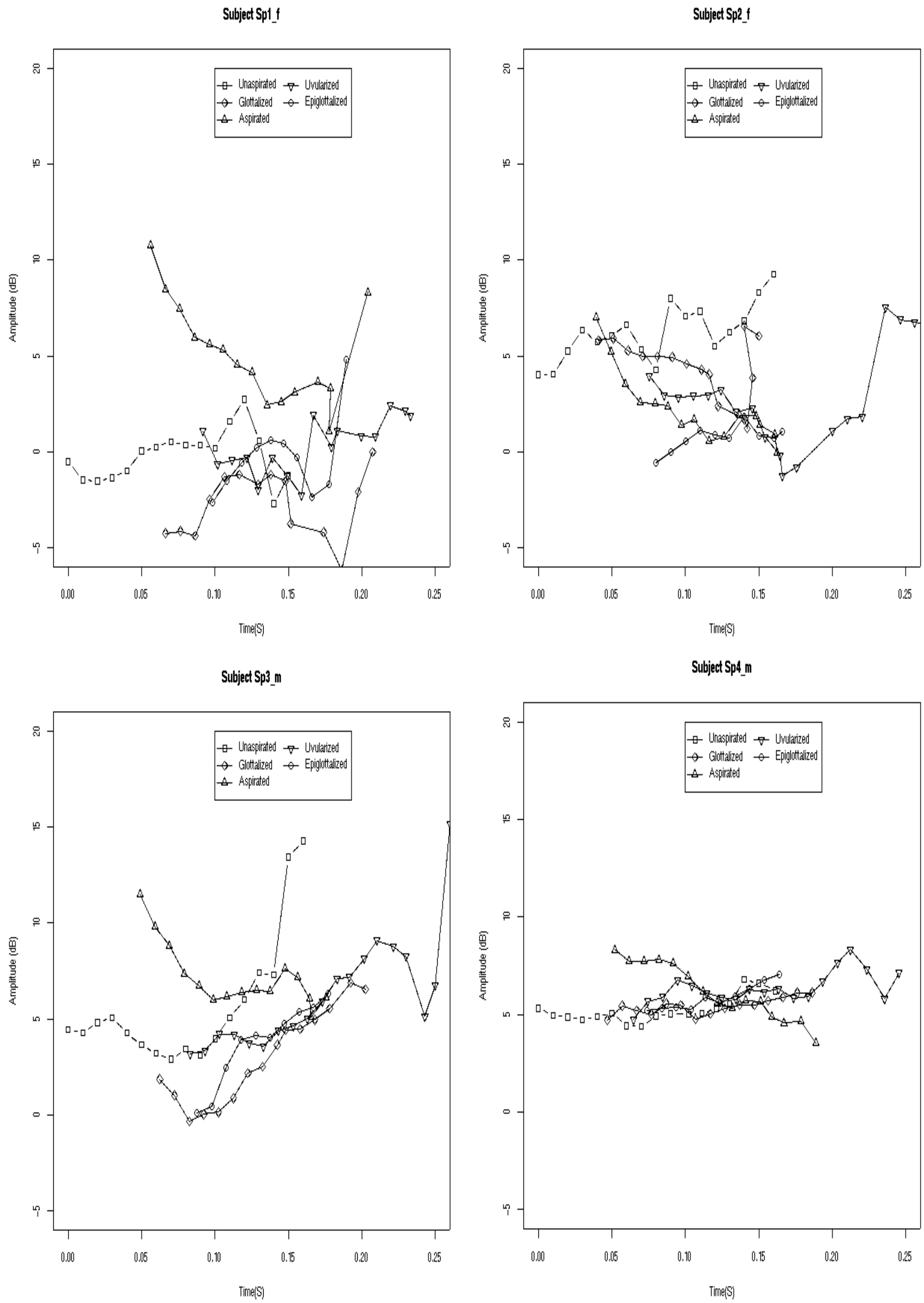


Figure 12: Median spectral slope values associated with voiceless consonant initial roots in the [a] context

Subject Sp4_m, on the other hand, has a clear ringing voice, and his productions exhibit a good separation on the cepstral measure. Thus, there is evidence that cue trading may be occurring on the dimension of the cepstral HNR measure, and the dimension of spectral slope, measured via H1-H2. However, the evaluation of spectral slope values of subject Sp3_m's productions needs to be balanced by the possibility of H2 boosting in vowels following voiceless unaspirated, uvularized and glottalized click consonant initial contexts (see Table X above). Boosting would decrease the difference between the amplitudes of H1 and H2 leading to lower spectral slope values in these contexts. For example, the low spectral slope values found following uvularized consonants for subject Sp3_m can be attributed to the boosting of H2 in this context, leading to a smaller H1-H2 value than expected.

The spectral slope values associated with vowels following voiced guttural consonants are largely similar to those found on vowels following voiceless guttural consonants, although subject Sp2_f's productions exhibit greater spectral slope values for voiced aspirated consonant initial roots that fall throughout the vowel, as shown in Figure 13. The effect is slightly greater for the voiced aspirates than for the voiceless aspirates. The vowels following uvularized consonants and glottalized consonants do not seem to be much different on this dimension from the non-guttural unaspirated voiceless and voiced consonant initial roots. The other three subjects' productions display greater spectral slope values for voiced aspirated and voiced uvularized initial roots than unaspirated initial roots throughout the duration of the vowel. Thus, it appears that spectral slope, measured via H1-H2, is a stronger acoustic correlate of guttural coarticulation involving voiced guttural consonants than it is for guttural coarticulation involving voiceless guttural consonants. In order to achieve pre-voicing, the glottis must be fairly adducted until right up to the oral constrictions involved in the clicks. This may result in later opening of the glottis to achieve aspiration and uvularization noise, which could also coincide with later closing of the glottis and increased acoustic correlates of the glottal aperture associated with the release properties on the following vowel.

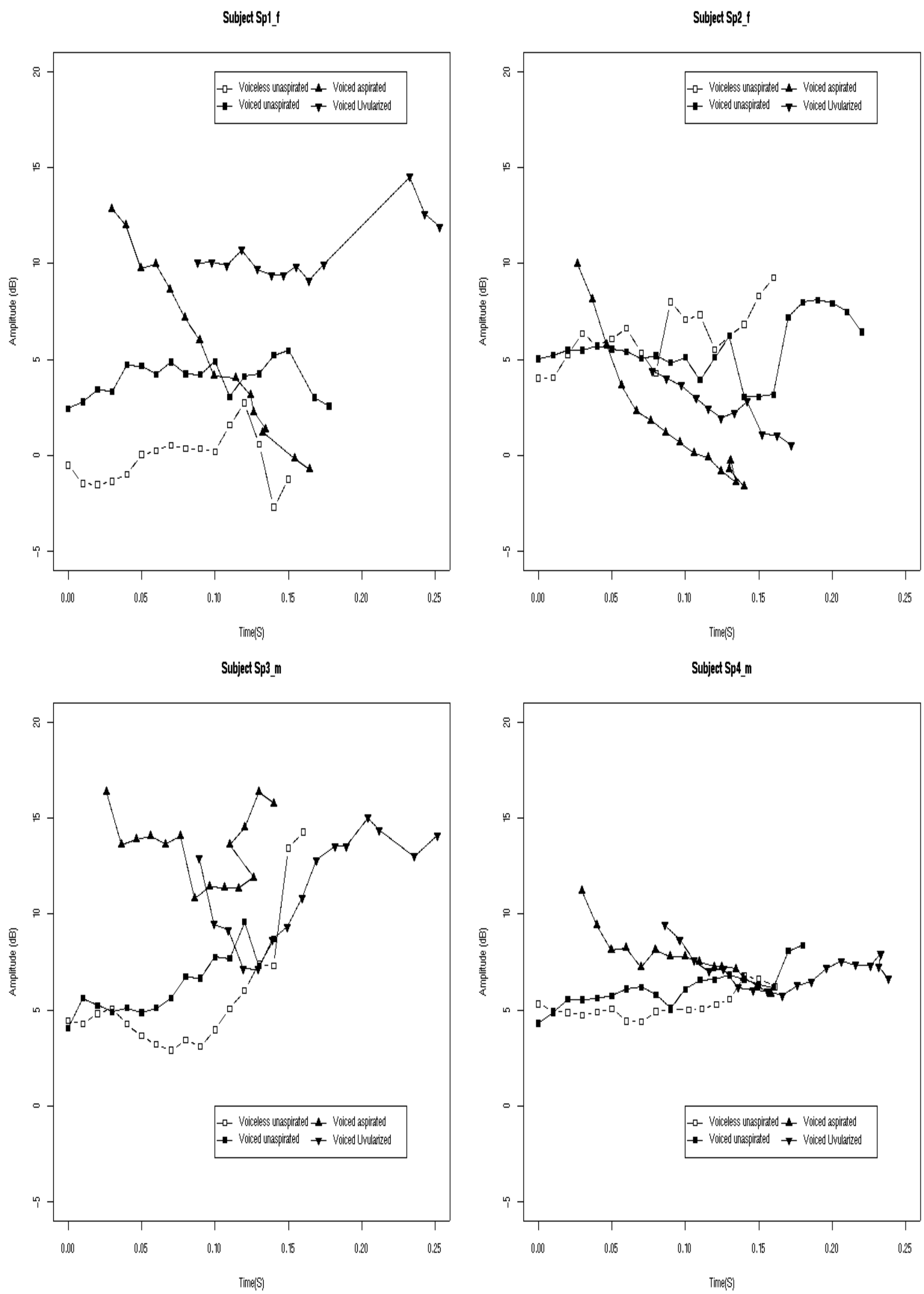


Figure 13: Median spectral slope values associated with voiced consonant initial roots in the [a] context

It is also interesting to note that voiced aspirates have shorter VOT lag values than voiceless aspirates, as can be seen by the earlier recognition of epochs by the ESPS *epochs* program (which also implies earlier positive probability of voicing values since the voiceless portions of the signal were gated out using probability of voicing values). Voiced aspirates also display higher H1-H2 values at the earlier recognized points, but in fact, at comparable time points, vowels following voiced aspirates display lower H1-H2 values than vowels following voiceless aspirates. This is consistent with the expectation that the vocal cords should be more closely approximated in the articulation of voiced aspirates in order to achieve voicing, and with the more closely approximated vocal folds seen following voiced aspirates than following voiceless aspirates in my own articulations (Miller-Ockhuizen In Progress).

Similar differences in VOT are not found between voiceless and voiced uvularized consonants. In fact, the VOT values for both consonant types are comparable for all subjects. In addition, the H1-H2 values at comparable time points are actually higher in vowels following voiced uvularized consonants than in vowels following voiceless uvularized consonants for subjects Sp1_f, Sp3_m and Sp4_m. They are quite similar for subject Sp2_f. Subject Sp1_f does show a somewhat greater difference in the H1-H2 values found following voiced unaspirated clicks and voiceless unaspirated clicks, but the other subjects do not show any marked difference.

5.2.2 HNR (R1)

Figure 14 plots the median gamnitude of the first rahmonic peak within the cepstrum on the vertical axis, and real time on the horizontal axis for all four subjects' productions of words with initial voiceless guttural click consonants. As can be seen in the figure, vowels following voiceless guttural consonants all display upward sloping HNR curves over the first part of the vowel, as measured by R1. For subjects Sp2_f, Sp3_m and Sp4_m, the gamnitudes of the first rahmonics associated with vowels following all guttural consonants are well below those found following the voiceless unaspirated consonants. For subject Sp1_f, vowels following all guttural consonants

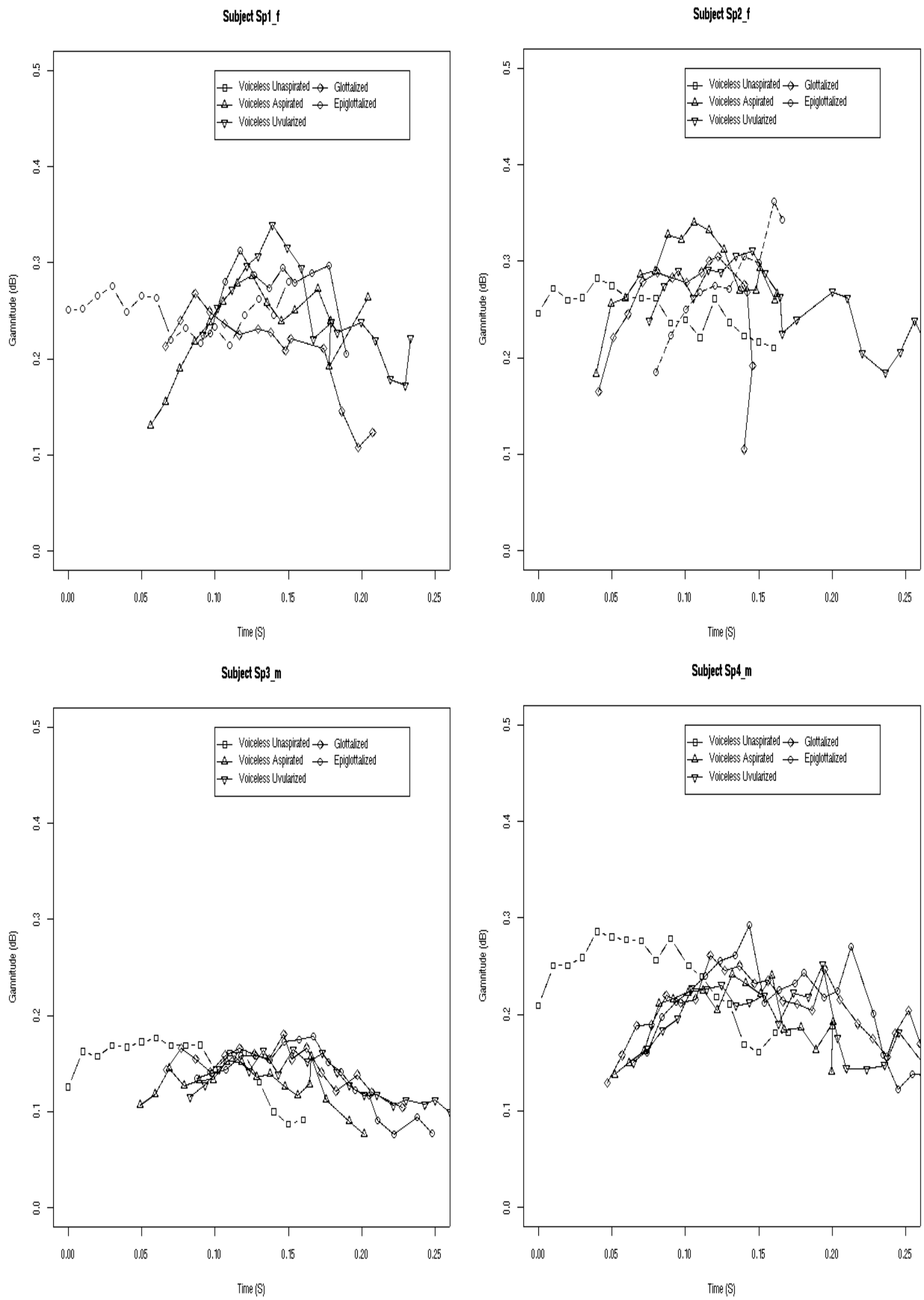


Figure 14: Median gamnitude of first rahmonic peak (R1) in voiceless guttural consonant initial roots in the [a] context

show an upward sloping R1 value that differs markedly from the flat, stable R1 values found in vowels following unaspirated consonants. However, only the vowels following voiceless aspirated consonants begin well below the R1 values found following voiceless unaspirated consonants. For subjects Sp2_f and Sp4_m the results are even clearer than they are for subject Sp1_f. All of the voiceless guttural click initial roots start much lower in gamnitude much later in the root.

Vowels following voiced guttural click consonants also display an upward slope in R1 over the initial part of the vowel, as can be seen in Figure 15 for all subjects. Additionally, the vowels following voiced aspirated and voiced uvularized clicks that behave as gutturals are much lower in gamnitude initially than the vowels following voiced unaspirated clicks. While vowels following voiced unaspirated clicks also display somewhat lower gamnitude values than are found following voiceless unaspirated clicks, the difference is not as marked as that found in vowels following guttural clicks. Both voiceless and voiced aspirated click initial roots start much lower in gamnitude than the voiceless and voiced unaspirated click initial roots. Both aspirated root types rise in gamnitude throughout the vowel, with the voiceless aspirated initial roots surpassing the voiceless unaspirated initial roots in gamnitude about a third of the way through the root. The voiced aspirated initial roots take longer to rise, but probably attain peak gamnitude at around the same time as the voiceless aspirated initial roots. The voiceless uvularized click initial roots start out at about the same gamnitude as found with the voiceless unaspirated click initial roots, but they rise in gamnitude over the first half of the root, and fall again at the end. This contrasts with voiceless unaspirated click initial roots, which maintain a stable gamnitude level throughout the vowel.

The voiced unaspirated click initial roots also maintain a steady gamnitude level throughout the root, although they are slightly lower in overall gamnitude than their voiceless counterparts. The voiced uvularized click initial roots are also fairly steady in gamnitude throughout the root, falling slightly throughout the root, but they are much lower in gamnitude throughout their duration when compared with voiceless and voiced unaspirated click initial roots. In general, the guttural click initial roots all rise from a lower gamnitude at the beginning of the root, or stay at a very low gamnitude throughout

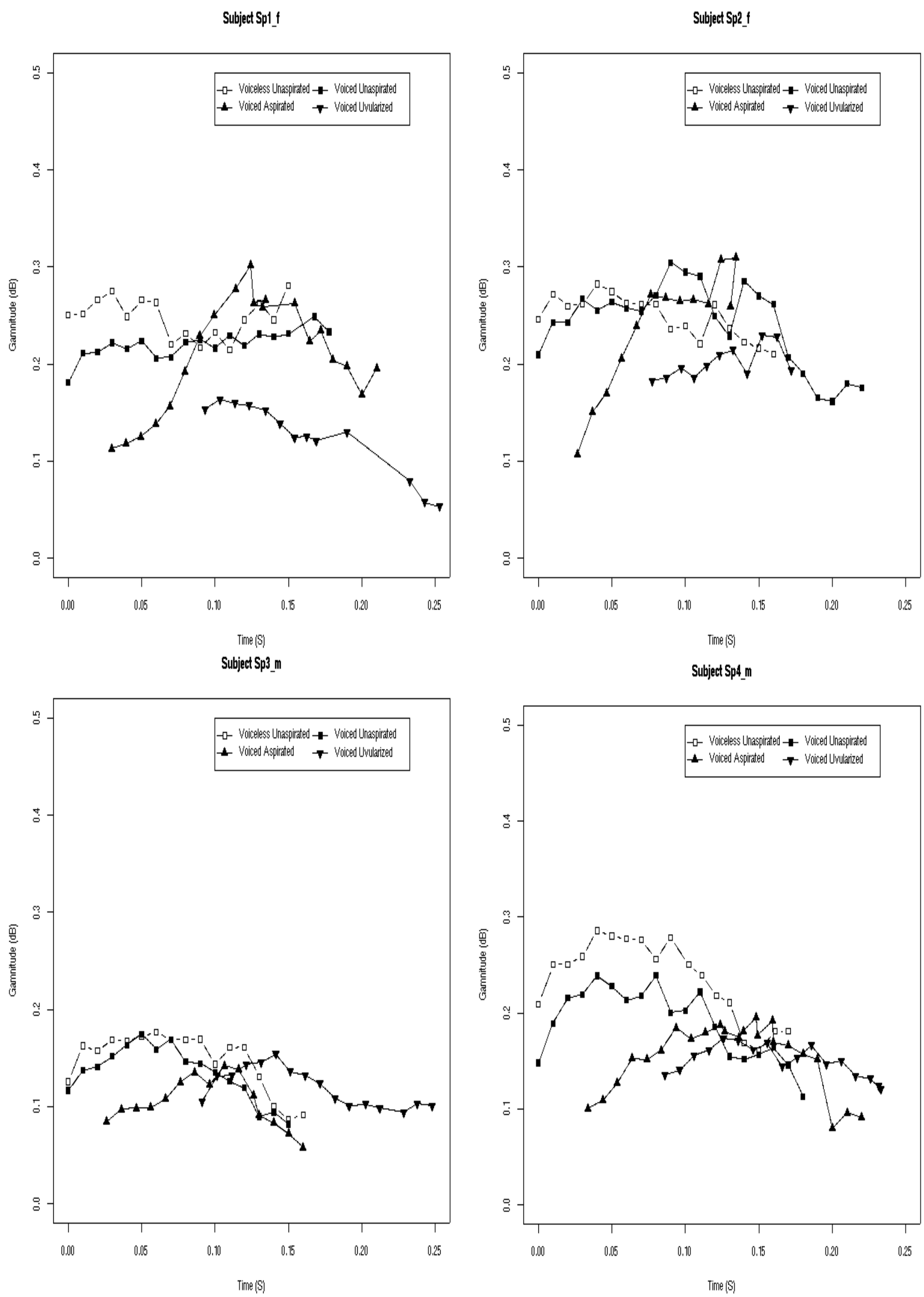


Figure 15: Median gamnitude of first rahmonic peak (R1) in voiced consonant initial roots in the [a] context

the root. In contrast, both voiceless and voiced unaspirated click initial roots maintain a steady relatively high gamnitude throughout the root, with the voiceless unaspirated click initial roots displaying the highest gamnitude initially.

5.2.3 Jitter (PPQ)

Graphs showing PPQ on the vowel following guttural consonant types for all four subjects are provided in Figure 16. Regression lines indicate the degree that change in F0 over the first half of the vowel can predict the amount of jitter, measured via PPQ, over the same period. As can be seen visually, the change in F0 can account for quite a bit of the variation, but consonant phonation type does not seem to be controlling the amount of PPQ found. In addition, there is a level of variability not captured here, since the data is somewhat loosely spread around the regression line. These results show that even the PPQ measure of jitter, which involves smoothing over 5 cycles, is still too dependent on changes in F0, and thus probably is not a good measure of linguistic phonation type contrasts. We have seen that even in level toned roots, the F0 changes associated with roots containing initial guttural consonant type contrasts are too great to allow the measure to be useful. A smaller window that started after the rise in F0 associated with guttural consonant types would be too small to attain useful results, and might also possibly be too late to catch any jitter present due to C-V coarticulation. That is, jitter, which has proven to be useful in measuring differences in sustained phonation type in the clinical literature, is not very useful for capturing more transient jitter due to guttural coarticulation, particularly when these phonation types are also associated with changes in F0.

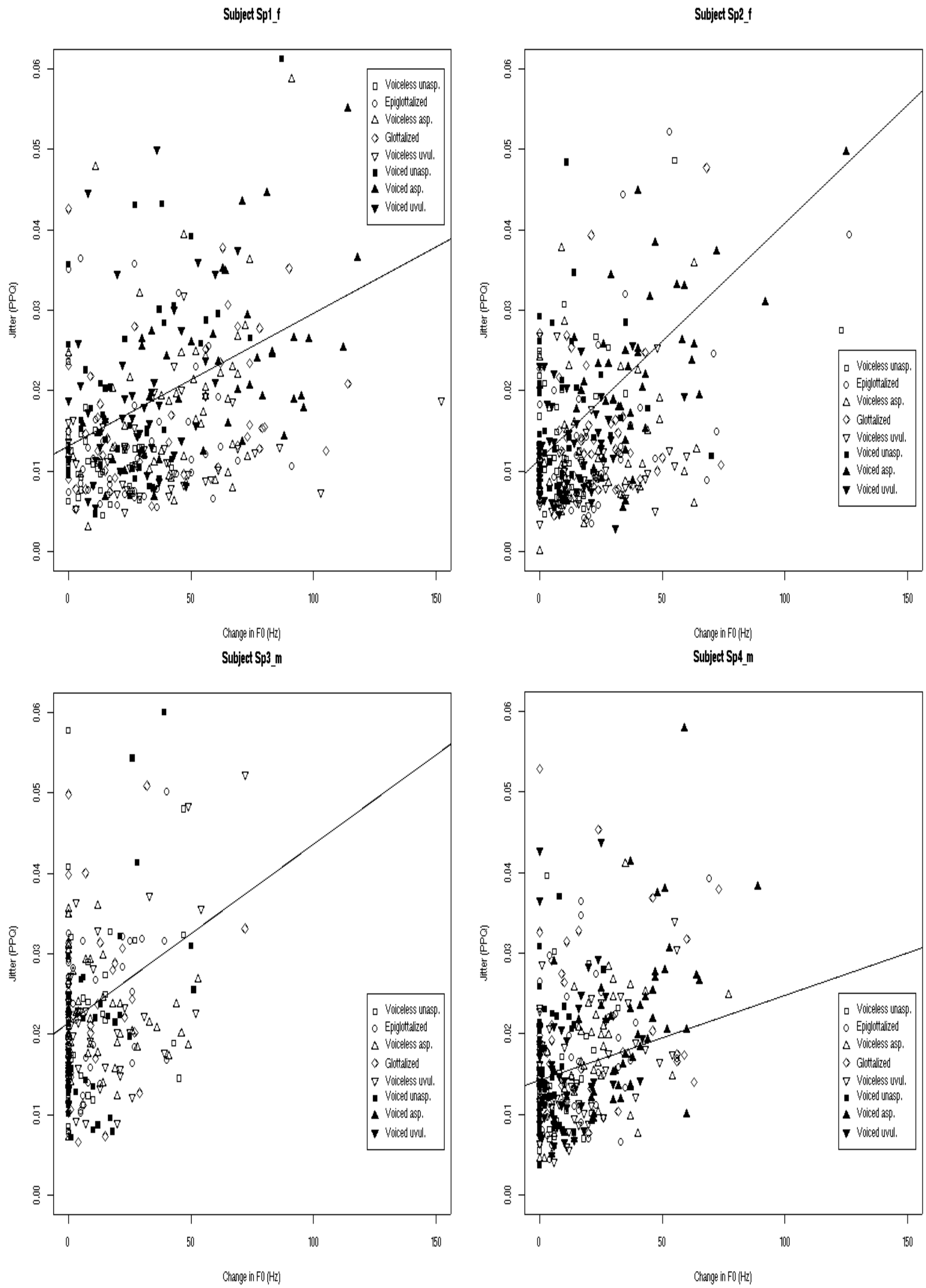


Figure 16: Change in F0 against PPQ associated with guttural consonant types over first half of vowel

The same two regressions reported above for the vocalic data were calculated for the consonant data in order to determine whether change in F0 over the vowel and the consonant phonation types predict the variance in PPQ over the first half of the following vowel. For these calculations, voiceless and voiced aspirates are grouped together as “aspirated” consonants, and voiceless and voiced uvularized consonants are grouped together as “uvularized” consonants. The results are reported in Table XI. The first regression showed a substantial contribution of change in F0 to PPQ for all subjects. The difference between the contribution of change in F0 to PPQ for different subjects for guttural vowels and vowels following guttural consonants is not consistent. For subject Sp1_f, the prediction of PPQ from change in F0 shows an increase of about 5% in predicting the variability in vowels over consonants. However, for the other three subjects, there is about a 5% decrease in predictability of PPQ from change in F0 for consonantal phonation type when compared with vocalic phonation type.

For all subjects, consonantal phonation type is also a significant contributor to PPQ over the first half of the vowel, but adding consonantal phonation type as a predictor only significantly increases prediction of PPQ for subjects Sp2_f and Sp4_m. Different consonantal phonation types also contribute to the prediction. For subject Sp2_f, uvularized and epiglottalized clicks aid in determining the amount of PPQ present, while aspiration and glottalization increase prediction for subjects Sp3_m and Sp4_m.

CONSONANTAL PHONATION TYPES Pr > /t/				
Subject	Sp1_f	Sp2_f	Sp3_m	Sp4_m
Change in F0	<.001	<.001	<.001	<.001
R^2 (model A)	0.10	0.08	0.10	0.09
Change in F0	<.001	<.001	<.001	<.001
Aspirated			<.05	
Uvularized		<.01		
Glottalized				<.001
Epiglottalized		<.05		
R^2 (model B)	0.11	0.11	0.12	0.14
Pr (> F)		<.001		<.001

Table XI. Regression results for PPQ associated with consonantal phonation type over the first half of the vowel

It is not surprising that consonantal phonation type is not a good predictor of the change in F0 over the first half of the following vowel, as shown by the regression results reported in Table XII. It is likely that consonantal phonation type would be a good predictor of the shape of the F0 curve right after the consonant, but that these differences are more localized than the total change in F0 over the first half of the vowel is able to capture. That is, the local differences may be great, but they go unnoticed when the change in F0 is calculated over the first half of the vowel, due to the fact that the same target peak F0 associated with the lexical tone is reached for roots with all consonantal phonation types in Ju|'hoansi.

Subject	**Sp1_f**	**Sp2_f**	**Sp3_m**	**Sp4_m**
Aspiration	<.001	<.001	<.001	<.001
Uvularization	<.01	<.05		<.001
Epiglottalization	<.05	<.001		<.001
Glottalization	<.001	<.001		<.001
R^2 (Model C)	0.20	0.17	0.14	0.26

Table XII. Change in F0 as a predictor of consonantal phonation type

A fourth regression was calculated to look at the independent relationship between PPQ and consonantal phonation type. These results are reported in Table XIII.

Subject	**Sp1_f**	**Sp2_f**	**Sp3_m**	**Sp4_m**
Aspiration	<.01			<.01
Uvularization		<.05		
Epiglottalization				
Glottalization				<.001
R^2 (Model D)	.04	.05	.08	.08

Table XIII. PPQ as a predictor of consonantal phonation type

Given that the change in F0 is itself correlated with the consonantal phonation type, I also calculated a partial regression to ascertain whether consonantal phonation type predicts a substantial proportion of the jitter once the mutual correlation with the F0 change is taken into account. I calculated this partial regression by correlating the residuals between model A and model D. The correlation between the residuals of the two models is substantial for all subjects, as shown by the results provided in Table XIV. Interestingly, the correlation is the largest for subjects Sp1_f and Sp3_m, the same subjects for whom consonantal phonation type does not contribute much to the prediction of PPQ (model D).

	Sp1_f	Sp2_f	Sp3_m	Sp4_m
Correlation	0.28	0.26	0.35	0.25

Table XIV. Correlation of residuals between model A (change in F0 ~consonantal phonation type) and model D (PPQ ~ consonantal phonation type)

5.3 Discussion

Results of experiment 2 show that aspirated consonants (both voiced and voiceless) and glottalized consonants in Ju|'hoansi all display laryngeal coarticulation. That is, vowels following these consonants all display low HNR values relative to vowels following unaspirated consonants, just as shown to exist in other languages. Similarly, vowels following both voiceless and voiced aspirated consonants display relatively high spectral slope values, and vowels following glottalized consonants display relatively low spectral slope values relative to vowels following both voiceless and voiced unaspirated consonants. Since the consonants in this study all involve click initial roots, these findings also show that aspirated and glottalized clicks display laryngeal coarticulation that is similar to laryngeal coarticulation found with pulmonic consonants.

In addition to extending laryngeal coarticulation to click consonants, this study also extends laryngeal coarticulation to include uvularized and epiglottalized consonants. That is, results of experiment 2 have shown that uvularized and epiglottalized consonants show similar spreading of voice quality cues to the following vowel that is found with aspirated and glottalized consonants. Vowels following uvularized click consonants, that are articulated with a spread glottal configuration, display high H1-H2 values relative to vowels following unaspirated click consonants. Vowels following epiglottalized click consonants display low H1-H2 values relative to vowels following unaspirated click consonants. Vowels following both uvularized and epiglottalized clicks exhibit low HNR values relative to vowels following unaspirated clicks. These results show that consonants articulated with a constriction in the pharynx spread similar voice quality cues to the following vowel as laryngeally articulated consonants do. The results are consistent with terming the process guttural coarticulation, following the use of the term guttural in Semitic and Salishan language phonologies to encompass laryngeally and pharyngeally articulated consonants.

Jessen and Roux (2002) show that vowels following voiced clicks in Xhosa have higher H1-H2 values than vowels following voiceless clicks for half (four out of eight) of their subjects. It is interesting to note that only one out of four Ju|'hoansi subjects included in the current study (female subject Sp1_f) shows a somewhat greater difference in the H1-H2 values for vowels following voiceless and voiced unaspirated clicks. Voiced unaspirated initial roots also display slightly lower HNR values relative to vowels following voiceless unaspirated clicks, which also remain stable throughout the root. This stability differs markedly from the large rise in HNR associated with aspirated, glottalized, uvularized and epiglottalized click consonants that pattern together in the OCP constraint and tonal depression. The smaller differences in spectral slope and HNR found in Ju|'hoansi voiced and voiceless unaspirated clicks is likely tied to the fact that Ju|'hoansi, like other Khoesan languages, contrasts voiced unaspirated and voiced aspirated clicks, while Xhosa only contains one type of voiced click. The lack of voiced aspirated consonants in Xhosa makes breathiness available as a cue to voicing in that language. Additionally, Ju|'hoansi voiced clicks do consistently exhibit closure voicing that is lacking in Xhosa voiced clicks. The consistent availability of closure voicing as a cue to voicing in clicks in Ju|'hoansi probably lessens the import of additional cues such as high spectral slope to mark the contrast between voiced and voiceless unaspirated clicks.

The results of experiment 2 also show that consonantal phonation types have a low degree of PPQ associated with them in Ju|'hoansi. Thus, while one of the acoustic effects of guttural coarticulation is variation in the duration of pitch periods, the jitter associated with coarticulation is beyond the more controlled changes in F0 that are themselves associated with consonantal phonation types. However, given the small amount of jitter present, this can only be a small part of the irregularity picked up by the cepstral HNR measure.

6. Experiment 3: Three-way phonation type contrasts in a single syllable

6.1 Introduction

In this section, I report the results and discussion for experiment 3, designed to explicitly examine the differences between partially breathy and fully breathy vowels, and to compare the timing of aspirated consonants with the two types of breathy vowels. The first hypothesis tested is that the two types of breathy vowels found in the language are differentiated solely by the timing of voice quality cues associated with them, and that the partially breathy vowels can be thought of as diphthongs in voice quality. A second hypothesis tested is that all three types of syllables: (1) monosyllabic roots containing an initial aspirated consonant followed by a modal vowel, (2) monosyllabic roots containing unaspirated consonants followed by partially breathy vowels, and (3) monosyllabic roots containing unaspirated consonants followed by fully breathy vowels, all contain marked voice quality cues in the C-V transition. A third hypothesis tested is that the only cue that differentiates roots with aspirated consonants from those with breathy vowels is Voice Onset Time (VOT). In addition to breathiness, a similar three-way contrast in epiglottalization on Ju|'hoansi syllables is investigated. Results corroborate all three hypotheses.

6.2 Results

6.2.1 Spectral Slope (H1-H2)

Figure 17 provides spectral slope values for roots with both types of breathy vowels as well as roots with initial aspirated clicks, and unaspirated clicks for comparison. Comparing the spectral slope patterns of partially breathy and fully breathy vowels, we see that partially breathy vowels display very high H1-H2 values initially, but they fall throughout the root, while fully breathy vowels display stable high H1-H2 values throughout the root. These patterns are consistent with partially breathy vowels being diphthongs in voice quality. Comparing vowels following voiced aspirated clicks, with both types of breathy vowels, we see that the timing of H1-H2 values for partially breathy vowels and vowels following voiced aspirated consonants are highly similar. For subject Sp1_f, there is almost complete overlap between the spectral slope values throughout the vowel for both vowels following voiced and voiceless aspirated consonants, and partially breathy vowels. The other three subjects' productions exhibit comparable spectral slope values for breathy vowels and vowels following aspirated consonants right at the beginning of the vowel, but with the roots containing consonants and vowels being differentiated toward the end of the vowels. As expected, the aspirated consonants and partially breathy vowels have lower spectral slope values at the end of the vowel, while fully breathy vowels have greater spectral slope values throughout the entire vowel.

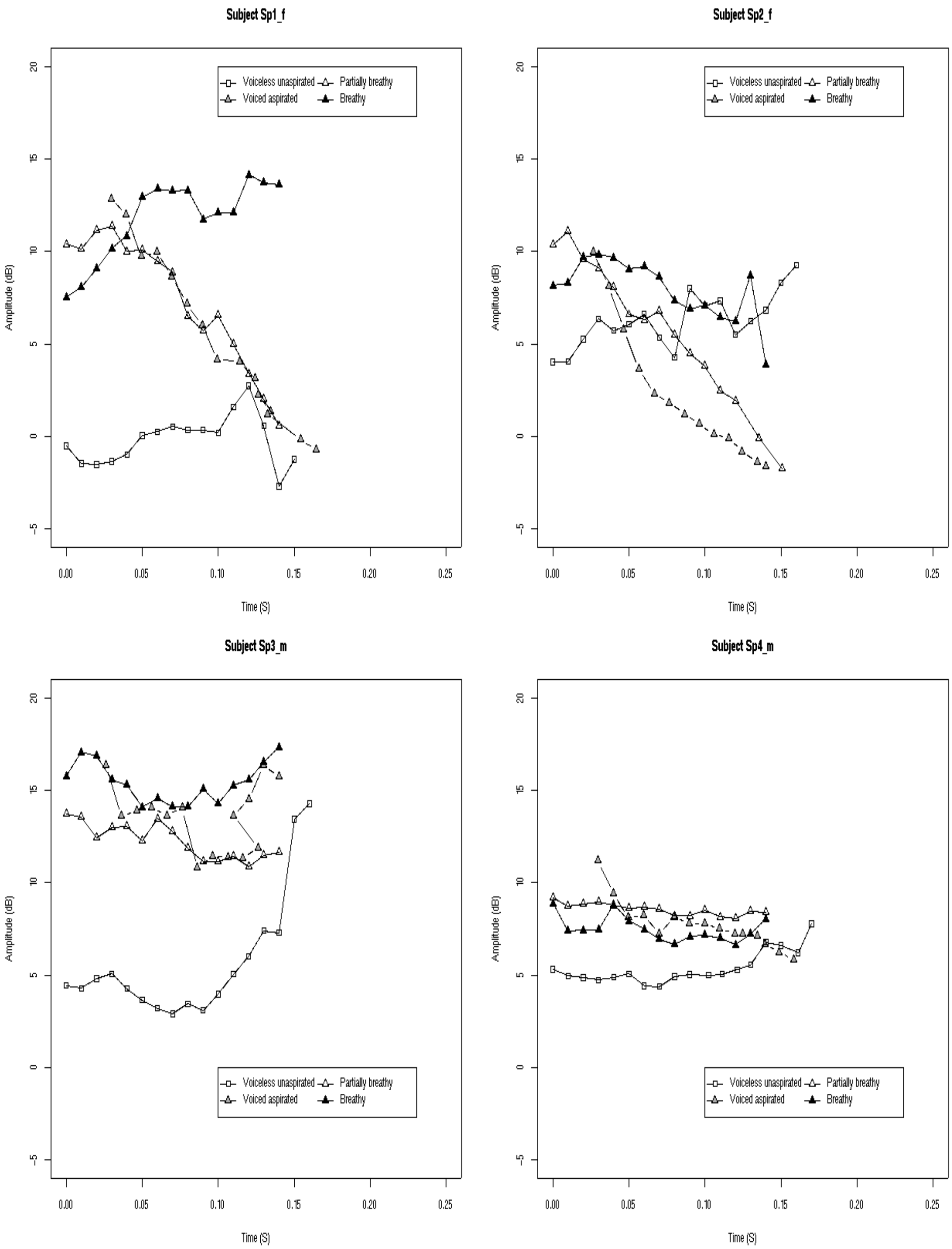


Figure 17: Median spectral slope values associated with the four-way aspiration contrast in the [a] context

Figure 18 displays H1-H2 values for roots with both types of epiglottalized vowels, epiglottalized initial clicks, and roots with initial voiceless unaspirated consonants for comparison. As can be seen, the fully epiglottalized vowels display low H1-H2 values that are stable throughout the vowel, while the partially epiglottalized vowels have rising H1-H2 values throughout the vowel, consistent with calling them diphthongs. Comparison of uvularized consonants and epiglottalized vowels for subject Sp1_f shows that vowels following voiceless uvularized clicks are distinct from vowels following epiglottalized clicks and roots with both types of epiglottalized vowels. However, considerable overlap occurs between vowels following epiglottalized consonants, and both types of epiglottalized vowels. For subject Sp2_f, vowels following uvularized consonants, epiglottalized consonants and epiglottalized vowels all display lower H1-H2 values than are found following voiceless unaspirated consonants. Subject Sp3_m shows the most overlap between epiglottalized click initial roots and roots containing epiglottalized vowels among the four subjects. Again, the guttural OCP constraint ensures that in all cases, roots with epiglottalized initial consonants and roots with partially epiglottalized vowels are differentiated by VOT, as they are in figure 18.

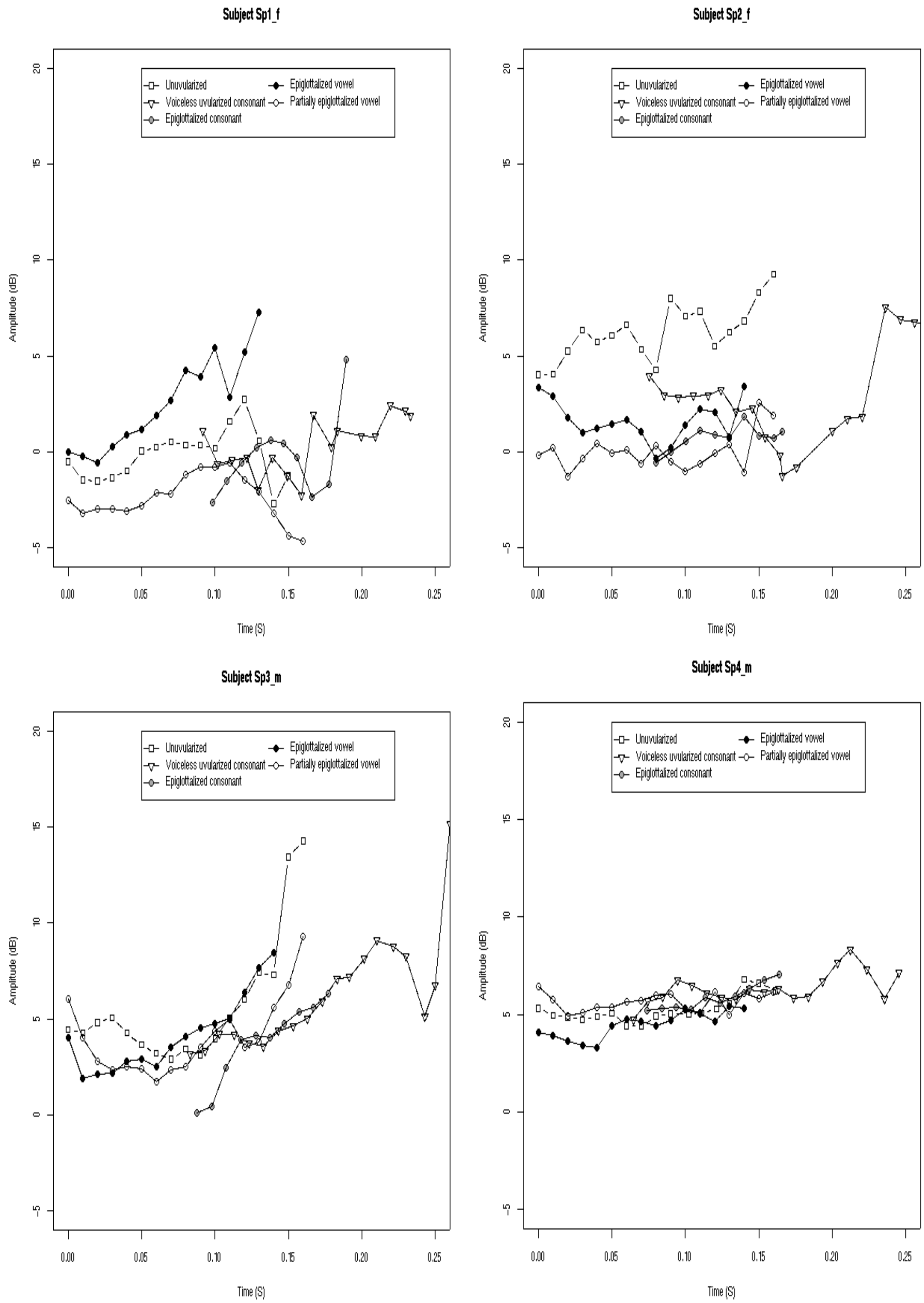


Figure 18: Median spectral slope values associated with the four-way epiglottalization contrast in the [a] context

Figure 19 provides spectral slope values for the entire duration of the vowel for glottalized click initial roots and roots containing glottalized vowels. As can be seen, both roots containing initial glottalized clicks and roots containing glottalized vowels, have low H1-H2 values throughout their duration relative to modal vowels in roots with initial voiceless unaspirated consonants. For subjects Sp1_f and Sp3_m roots containing initial glottalized clicks display lower H1-H2 values throughout their duration than roots containing glottalized vowels. On the other hand, subject Sp2_f's productions display lower H1-H2 values associated with glottalized vowels than in roots containing initial glottalized consonants. Subject Sp4_m does not differentiate roots containing initial glottalized clicks and roots containing glottalized vowels from roots with unaspirated consonants and modal vowels.

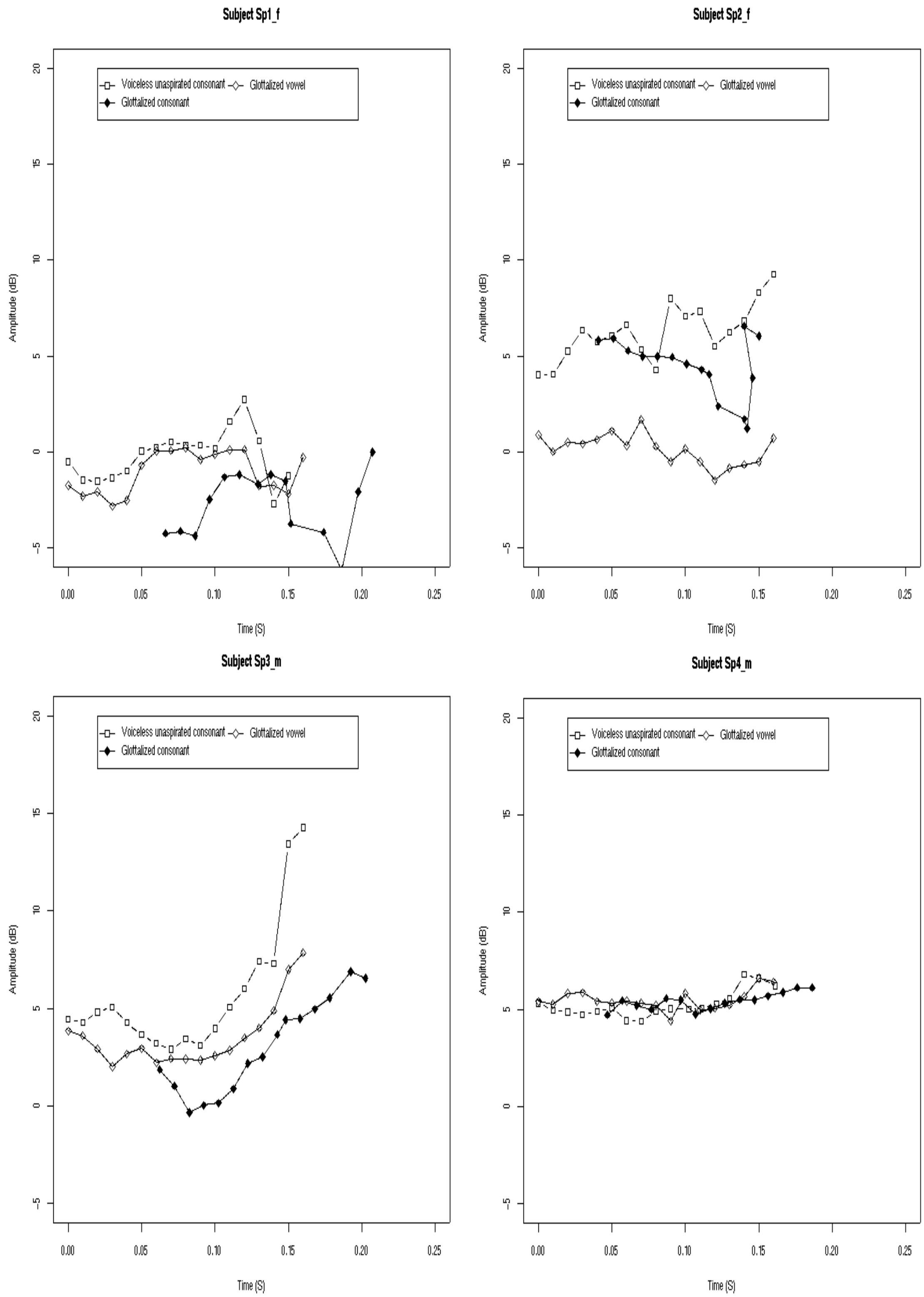


Figure 19: Spectral Slope values associated with the two-way glottalization contrast in the [a] context

6.2.2. HNR (R1)

HNR measured via R1 shows parallelism between guttural consonants and vowels that is striking. R1 associated with the four-way aspiration contrast is shown in Figure 20 for all four subjects. Female subject Sp1_f shows the greatest similarities, with the R1 values associated with both voiceless and voiced aspirated click initial roots and roots containing partially breathy vowels being almost exactly parallel over the entire duration of the vowel. The partially breathy vowels are very similar to the fully breathy vowels throughout the first half of the vowel, but rise in gamnitude to the level found for modal vowels by the end of the roots. In fact, the HNR values associated with partially breathy vowels overshoot the HNR levels found in modal vowels in roots with unaspirated initial clicks at the very end of the root (as can be seen by comparing figures 20 and10). The HNR results are thus also consistent with calling partially breathy vowels diphthongs in voice quality. The one clear acoustic difference between roots containing initial aspirated clicks and roots containing partially breathy vowels is VOT, seen by the time at which the first HNR values are displayed, which also differs for voiced aspirated and voiceless aspirated click initial roots. The two female subjects, Sp1_f and Sp2_f, show the greatest separation between modal and breathy vowels.

Subject Sp3_m's productions do not show as much separation as the other subjects' productions between modal vowels and breathy vowels or vowels following aspirated consonants, but the overall patterns are very similar. Subjects Sp2_f, Sp3_m and Sp4_m show a smaller difference in R1 between partially and fully breathy vowels than that found in subject Sp1_f's speech, which results in a slightly better separation between the noise levels associated with aspirated click initial roots and roots containing breathy vowels. Subject Sp3_m's speech also displays strong similarities between all of the types, although it is difficult to see with the small range he employs on this measure, due to his overall rough voice quality.

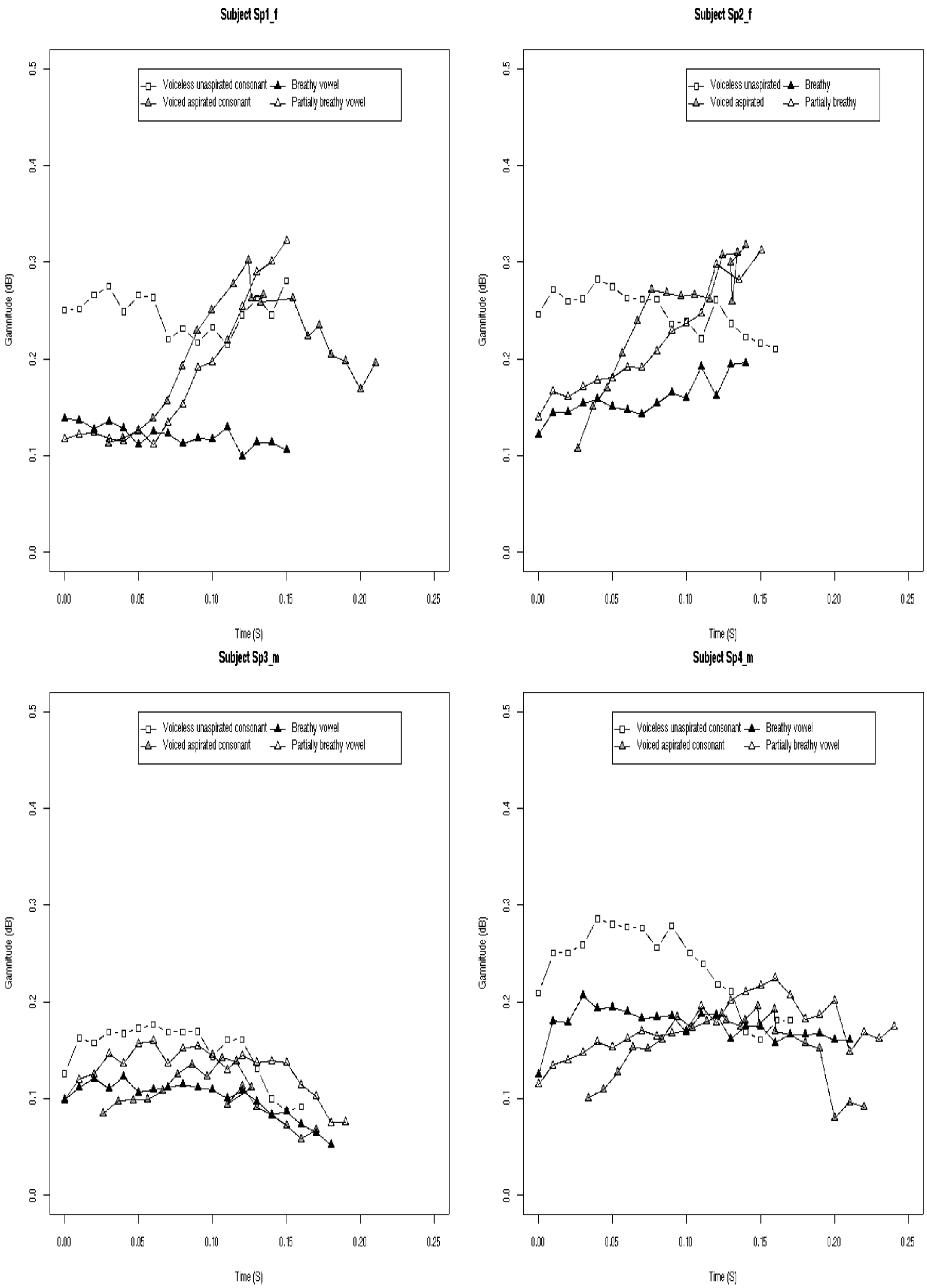


Figure 20: Median gamnitude of the first rahmonic peak (R1) in the cepstrum for the Ju|'hoansi four-way aspiration contrast in the [a] context

The four-way contrast in epiglottalization on the R1 dimension is provided for all subjects in Figure 21. For all four subjects, R1 associated with voiceless uvularized click initial roots and roots containing partially epiglottalized vowels are parallel, although the similarities are strongest for the two male subjects. The parallelism in R1 between voiceless uvularized click initial roots and roots containing partially epiglottalized vowels is striking for subject Sp1_f's productions. Subject Sp2_f and Sp4_m's patterns also show striking similarities. The pattern for partially epiglottalized vowels is similar to that found for fully epiglottalized vowels over the first half of the root, but the partially epiglottalized vowels overshoot the R1 values found in modal vowels by the end of the root. The timing of HNR values associated with partially epiglottalized vowels is consistent with them being diphthongs in voice quality. The really long VOT's associated with uvularized click initial roots are evident here.

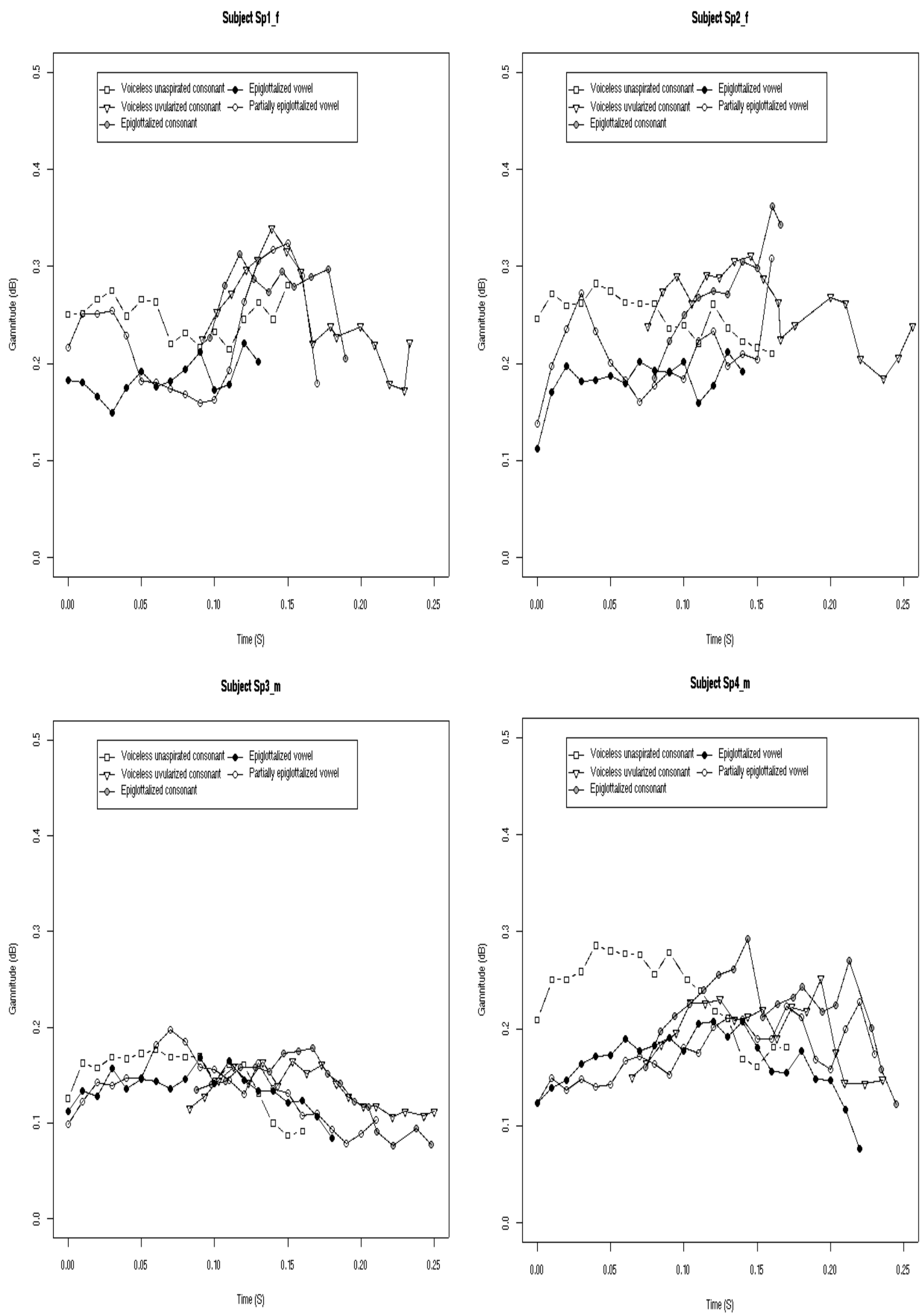


Figure 21: Median gamnitude of the first rahmonic peak (R1) in the cepstrum for the Ju|’hoansi four-way epiglottalization contrast in the [a] context

One last parallelism between vowels following guttural clicks and guttural vowels is the resemblance found between HNR of vowels following glottalized clicks and roots containing glottalized vowels. The R1 values associated with these contrasts for all four subjects are provided in Figure 22. For this two-way contrast, subjects Sp2_f, Sp3_m and Sp4_m show the highest degree of similarity on the HNR dimension, while subject Sp1_f shows the greatest separation. The HNR patterns associated with glottalized vowels show that these vowels are more similar to partially breathy and partially epiglottalized vowels then they are to fully breathy and fully epiglottalized vowels, and are thus actually phonetically diphthongs.

As noted in section 3.1.2, the HNR measure used here, the gamnitude of the first rahmonic peak in the cepstrum, is applicable across different vowel contexts. In figure 23 below, I provide the graphs of R1 over the entire vowel for the four-way aspiration contrast in the [o] context for all four subjects. As can be seen, the results are highly similar to the results shown in figure 20 above which shows the same contrasts in the [a] context. The main substantive difference that can be noted is that the overall gamnitude associated with the vowel following voiceless unaspirated consonants are higher in Sp1_f's productions in the [a] context than in the [o] context. Since the results displayed are medians, it could be that this subject produced a single loud production of the [o] words in the data. The high degree of similarity in the [a] and [o] contexts confirms our prediction that the cepstral measure of HNR used here is not at all dependent on first formant frequency differences. Of course, any measure that assesses amplitude, or a derivative thereof such as gamnitude, will be subject to differences in overall loudness of recorded tokens. The method used here of recording an entire wordlist in full before going on to the next repetition is the best way to avoid systematic amplitude differences found across different contexts. It will always be impossible to avoid sudden loudness changes in a particular token.

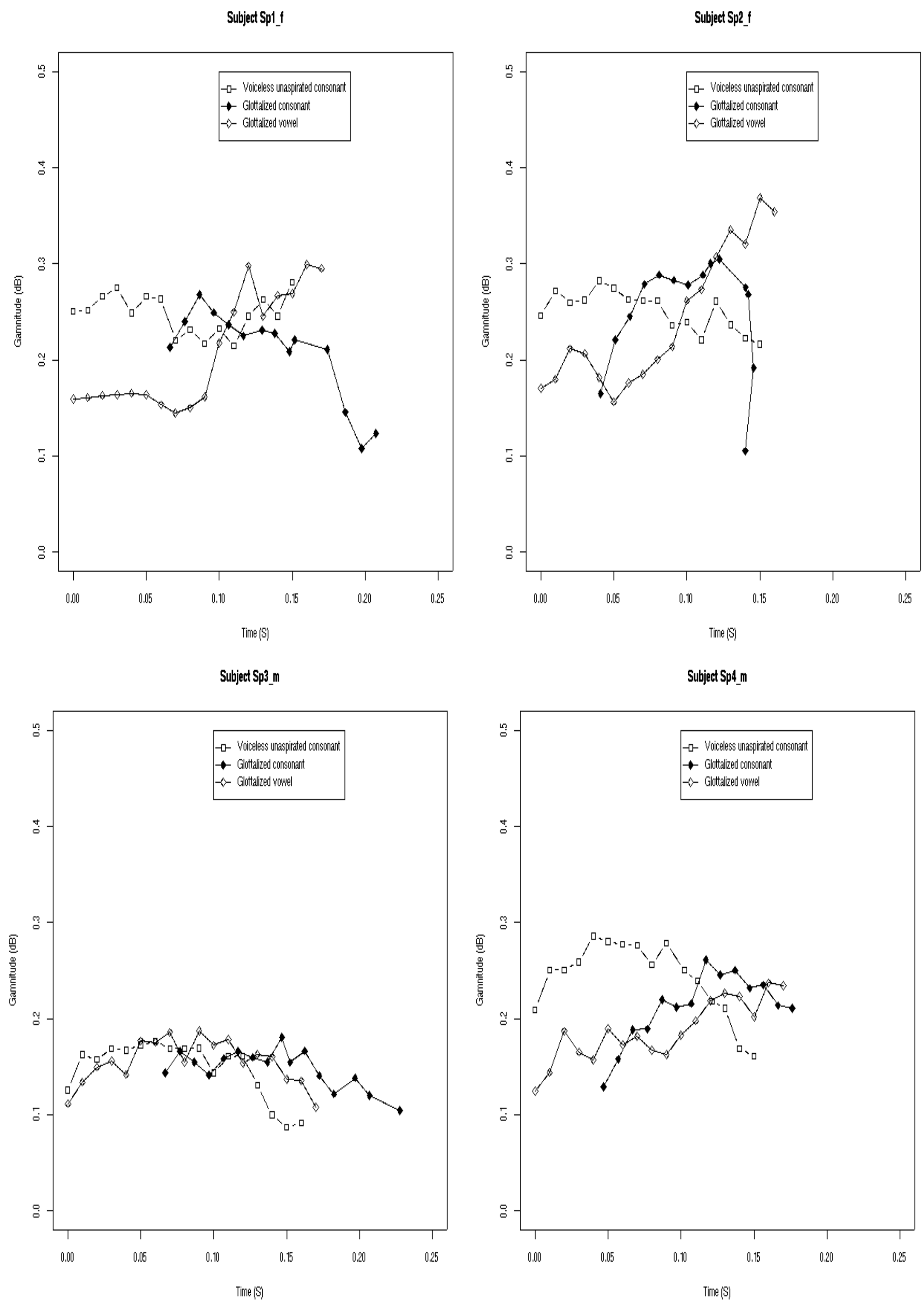


Figure 22: Median gamnitude of the first rahmonic peak (R1) in the cepstrum for the Ju|'hoansi two-way glottalization contrast in the [a] context

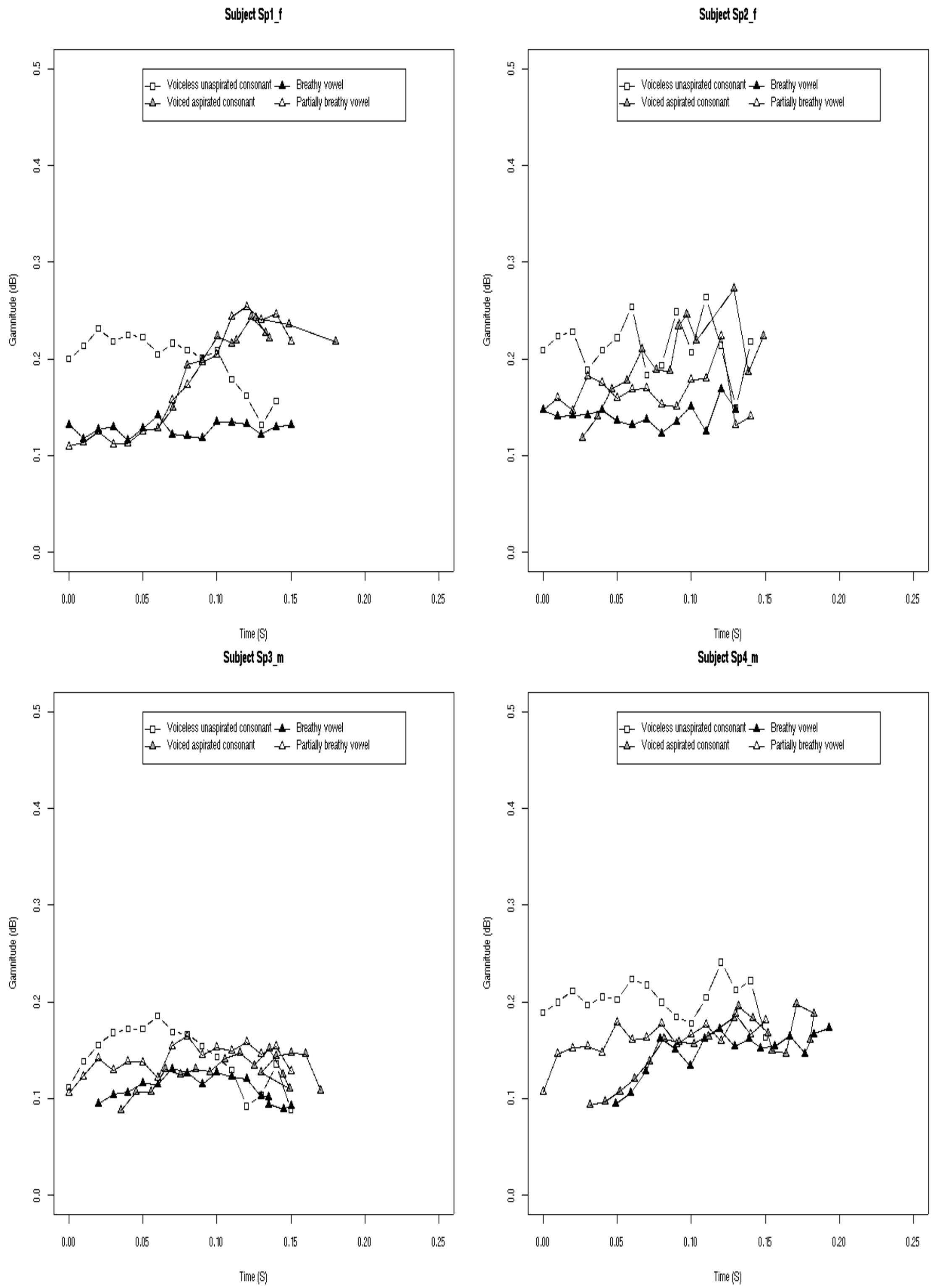


Figure 23: Median gamnitude of the first rahmonic peak (R1) in the cepstrum for the Ju|'hoansi four-way aspiration contrast in the [o] context

Voice Onset time differences among click types can also be noted by comparing the beginning points of the vowels in figures 20 and 22. As was noted earlier, unvoiced portions of the signal were gated out before inverse filtering of the signal, resulting in the first point showing up in the data reflecting the voice onset time of the consonant. Given that it was impossible to control for click type across all contexts in this study, the aspirated click initial root in the [a] context contains the [ǂʰ] click, while the aspirated click initial root in the [o] context contains the [!ʰ] click. As can be seen, the central post-alveolar [!] click type has a slightly longer VOT than the palatal [ǂ] click type. This is the most evident for subject sp1_m's productions. There is a consistent difference found for all of the data collected in this study between these two contexts.

Figure 24 shows the four-way epiglottalization contrast in the [o] context for all four subjects. Again, the results are highly similar to those shown above in figure 21 in the [a] context. The only major difference that can be seen in comparing figures 21 and 24 is that the separation between the HNR of vowels following unaspirated consonants and the roots containing epiglottalization is less clear in the [o] context. However, the vowels following unaspirated consonants still have a higher HNR, as measured by R1, over the first part of the vowel in the [o] context. There are no large gamnitude differences found between the [o] context and the [a] context with the epiglottalized sounds.

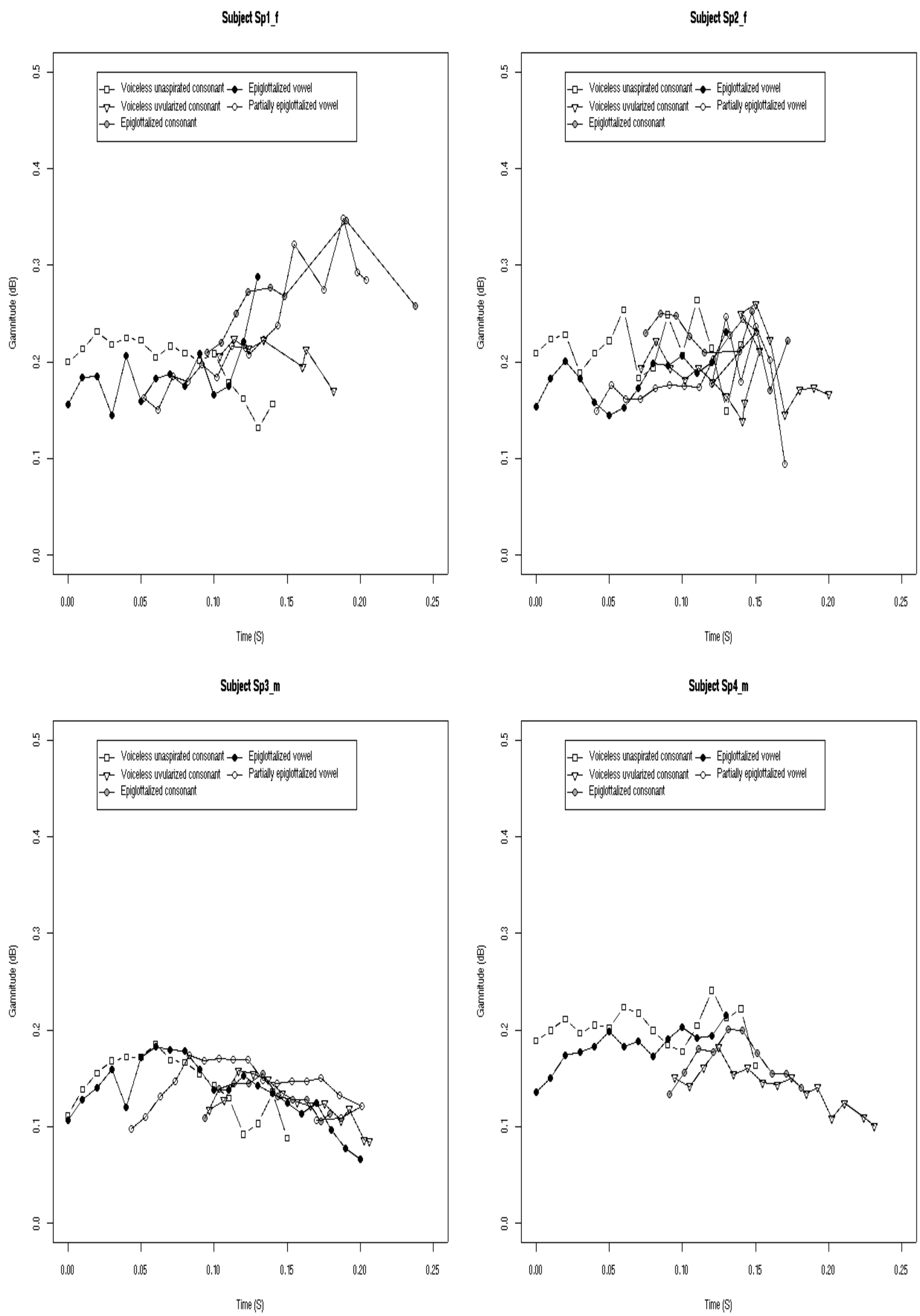


Figure 24: Median gamnitude of the first rahmonic peak (R1) in the cepstrum for the Ju|'hoansi four-way epiglottalization contrast in the [o] context

6.3 Discussion

Results of experiment three have provided explicit support for the contrast between partially breathy and fully breathy vowels, and partially and fully epiglottalized vowels. The evidence is consistent with the analysis provided in figure 1 in section 2.3. Evidence also shows that glottalized vowels are more similar to partially breathy and partially epiglottalized vowels in timing of voice quality cues. However, there is no phonological contrast in the timing of glottalization on vowels. This is then a case of a contrast that doesn't obey Silverman's (1997) principle that inventories should have maximally contrastive phasing of cues before more similar ones are allowed to occur. Again, it shows that VOT allows sufficient discriminibility between the paradigmatic contrast of roots containing initial glottalized clicks and roots containing a plain unaspirated click followed by a glottalized vowel.

Results of experiment three also show that partially breathy vowels are very similar in the timing of voice quality cues (high spectral slope values and low HNR values) to roots containing initial aspirated consonants. Similarly, HNR and spectral slope values associated with partially epiglottalized vowels and glottalized vowels are very similar in timing to roots containing initial epiglottalized consonants and initial glottalized consonants. Given that HNR values are similar across all guttural vowel types, and in vowels following all guttural consonant types through guttural coarticulation, HNR likely captures the level of acoustic similarity that is the basis of the guttural OCP constraint described in section 2.4 above. VOT values have been shown to differentiate roots containing guttural consonants from roots containing guttural vowels, given the presence of the OCP constraint which rules out the co-occurrence of a guttural consonant from a guttural vowel within the same root. If it were not for the OCP constraint, roots containing both an initial aspirated consonant and a partially breathy vowel would be free to occur in the language, and there would be no cues to differentiate such roots from roots containing either just a partially breathy vowel, or just an initial aspirated consonant.

VOT differences were also shown to occur with different click types. Since VOT differences are associated with pulmonic consonants associated with different places of articulation (Maddieson 1997), VOT differences found here point to differences in the place of articulation of the click just prior to the release of the posterior constriction, or to slight differences in the amount of overlap of the two constrictions in different click types that have been suggested to occur with the dental click by Traill (1997: 115) in order to account for the improved perception of [|] when appropriate formant transitions were available, and by Miller-Ockhuizen (2000) to account for the patterning of front clicks, [|] and [ǂ] disparately from the back clicks, [!] and [‖], with respect to the BVC. Further research will investigate these differences further, and compare them to VOT differences found with pulmonic consonants within a single language in order to determine what VOT tells us about place of articulation in click consonants.

7. Perceptual similarity of two phonological natural classes

The results reported here predict that there should be two levels of acoustic similarity among guttural consonants and vowels.[3] The spectral slope results suggest that uvularized and aspirated consonants should be perceptually similar, and glottalized and epiglottalized consonants and vowels should be perceptually similar, leading to one level of perceptual categorization. A second class predicted to emerge from perceptual similarity at the level of HNR includes all guttural consonants and vowels. These two levels of predicted perceptual similarity are shown in table XV.

	Unmarked Spectral Slope	Marked Low Spectral Slope	Marked High Spectral Slope
Low HNR		• Glottalization • Epiglottalization	• Breathiness • Uvularization
High HNR	• Modal voicing		

Table XV. Perceptually determined natural classes predicted by Spectral Slope and HNR patterns

[3] It is assumed herein that phonology interfaces with perception, not the acoustics directly as claimed by Hume and Johnson 2001.

In fact, patterns targeting each natural class are found both within Khoesan and in other language families. There are patterns showing both the perceptual similarity of aspirated and uvularized consonants that both have high spectral slope values, as well as the perceptual similarity of glottalized and epiglottalized sounds that both have low spectral slope values. For example, uvularized and aspirated consonants found in Khoekhoe have been merged from proto-Central Khoesan, which contains both uvularized and aspirated clicks, providing evidence for the perceptual similarity of aspiration and uvularization. In modern-day Khoekhoe, [!^{χ}] and [!^{h}] (using the [!] click type to symbolize all click types) are in free variation. Perceptually motivated diachronic changes from *k → x → χ → h in syllable initial position in German also provide evidence for the perceptual similarity of [χ] and [h]. Evidence for the similarity of glottalized and epiglottalized clicks, which both have low spectral slopes associated with them, comes from the diachronic change of *|x' > |' in Central Khoesan languages (Vossen 1997: 283-284). Spectral slope values are also important in differentiating different types of guttural consonants and vowels, which can not be distinguished on the basis of HNR.

It is important to recognize here that the spectral slope values are not simply a translation of the articulatory features [spread glottis] vs. [constricted glottis] into acoustic terms. There are articulatory to acoustic non-linearities which make the similarity in terms of the acoustic property more appropriate. For example, epiglottalized vowels in Ju|'hoansi are articulated with a constriction between the raised larynx, constricted false vocal folds and pharyngeal narrowing, not via a constriction at the glottis as with other types of glottalized consonants.

Guttural consonants and vowels are also all acoustically similar in that they display low gamnitude of the first rahmonic in the cepstrum, which is a measure of harmonics-to-noise ratio. That is, they display either random variation in the duration of the pitch period from cycle to cycle, or they contain noise in the low frequency range of the spectrum. Put simply, they are noisy. This level of perceptual similarity is targeted by the Guttural OCP constraint in Ju|'hoansi, which rules out the co-occurrence of all guttural consonants and vowels within the same root.

Cross-linguistic evidence for this type of similarity is also provided by laryngeal co-occurrence constraints found in Cuzco Quechua and Peruvian Aymara, where aspirated consonants and ejectives are banned from co-occurring within the same root (MacEachern 1999). While aspiration and glottalization involved in ejectives have opposite spectral slope values, both types of consonants are predicted to lead to low HNR of the following vowel, although this has not been explicitly tested for these languages. MacEachern (1999) proposes that VOT is the relevant acoustic attribute of ejected and aspirated consonants in these languages over which similarity is defined, and her hypothesis does indeed make the correct predictions. However, it is likely that HNR is also a factor in Aymara and Quechua phonotactic patterns. VOT can not be applied to the similarity involved in Ju|'hoansi gutturals, since it only applies to consonants, and the natural class of gutturals in Ju|'hoansi involves both consonants and vowels. By stating the Aymara and Quechua patterns in terms of HNR values, we would be able to capture both language patterns with a single acoustic attribute.

I have shown that there are two classes of acoustic similarity present in Ju|'hoansi gutturals, which are assumed to be relevant to the phonology at the level of perception, following Hume and Johnson (2001). The class of sounds having low HNR is the entire class of guttural consonants and vowels, as long as C-V coarticulation involving guttural consonants, which results in low HNR of the following vowel is considered. The natural class of guttural consonants and vowels is targeted by a guttural OCP constraint active in the language. I have also claimed that this level of perceptual similarity may also be a good predictor of some phonotactic sound patterns in other languages. The second natural class is determined by spectral slope values. Glottalized and epiglottalized consonants and vowels all have low spectral slope values associated with them, either on the vowel itself, or on the following vowel through C-V coarticulation found involving glottalized and epiglottalized consonants. I have motivated this natural class on the basis of several diachronic sound patterns found in Khoesan languages. Additionally, this acoustic attribute is important in maintaining the contrast between the different types of guttural consonants and vowels that can not be distinguished on the basis of HNR.

8. Conclusion

The results of the H1-H2 and HNR analyses presented in this paper show that guttural vowels all exhibit increased spectral noise below 1000 Hz. This is the first acoustic investigation of voice quality cues on epiglottalized vowels ever undertaken. The spectral slope (H1-H2) measure shows that epiglottalized vowels and glottalized vowels are acoustically opposite to breathy vowels. It is important to note that the articulation of epiglottalized vowels in Khoesan languages likely involves a constriction produced by the laryngeal sphincter (Traill 1986; Miller-Ockhuizen In progress), as seen in Arabic (Laufer and Condax 1981; El Halees 1986) and Esling's study of his own articulation of pharyngeal consonants (Esling 1999). Thus, the low H1-H2 values found in these vowels do not likely indicate a constriction at the glottis, but rather a constriction a bit higher up produced in a somewhat more complex manner, that has similar acoustic results. This thus provides evidence for the basis of the Guttural OCP constraint in Ju|'hoansi being the perceptually based class of consonants and vowels that have low HNR.

Cue trading between HNR and spectral slope measures suggests that these cues both play a part in offering the percept of guttural, although the cue weighting of these cues is left for future research. VOT is shown to differentiate roots containing initial guttural clicks from roots containing guttural vowels that have similar voice quality cues occurring in the same temporal domain (the C-V transition). Additionally, VOT differences have been shown to exist for different click types, suggesting differences in place of articulation of the constriction that is held just prior to release.

Guttural consonants have also all been shown to exhibit guttural coarticulation, resulting in decreased HNR. This provides further evidence for laryngeal coarticulation as shown to exist by Löfqvist and McGowan (1992), NíChasaide and Gobl (1997) and Blankenship (2002). Of interest is the extreme amount of coarticulation following aspirated and uvularized click consonants compared with pulmonic consonants in other languages. It is unclear whether this difference can be attributed to differences in velaric and pulmonic airstream mechanisms, or whether the difference should simply be attributed to linguistic differences in the degree of laryngeal coarticulation found in the various languages. The amount of coarticulation found with glottalized consonants is remarkably similar across languages, suggesting that the differences found with aspirates may be a linguistic difference.

The results also provide novel evidence that uvularized and epiglottalized consonants display coarticulation through voice quality cues similar to that found with laryngeal consonants. This is thus one area where the natural class of guttural sounds in Ju|'hoansi displays acoustic similarity. The epiglottalized consonants are also expected to show constrictions higher in the pharynx rather than at the glottis, although no investigations of these sounds produced by native subjects have been undertaken to date[4]. Given the fact that any constriction at or near the glottis results in decreased R1 and decreased H1-H2, the acoustic similarity of guttural consonants and vowels should be referenced in grouping these sounds, rather than articulatory properties such as spread or constricted glottal postures.

Given the temporal similarity of cues associated with guttural consonants and guttural vowels, VOT is the only cue that appears to reliably distinguish roots containing guttural consonants from roots containing guttural vowels. I have suggested that it is the presence of a Guttural OCP constraint in the language that allows this cue to be reliable, since its enforcement results in no roots that contain both guttural consonants and vowels. The results also show that one cue that differentiates uvularized consonants and aspirated consonants, which have such a similar effect on the spectral properties of the following vowel, is voice onset time, which is much longer for uvularized clicks than for aspirated clicks.

The availability of VOT to distinguish between paradigmatic contrasts associated with consonants and vowels is assured by the Guttural OCP constraint found in the language. The suggestion that VOT, a consonantal cue, is used to differentiate consonants from vowels is a novel idea, and is problematic for phonetic and phonological theories which assume that consonants and vowels are distinguished by different acoustic cues and different features. I have proposed that sufficient discriminability of sounds within an

[4] Investigation of Miller-Ockhuizen's productions of these sounds with high-speed fiberscopic technology showed severe tongue root retraction and false vocal fold constriction, but no glottal constriction. Ultrasound investigation of tongue root posture in clicks is currently underway (Miller-Ockhuizen, Namaseb, del Teso Craviotto and Iskarous In progress).

inventory can not be fully understood by looking at consonant and vowel inventories separately, because phonotactic constraints help shape the possible sequences of consonants and vowels that can occur within the language. Discriminability is only appropriately determined by looking at the possible strings that are allowed to exist. In this case, the OCP constraint helps rule out sequences that are not sufficiently discriminable. The presence of VOT cues that allow listeners to distinguish between roots containing consonants and vowels with parallel phonation type contrasts, allows phonetically conditioned coarticulation to be less gradient and confusable with phonologically conditioned phonemic contrasts in the language. Thus, the usual distinction found between more gradient phonetics, and categorical phonology (Cohn 1993; Zsiga 1995) is less tightly observed in such cases.

H1-H2 and HNR results suggest two different natural classes that can be identified by acoustic similarity involving these two measures. I have shown that both of these classes are targeted by different synchronic and diachronic sound patterns in Ju|'hoansi, Khoesan and cross-linguistically.

References

Abramson, A. S. and Luangthongkum, T. (2001) Phonation types in Suai, *Journal of the Acoustical Society of America,* 110 (5).

Alwan, A. (1989) Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants, *Journal of the Acoustical Society of America*, 86(2), 549-556.

Bessell, N. (1992) *Towards a Phonetic and Phonological Typology of Post-Velar Articulation.* Unpublished Ph.D. Thesis. University of British Columbia.

Bessell, N. (1998) Phonetic aspects of retraction in Interior Salish, In *Salish Languages and Linguistics: Theoretical and Descriptive Perspectives* (Ewa Czaykowska-Higgins and M. Dale Kinkade, editors), pp. 125-152. New York: Mouton de Gruyter.

Bickley, C. (1982) Acoustic analysis and perception of breathy vowels, *Speech Communication Group Working Papers: Research Laboratory of Electronics*, MIT, 71-82.

Blankenship, B. (2002) The timing of nonmodal phonation in vowels, *Journal of Phonetics,* 30, 163-191.

Card, E. (1983) *A Phonetic and Phonological Study of Arabic Emphasis*. Unpublished Doctoral Dissertation, Cornell University.

Cho, T., Jun, S. and Ladefoged, P. (2000) An Acoustic and Aerodynamic Study of Consonants in Cheju, *Speech Sciences*, March 2000, 109-137.

Cohn, A. (1993) Nasalization in English: phonology or phonetics, *Phonology,* 10, 43-81.

Davis, S. B. (1976) *Computer Evaluation of Laryngeal Pathology Based on Inverse Filtering of Speech*, Speech Communications Research Laboratory Monograph No. 13. New York: National Institute of Health and the Voice Foundation.

Dickens, P. (1994) *English - Ju|'hoan Ju|'hoan - English Dictionary*, Quellen zur Khoisan-Forschung, 8, Köln: Rüdiger Köppe Verlag.

Elgendy, A. (2001) *Aspects of Pharyngeal Coarticulation*, Amsterdam: *LOT 44* Netherlands Graduate School of Linguistics.

El-Halees, Y. (1986) Does the epiglottis function as an articulator in the production of pharyngeal sounds?, *Logopedics and Phoniatrics: Issues for Future Research: Proceedings of the XXth Congress of the International Association of Logopedicsand Phoniatrics, 3rd-7th of August, 1986, Sasakawa Hall and Miyako Inn, Tokyo, Japan*, Tokyo: Bulletin of the Research Institute of Logopedics and Phoniatrics.

Esling, J. H. (1996) Pharyngeal consonants and the aryepiglottic sphincter, *Journal of the International Phonetic Association,* 26(2), 65-88.

Esling, J. H. (1999) The IPA Categories 'Pharyngeal' and 'Epiglottal', *Language and Speech,* 42(4), 349-372.

Fischer-Jorgensen, E. (1967) Phonetic analysis of breathy (murmured) vowels, *Indian Linguistics*, 28, 71-139.

Fujimura, O. and Lindqvist, J. (1971) Sweep-tone measurements of vocal-tract Characteristics, *Journal of the Acoustical Society of America,* 49, 541-558.

Gobl, C. and NíChasaide, A. (1999) Voice source variation in the vowel as a function of consonantal context, In *Coarticulation: Theory, Data and Techniques* (W. J.

Hardcastle and N. Hewlett, editors), pp. 122-143. Cambridge: Cambridge University Press.

Goldsmith, J. A. (1976) An overview of autosegmental phonology, *Linguistic Analysis*, 2, 23-68.

Goldstein, L. (1994) Possible articulatory bases for the class of guttural consonants, In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (Patricia A. Keating, editor), pp. 234-241. Cambridge: Cambridge University Press.

Halle, M. (1995) Feature Geometry and Feature Spreading, *Linguistic Inquiry,* 26 (1), 1-46.

Halle, M. and Stevens, K. (1971) A note on Laryngeal Features, *RLE Quarterly Progress Report*, 101, 198-312.

Hayward, K.M and Hayward, R.J. (1989) Guttural: Arguments for a New Distinctive Feature, *Transactions of the Philological Society,* 87 (2), 179-193.

Hillenbrand, J., Cleveland, R.A. and Erickson, R.L. (1994) Acoustic Characteristics of Breathy Voice Quality, *Journal of Speech and Hearing Research*, 37, 769-778.

Honda, K. (1995) Laryngeal and extra-laryngeal mechanisms of F0 control, In *Producing Speech: Contemporary Issues* (F. Bell-Berti & L.J. Raphael, editors), pp. 215-232. New York: American Institute of Physics.

Huffman, M. K. (1987) Measures of Phonation in Hmong, *Journal of the Acoustical Society of America*, 81, 495-504.

Hulst, van der H. and Mous, M. (1992) Transparent consonants, In *Linguistics in the Netherlands 1992* (R. Bok-Bennema and R. van Hout, editors), pp. 101-112. Philadelphia: Benjamins.

Hume, E. and Johnson, K. (2001) A Model of the Interplay of Speech perception and Phonology, In *The role of Speech Perception in Phonology* (E. Hume and K. Johnson, editors), pp. 3-26. San Diego: Academic Press.

Jackson, M., Ladefoged, P., Huffman, M. and Antonanzas-Barroso, N. (1985a) Measures of spectral tilt, *UCLA WPP,* 61, 72-125.

Jackson, M., Ladefoged, P., Huffman, M. and Antonanzas-Barroso, N. (1985b) Automated measures of spectral tilt, *UCLA WPP*, 62, 77-88.

Javkin, H., Antonanzas-Barroso, N. and Maddieson, I. (1987) Digital inverse filtering for linguistic research, *Journal of Speech and Hearing Research,* 30, 122-129.

Jessen, M. and Roux, J.C. (2002) Voice quality differences associated with stops and clicks in Xhosa, *Journal of Phonetics,* 30, 1-52.

Kirk, P. L., Ladefoged, J. and Ladefoged, P. (1993) Quantifying Acoustic Properties of Modal, Breathy and Creaky Vowels in Jalapa Mazatec, In *American Indian Linguistics and Ethnography in Honor of Laurence C. Thompson (University of Montana Occasional Papers in Linguistics No. 10*) (A. Mattina and T. Montler, editors), pp. 435-450.

Klatt, D. H. and 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.

Krom, de G. (1993) A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals, *Journal of Speech and Hearing Research,* 36,254-266.

Krom, de G. (1995) Some Spectral Correlates of Pathological Breathy and Rough Voice Quality for Different Types of Vowel Fragments, *Journal of Speech and Hearing Research,* 38, 749-811.

Ladefoged, P. (1973) The Features of the Larynx, *Journal of Phonetics,* **1**, 73-83.

Ladefoged, P. and Maddieson, I. (1996) *The Sounds of the World's Languages*. Cambridge: Blackwell Publishers.

Ladefoged, P., Maddieson, I. and Jackson, M. (1988) Investigation of Phonation Types in Different Languages, In *Vocal Physiology: Voice Production, Mechanisms and Functions* (O. Fujimura, editor), pp. 297-317. New York: Raven Press.

Ladefoged, P. and Traill, A. (1984) Phonetic details of click consonants, *Language,* 60 (1), 1-20.

Laufer, A. and Condax, I.D. (1981) The function of the epiglottis in speech, *Language and Speech,* 24, 39-61.

Leben, W.R. (1973) *Suprasegmental phonology*, Unpublished Ph.D. Dissertation in Linguistics, MIT.

Lee, S. (1995) Orals, gutturals and the jaw," In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (B. Connell and A. Aravaniti, editors), pp. 343-360. Cambridge: Cambridge University Press.

Liljencrants, J. and Lindblom, B. (1972) Numerical simulations of vowel quality systems: the role of perceptual contrast, *Language*, 48, 839-62.

Lindblom, B. (1990) Explaining phonetic variation: A Sketch of the H & H Theory, In *Speech Production and Speech Modelling* (W. J. Hardcastle and A. Marchal, editors), pp. 403-439.

Lindblom, B. and Maddieson, I. (1988) Phonetic universals in consonant systems, In *Language, speech and mind* (L.M. Hyman and C.N. Li, editors), London and New York: Routledge.

Löfqvist, A. and McGowan, R.S. (1992) Influence of consonantal envelope on voice source aerodynamics, *Journal of Phonetics,* 20, 93-110.

MacEachern, M. R. (1999). *Laryngeal Cooccurrence Restrictions*. New York: Garland Publishers.

Maddieson, I. (1997) Phonetic universals, In *The Handbook of Phonetic Sciences* (W. J. Hardcastle and J. Laver, editors), pp. 619-639. Cambridge: Blackwell.

Maddieson, I., Ladefoged, P. and Sands, B. (1999) Clicks in East African Languages, In *African Mosaic: Festschrift for J.A. Louw* (R. Finlayson, editor), pp. 59-91. Pretoria: UNISA Press.

McCarthy, J. J. (1991) Semitic Gutturals and Distinctive Feature Theory, In *Perspectives on Arabic Linguistics III* (B. Comrie and M. Eid, editors), pp. 63-91. Amsterdam: Benjamins.

McCarthy, J. J. (1994) The Phonetics and Phonology of Semitic Pharyngeals, In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (Patricia A. Keating, editor), pp. 191-233. Cambridge: Cambridge University Press.

McDonough, J. (forthcoming 2002) *A phonetic study of Navajo.* Boston: Kluwer.

Miller-Ockhuizen, A. (1998) Towards A Unified Decompositional Analysis of Khoisan Lexical Tone, In *Language, Identity, and Conceptualization among the Khoisan* (M.

Schladt, editor), pp. 217-243. Quellen zur Khoisan-Forschung, 15. Köln: Rudiger Köppe Verlag.

Miller-Ockhuizen, A. (2000) C-V Coarticulation and Complex Consonants: Evidence for Ordering in Click Place Gestures, In *Proceedings of LP '98* (O. Fujimura, B. Joseph and B. Palek, editors), pp. 301-330. Prague: Charles University Press.

Miller-Ockhuizen, A. (2001) Grounding Ju|'hoansi Root Phonotactics: The phonetics of the Guttural OCP and other Acoustic modulations, Unpublished Ohio State University Dissertation in Linguistics.

Miller-Ockhuizen, A. (To appear) Coarticulatory fundamental frequency effects associated with Ju|'hoansi guttural consonants, In *Proceedings of The First Pan-American / Iberian Meeting on Acoustics*. Cancun, Mexico, 2-7 December 2002.

Miller-Ockhuizen, A. (Submitted) Gradient and Categorical F1 patterns in Khoesan languages, *Proceedings of ICPhS 2003*, Barcelona.

Miller-Ockhuizen, A. (In progress) The articulation of Uvularized consonants and epiglottalized vowels in Ju|'hoansi, MS, Cornell University.

Miller-Ockhuizen, A., Namaseb, L., del Teso Craviotto, M. and Iskarous, K. (In progress) Tongue Root Retraction in Ju|'hoansi Click consonants: The Phonetics Basis of the Back Vowel Constraint in Khoesan," Unpublished Manuscript, Cornell University, University of Toronto and Haskins Laboratories.

Nakagawa, H. (1998) A new cluster analysis of clicks and their accompaniments, Paper presented at LP 98, The Ohio State University, September 15th, 1998.

NíChasaide, A. and Gobl, C. (1997) Voice Source Variation, In *The Handbook of Phonetic Sciences* (W. J. Hardcastle and J. Laver, editors), pp. 427-461. Cambridge: Blackwell.

Noll, A. M. (1964) Short-Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection, *Journal of the Acoustical Society of America*, 36 (2), 296-302.

Obrecht, D. (1968) *Effects of the Second formant on the Perception of Velarized Consonants in Arabic.* Mouton: The Hague.

Ohala, J. (1993) Sound change as nature's speech perception experiment, *Speech Communication*, 13, 155-161.

Ohala, J. J. and Ohala, M. (1993) Phonetics of Nasal Phonology, In *Phonetics and Phonology: Nasals, Nasalization and the Velum* (M. K. Huffman and R. Krakow, editors), pp. 225-249. New York: Academic Press.

Pinto, N. and Titze, I. (1990) Unification of perturbation measures in speech signals, *Journal of the Acoustical Society of America*, 87(3), 1278-1289.

Qi, Y. (1992) Time normalization in voice analysis, *Journal of the Acoustical Society of America*, 92 (5), 2569 -2576.

Qi, Y. and Hillman, R.E. (1997) Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals, *Journal of the Acoustical Society of America,* 102 (1), 537 – 543.

Qi, Y., Weinberg, B., Bi, N. and Hess, W.J. (1995) Minimizing the effect of period determination on the computation of amplitude perturbation in voice, *Journal of the Acoustical Society of America,* 97 (4), 2525 – 2532.

Rose, S. (1996) Variable laryngeals and vowel lowering, *Phonology*, 13, 71-117.

Sands, B., Maddieson, I. and Ladefoged, P. (1996) The Phonetic Structures of Hadza, *Studies in African Linguistics*, 25(2), 171-204.

Shahin, K. N. (1997) *Postvelar Harmony: An Examination of its Bases and Crosslinguistic variation*, Unpublished Ph.D. dissertation, University of British Columbia.

Silverman, D. (1997). *Phasing and Recoverability*. New York: Garland.

Silverman, D., Blankenship, B., Kirk, P. and Ladefoged, P. (1995) Phonetic Structures in Jalapa Mazatec, *Anthropological Linguistics*, 37(1), 70-88.

Snyman, J. W. (1975) Z*u/'hoasi Fonologie en Woordeboek*. (Communication of the University of Cape Town School of African Studies 37), Cape Town: A. A. Balkema.

Snyman, J. W. (1977) The Interrupted Juxtaposed Vowels of Zu/'hoasi, In *Khoisan Linguistic Studies 3* (A. Traill, editor), pp. 93-106. A.S. I. Communication #6.

Stevens, K. N. (1997) Articulatory-Acoustic-Auditory Relationships, In *The Handbook of Phonetic Sciences* (W. J. Hardcastle and J. Laver, editors), pp. 462-506, Oxford: Blackwell Publishers.

Stevens, K. N. (1998) *Acoustic Phonetics*. Cambridge: MIT Press.

Talkin, D. (1989) Voicing epoch determination with dynamic programming, *Journal of the Acoustical Society of America,* 85*, Supplement 1*.

Thomas-Vilakati, K. D. (1999) *Coproduction and Coarticulation in IsiZulu Clicks*, Unpublished UCLA Dissertation in Linguistics. Los Angeles: UCLA.

Traill, A. (1985) *Phonetic and Phonological Studies of !Xóõ Bushman*. Quellen zur Khoisan-Forschung, 1. Hamburg: Helmut Buske Verlag.

Traill, A. (1986) The laryngeal sphincter as a phonatory mechanism in !Xóõ Bushman, In *Variation, Culture and Evolution in African Populations (Papers in honour of Hertha deVilliers)(*R. Singer and J. K. Lundy, editors), pp. 123-131. Johannesburg: Witwatersrand University Press.

Traill, A. (1997) Linguistic phonetic features for clicks: Articulatory, acoustic and perceptual evidence, In *African Linguistics at the Crossroads: Papers from Kwaluseni* (R. K. Herbert, editor), pp. 99-117. Köln: Rüdiger Köppe Verlag.

Vossen, R. (1997) Die Khoe-Sprachen: ein Beitrag zur Erforschung der Sprach-geschichte Afrikas. *Quellen zur Khoisan-Forschung,* 12. Köln: Rüdiger Köppe Verlag.

Watkins, J. (1999) CQ of laryngeal gestures and settings in Wa, In *Proceedings of the 14th International Congress of Phonetics Sciences*, San Francisco, August 1999, pp. 1017-1020.

Wright, R., Maddieson, I., Ladefoged, P. and Sands, B. (1995) A phonetic study of Sandawe clicks, *UCLA WPP*, 91, 1-24.

Yumoto, E., W. Gould and T. Baer. (1982) Harmonics-to-noise ratio as an index of the degree of hoarseness, *Journal of the Acoustical Society of America,* 71, 1544-1550.

Zawaydeh, B. (forthcoming) The Interaction of the Phonetics and Phonology of Gutturals, In *Laboratory Phonology VI* (J. Local, R. Ogden and R. Temple, editors). Cambridge: Cambridge University Press.

Zsiga, L. (1995) An acoustic and electropalatographic study of lexical and post-lexical palatalization, In *Papers in Laboratory Phonology IV: American English,* *Phonology and Phonetic Evidence* (B. Connell and A. Aravaniti, editors), pp. 282-302. Cambridge: Cambridge University Press.

Appendix A: Word list for Experiment 1

	Voiceless unaspirated click initial	**Voiced unaspirated click initial**
1.	**Fully breathy**	
	a. !ȁ$^{\text{ɦ}}$ȁ$^{\text{ɦ}}$ 'red crested korhaan'	a. g‖ȁ$^{\text{ɦ}}$ȁ$^{\text{ɦ}}$ 'to press down'
	b. ‖ȍ$^{\text{ɦ}}$ȍ$^{\text{ɦ}}$ 'small annual clover'	b. g!ȍ$^{\text{ɦ}}$ȍ$^{\text{ɦ}}$ 'dust, smoke'
2.	**Fully epiglottalized**	
	a. !ȁ$^{\text{ʕ}}$ȁ$^{\text{ʕ}}$ 'iron, steel'	a. g‖à$^{\text{ʕ}}$à$^{\text{ʕ}}$ 'aunt'
	b. !ȍ$^{\text{ʕ}}$ȍ$^{\text{ʕ}}$ 'pan'	b. g!ȍ$^{\text{ʕ}}$ȍ$^{\text{ʕ}}$ 'male'
3.	**Glottalized**	
	a. !ȁ$^{\text{ʔ}}$à 'dry season'	a. gǂà$^{\text{ʔ}}$á 'wide '
	b. tʃò$^{\text{ʔ}}$ó 'to be unconscious'	b. g!ò$^{\text{ʔ}}$ó 'cough'

Appendix B: Wordlist for Experiment 2

1. **Voiceless unaspirated click**
   a. ‖áá 'to warm hands by fire'
   b. ǂòà 'sitting mat'
   c. !óó 'older brother'

2. **Voiceless aspirated click**
   a. ǂ^{h}àà 'path'
   b. |^{h}óá̰ 'to rub away'
   c. !^{h}òò 'forcibly'

3. **Voiced unaspirated click**
   a. g!àà 'rain'
   b. g!òà 'wet leaf'
   c. g!óó 'petrol'

4. **Voiced aspirated click**
   a. g|$^{\text{ɦ}}$àà 'place for drying meat'
   b. gǂ$^{\text{ɦ}}$òà 'dog'

8. **Glottalized click**
   a. !$^{\text{ʔ}}$àà^{n} 'to catch up'
   b. !$^{\text{ʔ}}$òà^{n} 'yawn'
   c. !$^{\text{ʔ}}$òò^{n} 'leadwood tree’

c. g!ʱòó 'to sit'

5. **Voiceless epiglottalized click**
   a. !ˤàà 'to dry out'
   b. ǁˤòà 'to work'
   c. !ˤòò 'to hate'

6. **Voiceless uvularized click**
   a. ǂᵡàà 'moist sand'
   b. ǂᵡòà 'to pacify'
   c. !ᵡòò 'to be unsuccessful'

7. **Voiced uvularized click**
   a. g!ʶàà 'to take out'
   b. g!ʶòà 'knee'
   c. g|ʶòò 'to give something to someone'

Appendix C: Wordlist for Experiment 3

1. **Partially breathy**
   a. ǂà̤ʱà 'plain'
   b. |ò̤ʱò 'to follow'

2. **Partially epiglottalized**
   a. ǁà̤ˤà 'to hold down'
   b. ǂò̤ˤó 'uterus'