QUECLARATIVES IN XHOSA : AN ACOUSTIC AND PERCEPTUAL ANALYSIS

by

Catherine Jacquelynn Julia Jones

Thesis presented in fulfilment of the requirements for the degree D.Litt at the University of Stellenbosch.

March 2001

Promoter: Prof. J.C. Roux

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or in part been submitted at any university for a degree.

D

SUMMARY

Key words:
acoustic speech analysis, speech synthesis, speech perception, copulative queclarative, linguistics, psycholinguistics, human language technology

This study investigates the notion of interrogativity in Xhosa as expressed in the form of Queclaratives. Queclaratives, or statements which are question-like in function, have been studied in many languages of the world. Unfortunately with regard to the Bantu languages, studies relating to interrogativity in general have largely been impressionistic in nature.

This research comprised two aspects of analysis. These included an acoustic and a perceptual analysis of data. The reason for this approach is that, without this combination the results could have been considered to be suspect and lacking in authenticity.

The acoustic analysis was conducted on 858 words in statement and queclarative pairs. Significant parameters were extracted and these were then statistically analyzed. The results revealed that duration on the penultimate vowel, pitch on the penultimate vowel and the overall raised pitch of queclaratives as opposed to statements were indeed the acoustically significant parameters differentiating statements from queclaratives. However as is well known, there is no one-to-one relationship between the acoustic signal and its perception and, therefore, it is imperative that such findings also be perceptually tested.

The perceptual testing of these parameters was conducted in an attempt to elicit whether they were perceptually significant and also at what point in the utterance listeners could differentiate between queclaratives and statements.

The next progression was the compilation of carefully designed perception tests on the acoustically significant parameters. Two experiments were compiled using stimuli that were manipulations of the original signal of one of the selected informant's utterances.

These tests were administered on multimedia computers in the Language Laboratory at the University of Stellenbosch using 64 subjects for the first experiment and 63 for the second.

The results of the perception tests showed that duration and pitch on the penultimate syllable are perceptually highly significant in differentiating statements from queclaratives. However the results also indicated very early recognition of the different forms with minimal speech segments from which the penultimate vowels were absent altogether. This then suggests that the perceptual judgements made earlier in the utterance may be either reinforced or overridden by the duration and pitch on the penultimate vowel.

These results have assisted in the validation of some impressionistic claims made within the Bantu and other languages, while refuting others. However, as this corpus of data has included research on copulative queclaratives, it appeals for further research on this subject using any other linguistic markers.

The results have also been evaluated in terms of their possible contribution to the related disciplines of psycholinguistics, linguistics and human language technologies. In so doing, the thesis makes an urgent appeal to researchers to pursue this experimental approach to language research. Another appeal is made for an awareness campaign as to the importance of this approach in harnessing the power of language for the development of language and society as a whole. The fertility of the South African society lies in its richness of multilingualism and the necessity for the improvement of the dissemination of information to all people of all languages and the improvement of communication between people in general, including those less fortunate in terms of literacy skills.

OPSOMMING

Sleutelwoorde:
akoestiese spraakanalise, spraaksintese, spraakpersepsie, kopulatiewe stelvraag, linguistiek, psigolinguistiek, taal-en-spraaktegnologie

In hierdie projek word die aard van vraagstelling in Xhosa ondersoek met betrekking tot die stelvraag-vorm. Stelvrae, of stellings wat ook as vrae kan funksioneer, is reeds bestudeer vir heelwat wêreldtale. Oor die algemeen was studies oor vraagstelling in die Afrikatale egter grootliks impressionisties van aard.

Hierdie navorsingsprojek het uit twee analisekomponente bestaan, naamlik 'n akoestiese analise van die data en 'n reeks persepsuele eksperimente. Sonder die kombinasie van die twee tipes analise sou die resultate van die navorsing minder kredietwaardig gewees het.

Die akoestiese analise is gedoen op 858 woordpare bestaande uit stellings en stelvrae. Die data is statisties ontleed en die relevante parameters is onttrek. Die resultate het daarop gedui dat die duur en toonhoogte van die voorlaaste vokaal sowel as die register van die hele woord belangrike parameters is in die onderskeid tussen stellings en stelvrae. Aangesien dit wel bekend is dat daar nie 'n een-tot-een verwantskap tussen die akoestiese klanksein en die persepsie daarvan is nie, is dit noodsaaklik om ook 'n persepsuele eksperiment uit te voer. Die persepsuele toetse is so opgestel dat bepaal kon word watter akoestiese parameters ook persepsueel relevant is en om die vroegste sillable te vind waar luisteraars reeds die onderskeid tussen die twee vorme kan maak.

Die volgende stap was om stimuli vir die persepsietoetse op te stel wat inderdaad bogenoemde resultate sou lewer. Stimuli is geskep deur die spraakdata van een spreker te manipuleer. Die persepsietoetse is toe uitgevoer op multimedia-rekenaars in die Taallaboratorium van die Universiteit van Stellenbosch.

Die resultate van die persepsietoetse het gewys dat die duur en toonhoogte van die voorlaaste sillabe ook persepsueel belangrik is in die ondersekeid tussen die

verskillende vorme. Dit was ook duidelik dat proefpersone die vorme van mekaar kon onderskei met minimale hoeveelhede inligting waar die voorlaaste en laaste sillabes heeltemal afwesig was. Dit dui daarop dat luisteraars persepsuele besluite baie vroeg in die woord neem, maar dat hierdie besluite óf versterk óf omgekeer kan word deur die duur en toonhoogte van die voorlaaste sillabe.

Die resultate van die navorsing het sekere impressionistiese stellings ten opsigte van Afrikatale ondersteun terwyl ander stellings as foutief bewys is. Een van die belangrike bevindings was dat die impressionistiese standpunt dat stellings 'n dalende intonasiekontoer en vrae 'n stygende intonasiekontoer tydens die afloop van die uiting het, 'n oorvereenvondiging is. Hierdie werk is gedoen op enkelwoord-kopulatief stelvrae en leen dit daartoe om uitgebrei te word na frases en sinne vir toekomstige navorsing.

Verder is die navorsingsresultate in verband gebring met verwante dissiplines soos psigolinguistiek, linguistiek en taal- en spraaktegnologie. 'n Pleidooi is gelewer vir 'n bewusmakingsveldtog om die belangrikheid van hierdie tipe navorsing te beklemtoon om die potensiaal van taal te benut vir die ontwikkeling van Suider-Afrikaanse tale en gemeenskappe. Die rykheid van ons gemeenskap lê in die veeltaligheid daarvan en bied besondere uitdagings om die verspreiding van inligting na alle mense van alle tale te verbeter en om die kommunikasie tussen mense in die algemeen, maar ook spesifiek vir diegene met laer vlakke van geletterdheid.

ACKNOWLEDGEMENTS

The completion of a thesis is not the work of one alone. It owes acknowledgement and thanks to all who have shown interest and genuine concern which in turn has provided the motivation, support and the infrastructure to reproduce work of this nature in it's final form.

I would like to acknowledge and thank my promoter Professor Justus Roux for his support, encouragement and assistance throughout this period of study and research.

Thank you to my colleagues in the Department of African Languages at UNISA (both young and more mature) for their concern and encouragement. Special tribute should be paid to Professor Sonja Bosch who at a very late stage gave me invaluable indicators. To Joyce (Department of Statistics), my most grateful thanks to you for your patience and guidance at a very busy time in the academic calendar.

My grateful thanks to my friends at the University of Stellenbosch, especially Surena, for her constant help and support. To Pumlani Sibula, who was so supportive and true to his word in organising Xhosa first language speakers as informants and listeners for the perceptual testing of data. Ndithi enkosi, Pum.

To my friend Hleziphi in the UNISA library, always pleasant and always so willing to help me. Hlez, ndiyabulela kakhulu wethu!

To my friend and colleague Rose, who has supported, encouraged and never given up on me over many years of studying - thanks Rose I am truly grateful to you.

To Pip, thanks for your unstinting help and support not only in the language laboratory but in so many other ways. Always willing, kind and supportive - thank you Pip.

To Jan, a young man whom I find difficult to thank. His knowledge, support, kindness and wonderful academic attributes I cannot adequately acknowledge either verbally or

in writing. Jan I shall always remember you for who you are, a young gentleman with so much to offer.

To my dear mother Glen, my family Jan and Noel, John and Ei and Gill and Pete for their evergreen support and interest in my work - my most grateful thanks.

To my sons Gwilym and Brynmor thanks for always being there for Dad and me and for your incredible patience over 23 years of studying.

Finally, to the most important person in my life – Donald. Without you this thesis nor anything else would ever have been accomplished. For your wonderful patience, support, understanding and love over all these years – this really is for you. Thank you.

I've learned –
that credentials on the wall
do not make you a decent human being.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ALASA	African Language Association of South Africa
ASL	Analysis/Synthesis Lab
ASR	automatic speech recognition
CSL	Computerized Speech Lab
DACST	Department of Arts and Culture, Science and Technology
DSP	digital signal processing
F0	pitch
HLT	human language technology
LPC	linear prediction coefficients
LTS	letter-to-sound (rules)
ms	milliseconds
NLP	natural language processing
PANSALB	Pan South African Language Board
RUEPUS	Research Unit for Experimental Phonology University of Stellenbosch
TD-PSOLA	time domain pitch synchronous overlap and add
TTS	text-to-speech (system)

SYMBOLS

F_s	sampling frequency in Hz
t_x	time instant in seconds of tag *x* within a word
d_x	duration of syllable *x* in seconds
p_x	pitch of syllable *x* in Hz
e_x	relative loudness of syllable *x*
E_x	absolute loudness of syllable *x*
$\overline{\mu}$	mean
$\widetilde{\mu}$	median
σ	standard deviation
$\frac{\overline{\mu}}{\sigma}$	mean divided by the standard deviation
$\left\lvert\frac{\overline{\mu}}{\sigma}\right\rvert$	magnitude of the mean divided by the standard deviation
Q_{wf}	feature *f* of word *w* for a queclarative token
S_{wf}	feature *f* of word *w* for a statement token
D_{wf}	difference feature *f* of word *w*
$n_{i,Q}$	number of queclarative responses for the *i*th stimulus
$n_{i,S}$	number of statement responses for the *i*th stimulus
$n_{i,U}$	number of undecided responses for the *i*th stimulus
$f_{i,Q}$	observed frequency for queclaratives for the *i*th stimulus
$f_{i,S}$	observed frequency for statements for the *i*th stimulus
e_i	expected frequency for the *i*th stimulus
$e_{i,Q}$	expected frequency for queclaratives for the *i*th stimulus
$e_{i,S}$	expected frequency for statements for the *i*th stimulus
χ_i^2	chi square value for the *i*th stimulus
Δ	change
$\approx$	approximately equal to
$\neq$	not equal to
Σ	sum
Δd	change in duration [s]

Δp	change in pitch [Hz]
f	feature
$f()$	function
S	significance value according to Wilcoxon test
t	time
v	value v
w	word w

CHAPTER 1

INTRODUCTION

Speech science is a discipline of investigation at one of the important frontiers of human inquiry challenging us all.

(Borden and Harris, 1984:275)

1.1 Speech, language and communication

This thesis *'Queclaratives in Xhosa: An acoustic and perceptual analysis'* focuses on a specific type of interrogative, called a *queclarative*. This category of interrogative may be defined as an interrogative which is declarative in form yet functions as a question. In languages of the world interrogatives are generally conspicuous by their use of some lexical or syntactic marker, while queclaratives, which do not exhibit such markers, by implication use other methods of conveying the intended message of questioning. This study therefore is concerned with the strategies used by humans in encoding and decoding this specific type of question, which by definition is devoid of linguistic markers.

The generation of speech does not simply require the articulation of sequential speech sounds that constitute an utterance but also includes other features which may or may not contribute to the meaning of an utterance. Speech, which is the production and conversion of language into audible sounds, may be analyzed or observed with due consideration, not only to articulatory features, but also to 'other' features. These other features include aspects like amplitude or loudness, speech tempo or rate of production, pitch or fundamental frequency and of course voice quality. Components like pitch or fundamental frequency may seem to be superimposed upon the linguistic segments of the utterance and are therefore referred to as suprasegmentals or as a suprasegmental or prosodic tier. The use of suprasegmental features may possibly add meaning, often add an emotion and may also assist in the process of disambiguation of the spoken word.

Speech therefore is the audible component of language and, according to Chomsky (1972), the relationship between language and speech resides in linguistic competence *vis à vis* linguistic performance. Linguistic competence therefore implies not only an ability to speak, to articulate and to produce sounds but also includes an understanding of the ordering of these sounds for purposes of intelligibility. The process of communication by a speaker therefore can be seen as one of transforming an intended phrase into an acoustic representation. A further transformation occurs from an acoustic level to the decoding of this representation by a listener, which is representative of a perceptual level. Therefore this research on queclaratives, which includes both acoustic and perceptual analyses, is concerned with this very process of communication, i.e. in this instance of communicating either a question or a statement.

1.2 Motivation

The majority of publications on research relating to interrogativity emphasize specifically the use of some lexical or syntactic marker for purposes of differentiation. An orientation of this nature can therefore also be considered to be purely linguistic. However, interrogativity, which may also be expressed without the use of such markers, may also be realized by employing some strategy other than meta-language.

The use of strategies which exclude the use of lexical or syntactic markers has been acknowledged by many scholars and Ultan (1978:219), in his study of the interrogative systems of some 79 languages, states that,

> ...Furthermore, although Q-intonation [i.e. Question intonation, CJJ] is often accompanied by some other Q-marker (e.g. inversion, QP, etc.) most languages (perhaps all?) also have Q-utterances distinguished from their corresponding declarative utterances solely by means of Q-intonation.

There is a vast corpus of publications which acknowledge the use of intonation as one of the 'other' strategies differentiating statements from questions and since this research is concerned specifically with identifying 'other' strategies, some of the relevant publications will be cited below. As this field of research, which includes an experimental approach, is relatively new in the South African context, the references cited will relate to languages beyond its borders since to date, according to the author's knowledge, there has been no study that addresses this issue within the South African Bantu languages.

The publications and comments on the use of prosody in the differentiation process may be classified into two distinct areas, namely those that are purely impressionistic in nature and those that include an experimental analysis of data. Impressionistic publications and comments include those made on languages such as Russian (Comrie, 1984; Restan, 1972), Mandarin (Harris, 1984), English (Fries, 1964), Dzamba, Likila and Lingala (Bokamba, 1976), Bengali (Shaha, 1984), Edo (Ọmọruyi, 1988), Engenni (Thomas, 1974) Palermo Italian (Grice 1995). However, there is also a substantial number of publications on the use of intonation in this differentiation process, which have adopted an experimental approach incorporating the analysis and interpretation of data. These investigations include research in Australian English (Allan, 1984), Kipare (Herman, 1996), Mandarin (Shen, 1990), English (Geluykens, 1987, 1988 and 1989), English (Hirschberg and Ward, 1992), Osaka Japanese (Miura and Hara, 1995), Japanese (Hinds, 1984) Korean (Jun and Oh, 1994, 1996), American English (Studdert-Kennedy and Hadding, 1973), to name but a few.

Some of the results relating to the use of fundamental frequency or linguistic tone in this process are presented below, showing that this strategy used for differentiation seems to be evidenced in many languages. It also becomes apparent that there are universal tendencies across different languages as to *how* and *where* within the utterance this strategy is used in this process of differentiation. For example it may be used initially, finally or over the entire contour, or there may be a combination of these tonal differences related specifically to interrogatives or to statements. This poses a question: in other words can interrogatives and declaratives be identified by a typical intonational specification?

According to Herman (1996) both declaratives and yes/no questions have final lowering, but the pitch range of yes/no questions is greatly expanded. Miura and Hara (1995) show that in Osaka Japanese, sentence initial F0 lowering and sentence final raising are some characteristics of rhetorical questions, which express disbelief or incredulity. Also according to Herman (1996), the expression of yes/no questions using the manipulation of pitch is not unique to Kipare. In Bolinger's survey of Hermann's (1942) work (which found higher pitch somewhere in the utterance in yes/no questions in 175 languages) he adds yet another 41 languages in which higher pitch is found in yes/no questions. Shen (1990) noted that the same combinations of elements, which exist in Kipare, are also found in Mandarin. These include non-expanded pitch range with final lowering, expanded pitch range with final lowering or expanded pitch range without final lowering. She notes that non-expanded pitch range (key or register) is used for assertive sentences, while expanded pitch range is used for interrogatives. This expanded pitch range or higher key, she states, is characteristic of lexically or grammatically unmarked questions. Jun and Oh (1996) show that in Korean, although not all speakers use the same strategy for expressing incredulity, a larger pitch range is one of the strategies used to differentiate between incredulity questions and 'wh'-questions. Eduard Hermann (1942) surveyed 175 languages, and identified a high interrogative pitch in some form or other everywhere. He stated that this high pitch not only implied a rising terminal but can also mean a relatively high overall pitch level. Hadding-Koch (1961) also indicated that high pitch level in questions was also evident in Swedish and this was corroborated by perceptual experiments.

The South African situation with regard to this topic of research on interrogativity using 'other' strategies is predominantly impressionistic in nature. This is evidenced by the comments related to interrogativity in Xhosa (Lanham, 1963; Riordan, 1969), in Zulu (Khumalo, 1981; Laughren, 1984; Nkabinde, 1999), in Northern Sotho (Lombard, 1980; Louwrens, 1987; Steyn, 1991), in Tswana (Mathibela, 1989; Khoali, 1993), in Sesotho (Doke and Mofokeng, 1957; Demuth, 1995).

According to Nkabinde (1999:379) the interrogative sentence has a quick tempo, a high pitch (without specifying at what point in the utterance) and no down-stepping. The declarative sentence is, however, characterized by a cadent intonation with unchecked length on the penultimate syllable of the only or last word in a sentence and it has a low pitch.

Steyn (1991:15), in her research on interrogativity in Northern Sotho, states that

> Dit blyk dus dat in die meeste Afrikatale die woordvolgorde van 'n vraagsin dieselfde is as dié van die ooreenstemmende stelsin, maar dat 'n verandering in suprasegmentele eienskappe soos intonasie en lengte, wel intree.

This impressionistic claim regarding the role played by intonation is also made by Demuth (1995:52) when she says that

> The word order in Sesotho yes-no questions is the same as that of declaratives; only prosodic phenomena serve to distinguish the two.

Mathibela (1989:16) also acknowledges the use of intonation in Tswana by stating that

> These questions in 15(a) and 15(b) do not have interrogative particles. They are formed by intonation.

The use of an experimental approach to research on this topic in the South African Bantu languages has been restricted and confined to very limited work in Xhosa by Louw (1968), Theron (1991) and Roux (n.d.).

Louw (1968), using a pitch extractor, did present experimental phonetic data on pitch contours in Xhosa and did touch on the differences between questions and statements. However, this research, while supportive of impressionistic claims, did not quantify it to any degree. Roux (n.d.), in his study on the prosodics of declarative and question sentences in Xhosa, makes a very relevant comment in his conclusion:

> Although these acoustic analyses present a wealth of information, the limitations are perfectly clear. This type of data alone will not necessarily bring us closer to an understanding of the relevant perceptual cues involved. The focus should now change to experimental perceptual investigation ...

Theron (1991) has to date, to the author's knowledge, presented the most extensive research of an experimental nature on Xhosa queclaratives. Her findings, after the acoustic analysis of data, proposed that three factors influence disambiguation in Xhosa, namely sentence duration and speech tempo, penultimate vowel duration, and tonal register and movement. However it is true to say that her study did not include any perceptual testing of data. While the need for an experimental approach, which includes an acoustic analysis of data, is acknowledged, such an approach to the exclusion of the role played by psycholinguistics cannot be accepted as conclusive. According to Studdert-Kennedy (1998:170)

> ...the units of the acoustic signal do not correspond one for one, with the units of perception...

Due to the impressionistic bias and the absence of perceptual testing of data it becomes obvious that a significant hiatus exists within this field of research in South Africa. It seems apparent therefore that the validity and authenticity of impressionistic claims should also be investigated.

The motivation for a study of this nature, which includes both acoustic and perceptual investigations, arises from the predominantly impressionistic claims, limited experimental investigation to date as well as a total disregard for any perceptual testing of data specifically in South Africa.

1.3 Objectives

This research comprises an acoustic and a perceptual component, with the following objectives:

1. To determine the typical acoustic differences which exist between statements and queclaratives in Xhosa.
2. To determine which acoustic parameters are actually relevant for mother tongue listeners of Xhosa to make a perceptual distinction between statements and queclaratives.
3. To determine whether there are human universals present in the disambiguation process which are applicable to Xhosa in differentiating statements from queclaratives.

The acoustic analysis of data is conducted in order to establish specifically the acoustic differences between the original signals of statements and those of queclaratives. In order to achieve the first objective which is to establish the acoustic differences between statements and queclaratives, measurements of duration, pitch and energy were calculated on each individual phoneme and syllable of all the words which comprised the prepared corpus of data used in the acoustic analysis. Having established these facts, further statistical analyses of these results is done using three non-parametric tests which reveal the areas of significant acoustic differences between these two entities in the relevant parameters.

The second objective which emanated from the acoustic analysis was to test the acoustically significant parameters perceptually in order to establish whether these acoustic differences were indeed relevant in changing people's perception of an utterance from a question to a statement and vice versa. It was hoped that the results obtained from the perceptual testing would then provide researchers with cues used by humans for purposes of differentiation or distinguishing between statements and queclaratives. This analysis is an attempt to validate or repudiate the impressionistic claims made by scholars with regard to languages in general and African languages in particular.

The third objective was the evaluation of the research results documented in the differentiation process for other languages as compared with those obtained in this experiment for Xhosa. This evaluation enabled the researcher to reflect on possible human universals and or language-specific strategies employed by listeners in the differentiation process.

The results obtained from this project not only provide linguistic, psycholinguistic and phonetic contributions but also make a contribution to the field of human language technologies (HLT[1]) within the South African context. Research in the field of human language technology in Europe, with its multilingual component, is well known, yet in South Africa, which enjoys the same milieu of multilingualism, it is as yet a relatively unknown field, albeit an area of tremendous potential. South African society, so diverse in

---

[1] http://www.hltcentral.org

cultural backgrounds, yet so rich in its multiple use of languages, is a prime example of where research and development in human language technologies may well serve to improve many facets of life for all.

1.4 Methodology

The methodology adopted for this research project is experimental in nature, and its use is limited to research within the African language context. This experimental research involves firstly an acoustic analysis of data. This is followed and accompanied by a perceptual analysis in the form of perceptual testing. On these results implications for the disciplines of psycholinguistics, linguistics and human language technology are then drawn.

Experiments related to duration, pitch, tempo and register are conducted and then evaluated and compared to prevailing impressionistic claims.

Experimental work on the topic of interrogativity includes Theron's (1991) contribution, which investigated sentence duration, tempo, length on the penultimate syllable, register and tonal down-drift. Her findings to a great extent supported impressionistic claims on the parameters analyzed. Roux (n.d.) also conducted experimental analyses on the nature of tonal declination in Xhosa in statement and echo-questions. While both researchers used acoustic analyses of a corpus of data, which in many instances seem to support impressionistic claims, neither have made any attempt at validating their findings from a perceptual perspective.

1.5 Summary of chapters

Chapter 1 of this thesis provides an introduction to the concepts of speech, language and the process of communication. As the title suggests, this research includes two interdependent scientific experimental approaches, namely an acoustic and a perceptual analysis. The term queclarative is defined and described, justifying its use for this investigation. This chapter also includes the motivation for and objectives of this type of research, which involves two sequential analyses, and explains, within broad parameters, the methodology and the instrumentation employed.

Chapter 2 provides an introduction to *acoustic analyses,* previously done specifically within the Bantu languages, and then an in-depth and comprehensive explanation of the acoustic analysis of the corpus of data selected for this project. The chapter also includes methods of collection of data, recording, preparation, analysis and statistical analysis of the acoustically analyzed data. The results of the acoustically analyzed data (of each analyzed parameter) is represented in tabulated form and included in Appendix A.

Chapter 3 introduces the concept of *perceptual testing* and comments on previous research in this field in general and specifically within the Bantu languages. Thereafter follows a description of the design and compilation of the perception tests on each parameter tested and the procedure used in the administration of these tests in the language laboratory. The perception test results form the conclusion to this chapter and the tabulated results on each parameter are presented in Appendix B.

A summary of the results of both the acoustic and perceptual analyses is presented in Chapter 4. The results and the implications of this project are then discussed, relating specifically to the topics of psycholinguistics, linguistics and human language technology.

Finally a comprehensive overview of the aims of this project and the results obtained relating to the original objectives is provided in Chapter 5. Then follows a motivation for the use of experimentally based research on the Bantu languages lending authenticity and credibility to research output. The chapter is concluded with a discussion on the envisaged future of this experimental approach within the South African context and the current activities in this field.

CHAPTER 2
ACOUSTIC ANALYSIS

"Holla your name to the reverberate hills,
And make the babbling gossip of the air
Cry out."

William Shakespeare, Twelfth Night (in Borden and Harris, 1984:25)

2.1 Introduction

This chapter provides a description of the initial part of this research project, which comprises an acoustic analysis of queclaratives in Xhosa. Research in the field of interrogativity has generally concentrated on pragmatic features of marking question formation in languages. This approach includes descriptions of the use of lexical or syntactic markers or tags used as cues to depict interrogativity. However, impressionistic claims have also been made for so-called 'other' strategies employed in the disambiguation process. These 'other' strategies include any other strategy that may be employed in this process i.e. any method that does not utilize any lexical or syntactic markers. This may therefore include prosodic features, which may contribute to this process. While publications and comments in languages other than Bantu languages have been impressionistic, a vast number also include an experimental approach to the topic of interrogativity. However, within Bantu languages, the publications contributing to this topic have not only been predominantly impressionistic in nature, but have often been merely an interpretation and description of data by scholars. These claims, both impressionistic and experimentally-based, made on the use of so-called 'other' strategies in encoding the message of questioning relate specifically to duration, intonation, tempo and loudness.

This research project therefore originated as a reaction to the predominantly impressionistic approach which has been adopted in the past and which renders

> research in this field suspect, contradictory in nature and even lacking in authenticity. ... This research therefore attempts to bring the study of African languages at least in line with generally accepted practice in the field of contemporary phonetics ... (Jones et al., 1998:5)

It is commonly acknowledged that human language, whether it is spoken or written, is ambiguous. Linguistically, ambiguity is considered to arise from three levels: lexical/semantic, syntactic, and pragmatic. Ambiguity at a lexical/semantic level has been investigated mainly by psychologists.

For syntactically ambiguous sentences, many acoustic analyses have been conducted, most of which concern the English language (Cooper, 1976; Klatt, 1976; Lehiste, Olive and Streeter, 1976; Lehiste, 1980; Garro and Parker, 1983). Many of these studies have revealed uniformity in their results, indicating that the fundamental frequency (F0) and segmental duration are important factors for disambiguating syntactically ambiguous sentences. According to Miura and Hara (1995:291-292), perceptual studies have also been carried out on these sentences, using both natural and synthetic speech, by Streeter (1978); Berkovits (1981), Scott (1982), Price, Ostendorf, Shattuck-Hufnagel and Fong (1991). These studies have endorsed the importance of fundamental frequency and segmental duration in perceptual disambiguation. Hirschberg and Ward in 1992 investigated English sentences pronounced with *uncertainty* and *incredulity* of a rise-fall-rise contour and came to the conclusion that F0 or fundamental frequency is a primary factor for differentiating between the two pragmatic meanings. According to Hirschberg and Ward (1992:247-248)

> ...the substitution of the "incredulity" pitch range (via substitution of F0) seemed to have influenced subject judgements toward "incredulity" more than the substitution of any other of the independent variables (amplitude, duration, and spectral features), with substitution of spectral features a distant second.

This research by Hirschberg and Ward (1992) seems to highlight the significance of fundamental frequency in the disambiguation process, especially as they found that amplitude and duration appeared to play no significant role in subjects' interpretations of the contour presented for perceptual testing. These results also stimulate thoughts on the possible importance of the role of F0 in the interpretation of other contours carrying intonational meaning.

Miura and Hara (1995) conducted an analysis on the production and perception of rhetorical questions in Osaka Japanese to determine whether F0 and duration contribute to the discrimination between literal and rhetorical questions. This was done after studies on syntactically ambiguous sentences in Japanese by Uyeno, Hayashibe, Imai, Imagawa and Kiritani (1980) and Azuma and Tsukuma (1990), which showed the use of F0 and pause as well as segmental duration in differentiating between ambiguous forms.

Research within Bantu languages has been predominantly impressionistic. Differentiating statements from questions includes comments on the use of *speech tempo* by Louw (1968) and Khumalo (1981), on *tonal register* by Riordan (1969) and Khumalo (1981) and also on *word and penultimate syllable duration* by Lanham (1963).

Although impressionistic claims have dominated this field of research in Bantu languages, limited experimental research in languages relating to the encoding of statements and questions has been done on Xhosa by Louw, (1968), Theron, (1991), and Roux, (n.d.) and on Chichewa, a Bantu tone language of Malawi, by Myers, (1999). All these experimental works included the use of data that did not show evidence of any linguistic, syntactic or morphological markers characteristically used to distinguish between statements and questions. Therefore the data used in each of these experimental studies is comparable with the data used in this study, as the questions analyzed here were identical in linguistic form to their statement counterpart. Based therefore on this external characteristic, these questions are referred to and defined here as queclaratives.

The claims, both impressionistic and those made after experimental research for the Bantu and other languages, have been acknowledged. This analysis, having taken cognisance of previous research in this field, attempts to identify the acoustic differences between the original signals of statements and queclaratives specifically in Xhosa. These results are then also statistically analyzed and then perceptually tested to increase the validity and authenticity of the research output.

This chapter provides an introduction to and reasons for the acoustic analysis of the paired Xhosa statements and queclaratives, followed by a description of the aims of this acoustic level of research. Thereafter follows a description of the corpus used, which is presented in Table 2.1, and an explanation of the preparation of the data for analysis, which includes the methods used for the recording and tagging of examples. Included here is an example of a screen layout used in the preparation (tagging) of the original signal within the CSL system.

Following this is an exemplification of the three tests used for descriptive (non-parametric) statistical analysis. The tests used were the sign test, the Wilcoxon signed ranked pairs test and the mean divided by the standard deviation. Where the Wilcoxon test returned equal significance values for different features, the mean divided by the standard deviation was then used to resolve this and to identify the more important feature.

Then an introduction to each analyzed parameter is provided, followed by an explanation of the calculations of duration, pitch and loudness. These three parameters were selected on the grounds of the claims made (for Bantu and other languages), whether based on impression or after instrumental and perceptual analyses either singularly or collectively, as being relevant in the encoding process. Each explanation of the relevant calculation is followed by the statistical results on each of these parameters. Finally the chapter is summarized, to give an holistic overview of this acoustic analysis.

Tables showing the significant syllables, derived from the numerical calculations within each analyzed parameter, are included in Appendix A under the relevant sub-headings. These are represented in bold typeface and subscripted, using numerical values, in the order of significance. Tables providing summarized statistical results of the acoustically significant syllables are presented under the relevant headings in the body of this text and tabulated pitch contour results appear where relevant.

Finally, in the summary, after evaluation of the statistical results, proposals are made concerning the subsequent perceptual testing of the acoustically significant parameters which have been verified as statistically relevant.

2.2 Aims

The aims of the acoustic analysis were:

1. to determine the nature and extent to which the **duration** of the syllables and the utterance in its entirety facilitate a distinction between a statement and a queclarative;
2. to determine the nature and extent to which the **pitch** on the vowels and the overall pitch level on the utterance facilitates a distinction between statements and queclaratives;
3. to determine to what extent, from the mean **loudness** levels calculated, loudness facilitates a distinction between statements and queclaratives;
4. to determine whether there is any **evidence of interdependence** between the features of duration, pitch and loudness between statements and queclaratives.

2.3 Preparation of data for analysis

2.3.1 Corpus

The corpus consisted of copulative queclaratives. Three copulative forms were selected for 13 noun classes in Xhosa. In Xhosa one can form complete predicatives and even complete sentences by using the copulative form. Such non-verbal predicatives are called copulatives and have the meaning ‘it is’. Copulatives are derived from words other than a verb by means of copulative formatives. Copulatives were selected as the corpus of choice because each is interpreted semantically as a phrase. The type of question that evolves from each of these forms, when posing a question may be said to be related to what in other languages are considered, incredulous questions. This corpus comprised 39 queclarative/statement pairs (Table 2.1), or 78 words per informant for each of the eleven speakers, resulted in a corpus for analysis of 858 words.

The impersonal identifying copulative form that was used for this corpus is formed by prefixing the copula formatives to the nominal of each class. The copula formative *ng-* is prefixed to the classes 1, 2, 3 and 6 (note that these class prefixes commence with *u-* or *a-* and classes 4 and 9 whose prefixes commence with *i-* have *y-* prefixed to the full class prefix). Classes 5, 7, 8, 10, 11, 14 and 15 have the copulative formative, the same in form as their subject concords prefixed to the real prefix of the nominal (with the pre-prefix elided). For example:

(2.1)	Class 1	umntu	→	ng + umntu	→	ngumntu	*It is a person.*
	Class 4	imithi	→	y + imithi	→	yimithi	*It is a tree.*
	Class 7	isitya	→	si +(i)sitya	→	sisitya	*It is a dish.*

Table 2.1 Data corpus.

Singular			Plural		
Class	Xhosa	English	Class	Xhosa	English
1	Ngumntu. Ngumfazi. Ngumfana.	*It is a person.* *It is a woman.* *It is a young man.*	2	Ngabantu. Ngabafazi. Ngabafana.	*They are people.* *They are women.* *They are young men.*
3	Ngumthi. Ngumlilo. Ngumfula.	*It is a tree.* *It is a fire.* *It is a rivulet.*	4	Yimithi. Yimililo. Yimifula.	*They are trees.* *They are fires.* *They are rivulets.*
5	Lilitye. Liliso. Lilifu.	*It is a stone.* *It is an eye.* *It is a cloud.*	6	Ngamatye. Ngamehlo. Ngamafu.	*They are stones.* *They are eyes.* *They are clouds.*
7	Sisitya. Sisifo. Sisilo.	*It is a dish.* *It is an illness.* *It is an animal.*	8	Zizitya. Zizifo. Zizilo.	*They are dishes.* *They are illnesses.* *They are animals.*
9	Yinja. Yintloko. Yinkomo.	*It is a dog.* *It is a head.* *It is a head of cattle. (singular)*	10	Zizinja. Ziintloko. Ziinkomo.	*They are dogs.* *They are heads.* *They are heads of cattle. (plural)*
11	Luluthi. Lulwimi. Lulwandle.	*It is a stick.* *It is a tongue.* *It is a sea.*			
14	Bubuso. Bubuhlanti. Bubusika.	*It is a face.* *It is a kraal. (cattle)* *It is winter.*			
15	Kukutya. Kukwindla. Kukunene.	*It is food.* *It is autumn.* *It is right sides.*			

2.3.2 Recording

Eleven mother-tongue speakers of Xhosa, without any known speech or hearing defects, were asked to read aloud the stimuli, which were each presented on a single card, in random order. The queclaratives were read in citation form with breaks between words, firstly as a statement and then as a queclarative. In order to avoid coarticulation effects no carrier phrase was used. Two recordings were made of each token and the clearer token was finally chosen for analysis. In total 858 words were eventually analyzed.

The subjects were all males below the age of 45 years whose mother-tongue was Xhosa. For purposes of consistency it was decided to use one gender, and, to have an age restriction to below the age of 45 years.

As the speakers were all urbanized men, there was no obvious evidence of any dialectal variations in their speech. However it is also important to note that the data were recorded as individual words and therefore dialectal differences, if present, would be difficult to detect.

Two methods of capturing the data were used. Initially, four of the eleven speakers were recorded on audio tape in a sound-proof studio at the University of South Africa. This data was then digitized using the Computerized Speech Lab (CSL) system[1] of the Phonetics Laboratory at the Research Unit for Experimental Phonology at the University of Stellenbosch (RUEPUS). Later seven additional speakers were recorded, this time directly onto the CSL in the same Phonetics Laboratory at the University of Stellenbosch. The data were recorded at a sampling frequency (F_s) of 10 kHz and with 16 bits resolution.

While recording, care was taken not to overload the system in order to avoid clipping the digitized signals. The input level was varied between speakers and therefore one can only compare the relative loudness levels of different speakers and not the absolute loudness levels.

2.3.3 Tagging

In order to calculate the prosodic features of phonemes, the boundaries between phonemes had to be tagged. Tags were therefore inserted to mark the start of every phoneme by means of the CSL system, which makes provision for annotating the speech signal in this way and facilitates duration measurement. These tags store the time instant of the tag as well as arbitrary ASCII text and therefore, by implication, simultaneously give rise to an analysis of duration.

---

[1] CSL model 4300B, software version 5.04 from Kay Elemetrics Corp.

Note that although the '/' symbol by convention denotes a phonemic transcription, in this document it is also used to indicate normal orthography.

The *Note* field of the CSL speech file format was used to store the orthographic transcription of the recorded speech. For example, if the word was /ngumfana/ the text 'Word=NGUMFANA' was stored in the Note field.

The tags were inserted manually in the CSL, while the speech signal was displayed in one window and the spectrogram in a separate and larger window (see Figure 2.1). LPC-derived formant tracks were superimposed on the spectrogram to aid in identifying phoneme boundaries. Tagging was done both visually, using the spectrogram and the signal, and auditorily, by listening to the marked speech segments. Each word for all eleven speakers was tagged in the same manner. The total number of tags for this corpus per speaker was 286.

NB

Single orthographic characters were used to represent the phonemes, except when a phoneme occurred more than once. In these cases the phonemes were numbered. The ']' character was used to indicate the end of the word. For example, the word /ngumfana/ had the following tags:

(2.2) n1 u m f a1 n2 a2]

Note that the [ŋg] is represented by a single character 'n'. This limitation is imposed by the software used. Furthermore only ASCII, and not IPA characters, could be used.
The tags for one speaker were checked thoroughly by hand. After all mistakes were corrected, the tags for the other 10 speakers were compared with those of the corrected speaker. The sequence of tags for a particular word was identical for all speakers. If a specific speaker did not pronounce a phoneme, the tag for that phoneme was still present, but located very close to the next tag. For example, speakers sometimes dropped the final vowel. Therefore the 'a2' tag was placed very close to the ']' tag so that:

(2.3) $t_{]} - t_{a2} \approx 0$

A list of the corpus with the labelled tags, together with the relevant identification of each example by using the noun class (A-M) and the example number (01-03 for each class), is given in Table A.1.

Figure 2.1 below shows the screen layout used for tagging the word /ngumfana/ with the CSL system. The signal and tag markers appear in the A window and the spectrogram in the B window.

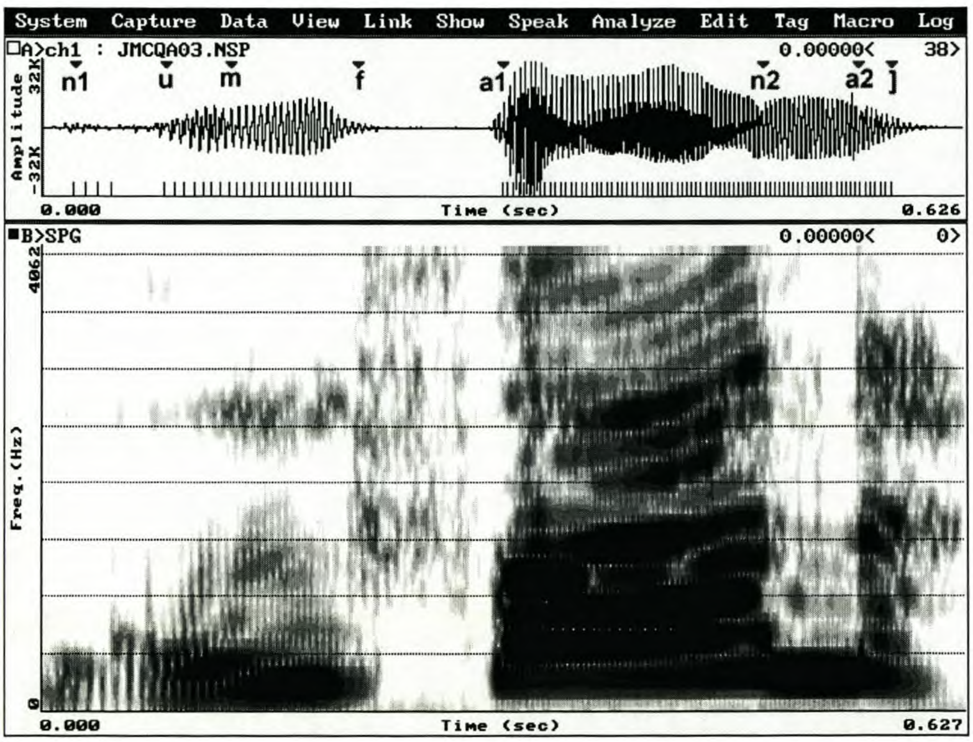


Figure 2.1 Screen layout used for tagging the word /ngumfana/ with the CSL system. The signal and tag markers appear in the A window and the spectrogram in the B window.

2.4 Statistical analysis

After the relevant measurements of duration, pitch and loudness had been taken, three types of descriptive statistical analyses were conducted in order to determine which features speakers use to distinguish between queclaratives and statements. These features should also be ranked in the order of significance for purposes of detecting the statistically most important. For every word, w, a number of features, f, were computed. For queclaratives the feature data vector Q_{wf} consisted of 11 samples, since there were 11 speakers. The corresponding feature data vector for statements was called S_{wf}.

As an example, consider the comparison between the duration of the queclarative and that of the statement form for the syllable /fa/ of the word /ngumfana/. The word w is /ngumfana/ and the feature f being compared is the duration of the syllable /fa/, i.e. d_{fa1}.

This notation is illustrated below.

(2.4) $Q_{\text{ngumfana } d_{fa1}}$ $S_{\text{ngumfana } d_{fa1}}$
w = ngumfana w = ngumfana
$f = d_{fa1}$ $f = d_{fa1}$

The following parameters were calculated for each data vector Q_{wf} and S_{wf} by applying a function $f(\,)$. The functions used are listed in 2.5 below.

(2.5) minimum
maximum
mean $\bar{\mu}$
median $\tilde{\mu}$
standard deviation σ

To illustrate the process the raw duration measurements for the Q_{wf} and S_{wf} vectors for the syllable /fa/ of the word /ngumfana/ are tabulated below. Thereafter follows Table 2.3 showing the functions that were applied and the numerical results.

Table 2.2 Raw duration measurements of the queclarative and statement forms for the syllable /fa/ of the word /ngumfana/.

Speaker	Queclarative: $Q_{\text{ngumfana } d_{fa1}}$	Statement: $S_{\text{ngumfana } d_{fa1}}$
A	325.8 ms	422 ms
D	245.4 ms	408.4 ms
F	221.3 ms	492.9 ms
J	300.3 ms	345.9 ms
M	276.3 ms	486.9 ms
N	310.8 ms	614.6 ms
P	328.6 ms	585.8 ms
S	250.6 ms	539.7 ms
T	363 ms	419.3 ms
X	264.5 ms	444 ms
Z	285 ms	349.9 ms

Table 2.3 Functions applied to the data vectors for the queclarative and statement forms of the syllable /fa/ of the word /ngumfana/.

Function	Queclarative: $f\left(Q_{\text{ngumfana d}_{\text{fa1}}}\right)$	Statement: $f\left(S_{\text{ngumfana d}_{\text{fa1}}}\right)$
minimum	221.3 ms	345.9 ms
maximum	363 ms	614.6 ms
mean $\overline{\mu}$	288.327 ms	464.491 ms
median $\widetilde{\mu}$	285 ms	444 ms
standard deviation σ	42.11	88.782

A statistical measure was required to determine whether Q_{wf} and S_{wf} originated from different distributions, or, in other words, whether Q_{wf} and S_{wf} were significantly different. If Q_{wf} and S_{wf} originated from the same distribution, that feature would not be significant in distinguishing between queclaratives and statements. In order to compare the two distributions a new data vector D_{wf} was defined as the difference between the queclarative and the statement data sets.

(2.6) $D_{wf} = Q_{wf} - S_{wf}$

Again using the word /ngumfana/, Table 2.4 below shows the vector $D_{\text{ngumfana d}_{\text{fa1}}}$.

Table 2.4 Differences in duration between the queclarative and statement forms for the syllable /fa/ of the word /ngumfana/.

Speaker	$D_{\text{ngumfana d}_{\text{fa1}}}$
A	-96.2 ms
D	-163 ms
F	-271.6 ms
J	-45.6 ms
M	-210.6 ms
N	-303.8 ms
P	-257.2 ms
S	-289.1 ms
T	-56.3 ms
X	-179.5 ms
Z	-64.9 ms

Parametric statistical methods, such as the normal distribution, are inappropriate for small sample sizes. In this data vector with queclarative/statement feature pairs of 11 samples each, two non-parametric methods were recommended, namely the *sign test* and the *Wilcoxon signed rank test*. These two tests and their application in finding the most significant features, are discussed and exemplified in §2.4.1 and §2.4.2. However, where the Wilcoxon signed rank test returned equal results or significance levels for different features, the mean divided by the standard deviation was used to resolve this problem and is discussed in §2.4.3.

The following parameters were calculated for each data vector D_{wf} by applying a function $f()$. The functions used are listed in 2.7 below with example values following in Table 2.5.

(2.7) mean $\bar{\mu}$
median $\tilde{\mu}$
standard deviation σ
$\frac{\text{mean}}{\text{standard deviation}} = \frac{\bar{\mu}}{\sigma}$
sign test
Wilcoxon signed rank test

Table 2.5 Functions applied to the difference data vector for the queclarative and statement forms of the syllable /fa/ of the word /ngumfana/.

Function	$f\left(D_{\text{ngumfana d}_{\text{fa1}}}\right)$
mean $\bar{\mu}$	-176.164 ms
median $\tilde{\mu}$	-179.5 ms
standard deviation σ	98.122 ms
$\frac{\text{mean}}{\text{standard deviation}} = \frac{\bar{\mu}}{\sigma}$	-1.79536
magnitude of $\frac{\bar{\mu}}{\sigma}$	1.79536
sign of $\frac{\bar{\mu}}{\sigma}$	-
sign test	1
Wilcoxon signed rank test	66

2.4.1 Sign test

The *sign test* is a non-parametric method, which tests the null hypothesis that the distribution of the differences between paired samples has a zero median value. The null hypothesis is rejected if the number of differences (i.e. Q_{wf} - S_{wf}) of one sign is too large or too small (Sachs, 1984:316). In our case the null hypothesis means that the samples are not sufficiently different to say that the specific feature is significant in discriminating between queclaratives and statements.

The procedure for the calculation of the sign test is outlined below:

1. Given two data sets Q_{wf} and S_{wf}, find the difference: $D_{wf} = Q_{wf} - S_{wf}$.
2. Find the number of positive samples n_p in D_{wf}.
3. Find the number of negative samples n_n in D_{wf}.
4. The effective sample size is $n = n_p + n_n$ since zero differences are ignored.
5. If $l \leq n_p \leq r$ then data sets Q_{wf} and S_{wf} do not differ that much and this feature is therefore *not* significant in distinguishing between queclaratives and statements. The null hypothesis is accepted. The lower and upper bounds l and r may be found in standard statistics textbooks. If $n_p < l$ or $n_p > r$ then data sets Q_{wf} and S_{wf} differ enough to say that this feature *is* significant in distinguishing between queclaratives and statements. The null hypothesis is rejected.

The sign test only indicates whether a feature is significant or not. It cannot rank features in order of significance. However the Wilcoxon signed rank test has this capacity.

2.4.2 Wilcoxon signed rank test

The *Wilcoxon signed rank test* is a non-parametric method, which can be used in a similar manner to the sign test to test the null hypothesis that the distribution of the differences between paired samples has a zero median value. However, the Wilcoxon signed rank test uses both the signs and magnitudes of the differences (Solomon, 1996:299). If the null hypothesis is rejected, it means in this application that the feature is significant. It is also possible to compute a magnitude of significance of S. The larger the value of S, the more significant a particular feature is.

The procedure for the calculation of the Wilcoxon signed rank test is outlined below:

1. Given two data sets Q_{wf} and S_{wf}, find the difference: $D_{wf} = Q_{wf} - S_{wf}$.
2. Assign a rank to each sample in D_{wf} by sorting the absolute values of D_{wf} in ascending order. The example below illustrates this step.

(2.8)

Index	1	2	3	4	5	6	7	8	9	10	11
D_{wf}	-0.7	0.2	-0.2	0.6	1.0	-0.9	0.3	-1.1	-0.1	1.3	0.4
$\|D_{wf}\|$	0.7	0.2	0.2	0.6	1.0	0.9	0.3	1.1	0.1	1.3	0.4
Rank	7	2	3	6	9	8	4	10	1	11	5
D_{wf} sign	–	+	–	+	+	–	+	–	–	+	+

(2.9)

Index	9	2	3	7	11	4	1	6	5	8	10
D_{wf} sorted by $\|D_{wf}\|$	0.1	0.2	0.2	0.3	0.4	0.6	0.7	0.9	1.0	1.1	1.3
Rank	1	2	3	4	5	6	7	8	9	10	11

3. Sum the ranks of the positive samples n_p in D_{wf}: $n_p = 2 + 6 + 9 + 4 + 11 + 5 = 37$.
4. Sum the ranks of the negative samples n_n in D_{wf}: $n_n = 7 + 3 + 8 + 10 + 1 = 29$.
5. T is the smaller value of n_p and n_n: $T = 29$.
6. The number of samples is eleven: $n = 11$.
7. If $T < t$ then the null hypothesis is rejected and feature D_{wf} is significant. Return a magnitude of significance S where S is the larger value of n_p and n_n: $S = 37$. The larger the value of S, the more significant feature D_{wf} is. The threshold t may be found in standard statistics textbooks.
8. If $T \geq t$ then the null hypothesis is accepted and feature D_{wf} is not significant. A zero significance value S is returned in this case: $S = 0$.

2.4.3 Mean and standard deviation

There are instances where the Wilcoxon test assigns equal significance values to different features. The magnitude of the mean divided by the standard deviation $\left|\frac{\overline{\mu}}{\sigma}\right|$ may be used to resolve which feature is more important. This value is larger for features that differ significantly. The sign of $\frac{\overline{\mu}}{\sigma}$ is also useful where positive values show that the numerical value of this feature is higher for queclaratives than for statements, while the negative implies the opposite.

2.4.4 Representation of significance results

The numeric results of the sign test, the Wilcoxon test and the $\overline{\mu}/\sigma$ were used to create symbolic tables of significance for the duration, pitch and loudness features. As an example Table 2.6 below shows the significance patterns that were derived from these three tests for the comparison of syllable duration for the word /ngumfana/.

Table 2.6 Parameters computed on the difference data vector for the queclaratives and statement forms of the syllable durations for the word /ngumfana/.

Parameter	$f\left(D_{\text{ngumfana}}\ d_{\text{ngu}}\right)$	$f\left(D_{\text{ngumfana}}\ d_{\text{m}}\right)$	$f\left(D_{\text{ngumfana}}\ d_{\text{fa}}\right)$	$f\left(D_{\text{ngumfana}}\ d_{\text{na}}\right)$	Significance
Sign test	0	0	1	0	ngu m $\mathbf{fa}_1$ na
Wilcoxon	0	61	66	0	ngu $\mathbf{m}_2$ $\mathbf{fa}_1$ na
$\bar{\mu}/\sigma$	-0.38187	-0.77402	-1.79536	-0.24778	ngu $\mathbf{m}_{2-}$ $\mathbf{fa}_{1-}$ na

Note that the $\bar{\mu}/\sigma$ yields only the relative importance of features. It does not indicate whether the calculated importance is in fact statistically significant. In this application a feature was taken to be insignificant if that was indicated by the Wilcoxon test (i.e. $S = 0$).

The conventions used in the creation and subsequent interpretation of these tables of significance is illustrated with an extract (Table 2.7) of the entire set of tables (as presented in Appendix A). Several aspects are illustrated in this example, namely:

1. The subscripts to the right of the syllables show the relative significance of those syllables, where '1' means most significant and '4' means least significant. Statistically significant syllables are printed in bold typeface, while insignificant syllables are printed in regular typeface and without subscripts.
2. The sign test does not indicate which syllable is more significant in the case of several syllables being significant.
3. The Wilcoxon test can rank syllables in order of significance. However, the test sometimes assigns equal significance values to different syllables. The $\bar{\mu}/\sigma$ may then be used to resolve the order of significance. The Wilcoxon test and the $\bar{\mu}/\sigma$ differed in the order of significance in a few instances. These results will be discussed in more detail under the relevant headings below.
4. The '+' or '−' sign following the subscripts in the $\bar{\mu}/\sigma$ column *does not* indicate that the significance is positive or negative. Rather, it shows that the *difference* is positive or negative.

Table 2.7 Significant difference between syllables based on statistical analyses of duration, pitch and loudness features for the word /ngumfana/.

Feature	Sign test	Wilcoxon test	$\bar{\mu}/\sigma$
Duration	ngu m $\mathbf{fa}_1$ na	ngu $\mathbf{m}_2$ $\mathbf{fa}_1$ na	ngu $\mathbf{m}_{2-}$ $\mathbf{fa}_{1-}$ na
Pitch	ng$\mathbf{u}_1$ $\mathbf{m}_1$ f$\mathbf{a}_1$ na	ng$\mathbf{u}_1$ $\mathbf{m}_1$ f$\mathbf{a}_1$ n$\mathbf{a}_2$	ng$\mathbf{u}_{3+}$ $\mathbf{m}_{2+}$ f$\mathbf{a}_{1+}$ n$\mathbf{a}_{4+}$
Relative loudness	ngu m fa na	$\mathbf{ng}_2$u m $\mathbf{f}_1$a na	$\mathbf{ng}_{2+}$u m $\mathbf{f}_{1-}$a na
Absolute loudness	ng$\mathbf{u}_1$ $\mathbf{m}_1$ f$\mathbf{a}_1$ $\mathbf{n}_1$a	ng$\mathbf{u}_1$ $\mathbf{m}_2$ f$\mathbf{a}_3$ $\mathbf{n}_3\mathbf{a}_4$	ng$\mathbf{u}_{2+}$ $\mathbf{m}_{1+}$ f$\mathbf{a}_{4+}$ $\mathbf{n}_{3+}\mathbf{a}_{5+}$

The tables of significance of features provided a compact representation of the large amount of numeric data and lent themselves towards the extraction of local as well as global patterns. To illustrate, Table 2.8 below shows the observations that were made about the duration, pitch and loudness differences for only one phoneme, namely the syllabic /m/ of the utterance /ngumfana/.

Table 2.8 Observations made for the syllabic /m/ of the utterance /ngumfana/.

Feature	Observation	Conclusions
Duration	$\mathbf{m}_{2-}$	• duration of queclarative /m/ < duration of statement /m/ • second most significant syllable duration feature
Pitch	$\mathbf{m}_{2+}$	• pitch of queclarative /m/ > pitch of statement /m/ • second most significant phoneme pitch feature
Rel. loudness	m	• relative loudness is not significant for the /m/ phoneme
Abs. loudness	$\mathbf{m}_{1+}$	• absolute loudness of queclarative /m/ > absolute loudness of statement /m/ • most significant absolute loudness feature

The notion a pattern search was adopted to further extract trends from the available data. A *pattern* was defined as a search string against which observations could be tested. In this case the capital 'S' represents any syllable. As mentioned earlier the subscripts to the right of the 'S' show the relative significance of those syllables where '1' means most significant and '4' means least significant. If the actual value of the subscript is not required, that subscript is replaced by a variable denoted as a lowercase letter (e.g. *a*, *b*, *c*). The permitted ranges of the variables will be specified. Table 2.9 below illustrates these concepts.

Table 2.9 Examples of defined patterns.

Pattern	Observation
S	Any statistically insignificant syllable
$\mathbf{S}_2$	Any syllable with a statistical order of significance of 2
$\mathbf{S}_a$ ($a > 2$)	Any syllable with a statistical order of significance greater than 2 $\mathbf{S}_3$ and $\mathbf{S}_4$ are possible instances
$\mathbf{S}_a$ ($a \neq 1$)	Any syllable with a statistical order of significance not equal to 1 S (= $\mathbf{S}_0$), $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$ are possible instances

For example, the pattern $\mathbf{S}_a\ \mathbf{S}_1\ \mathbf{S}_b$ ($a,b\neq1$) could be any one of the following:

(2.10)	S $\mathbf{S}_1$ S	$a = 0$	$b = 0$
	S $\mathbf{S}_1$ $\mathbf{S}_2$	$a = 0$	$b = 2$
	$\mathbf{S}_2$ $\mathbf{S}_1$ S	$a = 2$	$b = 0$
	$\mathbf{S}_2$ $\mathbf{S}_1$ $\mathbf{S}_3$	$a = 2$	$b = 3$
	$\mathbf{S}_3$ $\mathbf{S}_1$ $\mathbf{S}_2$	$a = 3$	$b = 2$

The pattern $\mathbf{S}_a\ \mathbf{S}_1\ \mathbf{S}_b$ exemplified above includes all three syllable words with the penultimate syllable statistically most significant. Note that the actual significance of the first and last syllables are not considered, hence the use of the variables a and b for these syllables. The following properties of the variables are evident and will not be indicated explicitly:

1. Variables cannot have equal non-zero values ($a\neq b$).
2. The minimum value is zero ($a\geq0$ and $b\geq0$).
3. The maximum value is the number of syllables ($a\leq3$ and $b\leq3$).
4. Values cannot be skipped (if a=0 then $b\leq1$).
5. $a,b\neq1$ means $a\neq1$ and $b\neq1$.

2.5 Duration

2.5.1 Introduction

Duration has generally been one of the parameters nominated as being significant in the differentiation between statements and questions. It has also been noted as a factor used in the disambiguation of 'incredulous' as opposed to 'uncertain' utterances (Hirschberg and Ward, 1992) and also 'rhetorical' and 'literal' questions (Miura and Hara, 1995). As indicated in the introduction a number of acoustic analyses (Cooper, 1976; Klatt, 1976; Lehiste, Olive and Streeter, 1976; Lehiste, 1980; Garro and Parker, 1982 and 1983) and perceptual experiments have been done on English in an attempt to identify the factors responsible for differentiating between syntactically ambiguous sentences. These studies, both acoustic and perceptual, have indeed shown that segmental duration is one of the significant factors influencing the disambiguation process in English.

Hirschberg and Ward (1992), while researching the interpretation of the rise-fall-rise intonation contour in English using the utterance 'Eleven in the morning', found that there were indeed differences between productions in duration in the utterance conveying *uncertainty* (longer) than the same phrase used to convey *incredulity*. Research showed that the uncertainty reading was somewhat longer than the same phrase uttered to convey incredulity. Such incredulous utterances may be compared to the copulative queclarative investigated here, which showed shorter segmental duration than its statement counterpart. This once again emphasizes the relevance of this parameter in the disambiguation process. Miura and Hara (1995) in their research on Osaka Japanese, using the utterances 'you eat peaches' and 'you eat cucumbers' as literal and rhetorical questions, both of which were devoid of any linguistic markers, observed that there were also acoustic differences in duration between these two linguistic forms. In summary, rhetorical questions show longer segmental duration than literal questions. Therefore segmental duration differentiates between these two linguistic forms of interrogative.

Jun and Oh (1996:52), in their prosodic analysis of three types of /wh/ phrases in Korean, found that incredulity questions were produced with similar duration. In contrast, English incredulity sentences are significantly shorter than their non-surprised versions (Ward and Hirschberg, 1988; Hirschberg and Ward, 1992).

Within Bantu languages, impressionistic claims on duration as a contributory factor to disambiguation have been made by Steyn concerning Northern Sotho (1991). Steyn acknowledges briefly that a change in suprasegmental features such as tone and length play a role in distinguishing between a declarative and a question (which are identical in form). Nkabinde (1999), in his research on Zulu claims that there is a quick tempo for questions and unchecked length on the penultimate syllable of the only or last word in a sentence for statements.

Doke and Mofokeng (1957:433-434), in commenting on interrogative constructions in Southern Sotho, state that

> The important point is that there is a change in length of the syllables of the last word, especially the penult of the sentence, which is very short, while the final

> syllable is clearly clipped. The effect on the ear is that a question sounds like a statement quickly spoken.

Lombard (1980:45-46) analyzed the occurrence and function of length in 9 types of Northern Sotho sentences, including interrogatives. His findings were that the penultimate syllable of declaratives may be classified as long syllables; in the case of interrogatives, however, the penultimate syllable is categorized as half-long or a short syllable. Ziervogel (1976:123), in his study of Northern Sotho, also pointed out that interrogative sentences lose length on the penultimate syllable.

In Xhosa, impressionistic claims relating to duration include those made by Lanham (1963). He regards the duration of the penultimate vowel as being important and states that a relatively long penultimate vowel is typical of questions, whereas an extra-long vowel signalizes a statement. Riordan (1969) is in agreement with these views on duration, as is Louw (1968).

After his experimental work on three queclarative examples, Louw (1968:88) states that

> The tone sequences stay basically the same in statements and questions, but in the last, the relative key in which all these tones are spoken is raised very high and the long length which occur [sic] as boundary marker in the statements, becomes much shorter...

Theron (1991), after her experimental study in Xhosa on this subject, returned results that make it very difficult to support impressionistic data that penultimate vowel shortening is significant or perceptually relevant in differentiating statements from questions. She showed, in fact, that for one speaker, a longer penultimate syllable for questions was produced in 33.4% of instances and for the second speaker, in 41.67% of instances. Theron (1991:59) did, however, indicate that there was a distinct shortening of sentence duration in questions as opposed to statements within the corpus of data analyzed. Statements on duration implicitly effect the *rate* of the utterance production, which, in turn, is directly related to *tempo*. Khumalo, in his impressionistic claims differentiating between statements and questions in Zulu, says of tempo:

Statement intonation is unmarked for tempo (i.e. it is at "average" speaking tempo), while question intonation is at a much faster tempo (1981:92).

Theron (1991:61), in her experimental work on Xhosa, found evidence of some variation between speakers with regard to tempo, although both informants produced statements at a slower rate of articulation as compared with corresponding questions.

These impressionistic claims made by scholars relating to duration and tempo and the experimental work done in Xhosa by Theron (1991), Louw (1968) and Roux (n.d.) appear to support the notions that questions are indeed of a shorter duration than statements and that they are, therefore, articulated at a quicker tempo.

This study revisits these claims on duration as a parameter contributing to the disambiguation process, both from an experimental and perceptual perspective and the procedure followed in the experiment and the results of the statistical analyses are now presented under the relevant headings.

2.5.2 Calculation of duration

The tag positions, which were inserted in preparation for analysis of the corpus, mark the boundaries of the phonemes and the utterance. Software[2] was then developed, using the parameters below in order to determine the duration of phonemes and syllables to extract these specific calculations. Duration measurements taken for the word /ngumfana/ are used in the text below to illustrate the concepts.

Table 2.10 Position of tags inserted for the queclarative form of the word /ngumfana/ for one speaker.

Phoneme	/ng/	/u/	/m/	/f/	/a/	/n/	/a/	
Tag	t_{n1}	t_u	t_m	t_f	t_{a1}	t_{n2}	t_{a2}	t_l
Position [ms]	21.5	83.1	127.3	214.6	313.2	490.9	556.5	579.1

The following measurements were taken for each word in the corpus:

[2] Developed by Mr J.A.N. Louw, RUEPUS, 1997.

(i) The duration of the whole word, which is the time difference between the first tag and the ']' tag. The duration of /ngumfana/ for example would be:

(2.11) time(']') - time('n1')

written as follows:

(2.12) $d_{ngumfana} = t_{]} - t_{n1} = 579.1$ ms - 21.5 ms = 557.6 ms

(ii) The duration of a phoneme (e.g. 'a1'), which is the time difference between the tag for that phoneme and the next tag:

(2.13) $d_{a1} = t_{n2} - t_{a1} = 490.9$ ms - 313.2 ms = 177.7 ms

Table 2.11 Phoneme duration values calculated from tags for the queclarative form of the word /ngumfana/ for one speaker.

Phoneme	/ng/	/u/	/m/	/f/	/a/	/n/	/a/
Duration	d_{n1}	d_u	d_m	d_f	d_{a1}	d_{n2}	d_{a2}
Value [ms]	61.6	44.2	87.3	98.6	177.7	65.6	22.6

(iii) The duration of the /fa/ syllable would then be given as:

(2.14) $d_{fa1} = t_{n2} - t_f = 490.9$ ms - 214.6 ms = 276.3 ms

Table 2.12 Syllable duration values calculated from tags for the queclarative form of the word /ngumfana/ for one speaker.

Syllable	/ngu/	/m/	/fa/	/na/
Duration	d_{n1u}	d_m	d_{fa1}	D_{n2a2}
Value	105.8 ms	87.3 ms	276.3 ms	88.2 ms

The following set of duration features were calculated for the word /ngumfana/ for example:

(2.15) Phonemes: d_{n1} d_u d_m d_f d_{a1} d_{n2} d_{a2}
Syllables: d_{n1u} d_m d_{fa1} d_{n2a2}

The total duration of the word ($d_{ngumfana}$) was also calculated. All duration measurements are in seconds (small values are shown in milliseconds). These feature sets were calculated for all speakers and all queclaratives and statements, and they correspond to the Q_{wf} and S_{wf} feature sets respectively (e.g. $d_uQ \in Q$, $d_uS \in S$, $d_uQ - d_uS \in D$).

(iv) The word tempo was calculated as:

$$\text{(2.16)} \quad \text{tempo} = \frac{\text{number of syllables}}{\text{total duration}} = \frac{4}{0.5576} = 7.17 \text{ syllables/s}$$

Tables 2.13 and 2.14 below are provided as an example of the duration features and values calculated for the word /ngumfana/. These values were calculated for each phoneme, syllable and word for each analyzed parameter for the entire corpus of data used in this research project.

Table 2.13 Mean phoneme duration and mean total duration for the utterance /ngumfana/ [in ms].

	d_{n1}	d_u	d_m	d_f	d_{a1}	d_{n2}	d_{a2}	$d_{ngumfana}$
Queclarative	44.6	57.1	107.2	105.6	182.7	79.2	66.9	643.5
Statement	61.2	61.7	120.6	131.1	333.4	86.7	66.9	861.7

Table 2.14 Mean tempo for the utterance /ngumfana/.

	Tempo
Queclarative	6.22 syllables/s
Statement	4.64 syllables/s

2.5.3 Results

The results of the statistical analysis on duration are presented next. The amount of numerical data obtained for the entire corpus was quite substantial. Therefore only the mean (average) values of the phoneme duration, word duration and word tempo are given in Tables A.2 and A.3 in Appendix A. Patterns of significance were derived from Table A.4.

Table 2.15 Average duration measurements over all words.

Parameter	Queclaratives	Statements	Percentage change
Average word duration	667.6 ms	895.8 ms	$\frac{895.8 - 667.6}{667.6} 100\% = 34.2\%$
Average word tempo	4.89 syllables/s	3.65 syllables/s	$\frac{4.89 - 3.65}{3.65} 100\% = 34.0\%$

Table 2.16 Patterns observed from the Wilcoxon and $\overline{\mu}/\sigma$ analyses of syllable duration

features (derived from Table A.4).

Observation	Pattern	Occurrences
Penultimate syllable duration is the *most* significant duration feature while other syllables are of *secondary* importance.	$\mathbf{S}_1\ \mathbf{S}_2$	1 / 1 (100%)
	$\mathbf{S}_a\ \mathbf{S}_1\ \mathbf{S}_b$ $(a,b\neq1)$ $(a>0$ or $b>0)$ $(a,b\leq3)$	13 / 27 (48.2%)
	S $\mathbf{S}_a\ \mathbf{S}_1\ \mathbf{S}_b$ $(a,b\neq1)$ $(a>0$ or $b>0)$ $(a,b\leq3)$	9 / 11 (81.8%)
		Total: 23 / 39 (59%)
Penultimate syllable is the *only* significant duration feature.	$\mathbf{S}_1$ S	0 / 1 (0%)
	S $\mathbf{S}_1$ S	11 / 27 (40.7%)
	S S $\mathbf{S}_1$ S	2 / 11 (18.2%)
		Total: 13 / 39 (33.3%)
Penultimate syllable duration is the *most* significant duration feature.	$\mathbf{S}_1\ \mathbf{S}_2$	1+0=1 / 1 (100%)
	$\mathbf{S}_a\ \mathbf{S}_1\ \mathbf{S}_b$ $(a,b\neq1)$ $(0\leq a,b\leq3)$	13+11=24 / 27 (88.9%)
	S $\mathbf{S}_a\ \mathbf{S}_1\ \mathbf{S}_b$ $(a,b\neq1)$ $(0\leq a,b\leq3)$	9+2=11 / 11 (100%)
		Total: 36 / 39 (92.3%)
Penultimate syllable duration is *not* the most significant duration feature.	$\mathbf{S}_a\ \mathbf{S}_b$ $(a\neq1)$ $(0\leq a,b\leq2)$	0 / 1 (0%)
	$\mathbf{S}_a\ \mathbf{S}_b\ \mathbf{S}_c$ $(b\neq1)$ $(0\leq a,b,c\leq3)$	3 / 27 (11.1%)
	S $\mathbf{S}_a\ \mathbf{S}_b\ \mathbf{S}_c$ $(b\neq1)$ $(0\leq a,b,c\leq3)$	0 / 11 (0%)
		Total: 3 / 39 (11.1%)
Other patterns on these words	S $\mathbf{S}_a\ \mathbf{S}_b$ $(0\leq a,b\leq2)$	18 / 27 (66.7%)
	S $\mathbf{S}_1\ \mathbf{S}_2$	6 / 27 (22.2%)
	$\mathbf{S}_3\ \mathbf{S}_1\ \mathbf{S}_2$	5 / 27 (18.5%)
	S $\mathbf{S}_2\ \mathbf{S}_1$ S	7 / 11 (63.6%)

The interpretations of the results of the analysis of duration of both the utterance and the syllables in this corpus of Xhosa copulative queclarative / statement pairs are given below:

1. The **duration** of the word was significantly longer for all the statements than for all the queclaratives. The statements were on average 34.2% longer than the queclaratives (Table 2.15).
2. The **tempo** in syllables per second was on average 34.0% higher for queclaratives than for statements (Table 2.15).
3. Every syllable of every queclarative was found to be **shorter** than the corresponding syllable of the statement according to the negative sign of the $\overline{\mu}/\sigma$, provided that the difference was statistically significant (Table A.4).
4. The **penultimate syllable** was longer for statements than for queclaratives. For 92.3%

of the words the penultimate syllable duration was the *most* significant syllable duration feature according to the Wilcoxon signed rank test in distinguishing between queclaratives and statements (Table 2.16).

5. The duration of the penultimate syllable was the **only** significant syllable duration feature for 33.3% of the words (Table 2.16).
6. For 59.0% of the words the duration on the penultimate syllable was the most significant syllable duration feature, while other syllables were found to be of **secondary** importance (Table 2.16).
7. In the 27 **three-syllable** words, the duration of the penultimate syllable was the most significant syllable duration feature used for disambiguation in 88.9% of the cases. For the 11 **four-syllable** words the penultimate syllable was always the most significant (Table 2.16).
8. For three of the words (7.7%) the penultimate syllable duration was **not** the most significant syllable duration feature, although it was rated as the second most important feature. Two of these words have the syllabic /m/ as the penultimate syllable, namely '**ngu**$_1$ **m**$_2$ ntu' and '**ngu**$_3$ **m**$_2$ **thi**$_1$' (Table A.4).
9. The most significant syllable as determined by the **sign test** corresponded with that identified by the **Wilcoxon test** and the $\overline{\mu}/\sigma$ test in 92.3% of the cases. However, the sign test tends to show fewer significant syllables than the Wilcoxon test (Table A.4).
10. The order of significance obtained from the **Wilcoxon test** and the $\overline{\mu}/\sigma$ was identical for 82.1% of the words. The other 17.9% consisted of five words (12.8%) where the Wilcoxon test returned equal significance values and two words (5.1%) where the order of significance differed (Table A.4).

2.5.4 Discussion

The introduction to the topic of duration includes examples where duration is said to have contributed to syntactic disambiguation. For example Hirschberg and Ward (1988 and 1992) demonstrated this for *incredulity* as opposed to *uncertainty* readings. Likewise Miura and Hara (1995) have similar findings for Osaka Japanese *rhetorical* as opposed to *literal* questions.

Within Bantu languages there are numerous examples where scholars claimed that duration was a factor used in the differentiation process between statements and questions. These included scholars of Northern Sotho such as Steyn (1991), Lombard (1980) and Ziervogel (1976); for Zulu Nkabinde (1999) and Khumalo (1981); for Southern Sotho Doke and Mofokeng (1957) and for Xhosa Lanham (1963), Riordan (1969). Louw (1968) and Theron (1991) reached similar findings after their experimental contribution to the subject.

From these results it is evident that acoustically the duration both of the syllable and the utterance as a unit played an extremely important role in differentiating between statements and queclaratives in this set of Xhosa data. The duration of every word and that of every syllable was longer for statements than for queclaratives.

As the duration was found to be reduced for queclaratives, by implication therefore the tempo in syllables per second was also higher for queclaratives. Furthermore, the statistical results showed that the duration of the penultimate syllable was the most significant syllable duration feature for 92.3% of the words, although other syllables were found to play a secondary role in disambiguation in 59.0% of the cases.

In view of these results after this acoustic analysis of Xhosa data, it is apparent that in Xhosa duration is also used as a factor which contributes to the disambiguation process and in this instance specifically between statements and queclarative forms. Therefore Xhosa seems to use the same cues as other languages. These findings also show the concurrence with impressionistic claims relating to duration for all the Bantu languages, as cited above.

2.6 Pitch

2.6.1 Introduction

The use of intonation as a means of differentiating between statements and questions has been acknowledged by many prominent scholars. Many of the claims made about the Bantu languages have, however, once again, been purely impressionistic in nature and have no substantiation in terms of scientific experimental evidence or perceptual testing of data.

This however is not the case for other languages, where a vast number of acoustic and perceptual analyses have been conducted, resulting in validated quantitative research.

Wales and Taylor (1987:199), in their experimental study on the perception of intonation cues to questions and statements for Australian English, say:

> In the absence of mitigating syntactic or pragmatic cues, the intonational cues most commonly assumed to be used in perceptually differentiating spoken questions from statements consist of a final rise in fundamental frequency (F0) for yes/no question, and a declination for statements (Allan 1984; Brown, Currie and Kenworthy, 1980; Cooper and Sorenson, 1981 Guy and Vonwiller, 1984; Lieberman, 1967).

In their conclusions to this research project, Wales and Taylor (1987:209) state that:

> The results suggest that coupled or sequential F0 sweeps serve as important cues to intonation-based question/statement differentiation, with other cues playing less important roles.

According to Ohala (1983:1) it is common that

> There are probably few sound-meaning correlates that are more strikingly similar across languages and cultures than the association between certain pitch (fundamental frequency) contours and the linguistic categories STATEMENT and QUESTION [Hermann 1942; Bolinger, 1964, 1978; Ultan 1969; Cruttenden,1981]. When the same utterance can be produced as a question or statement using only intonation, i.e., without any lexical or syntactic markers, it is almost invariably the case that high or rising pitch signals the former, whereas low or falling pitch the latter.

Crystal (1985:162) points out that intonation performs several functions in language. One is to highlight the contrast between grammatical structures such as statements and questions. Ultan (1978:219), in his publication on research of interrogative systems of 79 randomly chosen languages, makes these comments on the contribution of intonation:

> Furthermore, although Q-intonation is often accompanied by some other Q-marker

> (e.g. inversion, QP, etc.), most languages (perhaps all ?) also have Q-utterances distinguished from their corresponding declarative utterances solely by means of Q-intonation.

Furthermore, in this study Ultan (1978) found that while there certainly was evidence of a terminal rising contour, Q-intonation and a terminal rising contour were not synonymous. Additional contours were also present; however, the questions all shared one feature:

> ...higher pitch (register or glide) or more prominent stress at some point in the contour, usually towards the end, as opposed to falling or low- pitched ultima in simple declarative statements (1978:220).

Studdert-Kennedy and Hadding (1973:312) in their research (which is both experimentally and perceptually based) on intonation contours using the American English utterance 'November', said that

> Questions tended to have a terminal rise, usually from level 2, or a fall ending comparatively high. Questions were also generally spoken with an overall high F0 compared to statements, a phenomenon that, according to the literature, occurs in many languages (Hermann, 1942; Bolinger, 1964).

Shen, in her study of the prosody of Mandarin Chinese (1990:10), quotes from De Francis, who claims that what characterizes interrogative intonation in Mandarin Chinese is that the whole pitch level is shifted slightly upwards and, therefore, the entire sentence and not simply part of it is the disambiguating factor. This claim was made, however, after auditory impressions. The results of instrumental analyses carried out by Ho (1977) and J. Shen (1985) corroborate De Francis's impressionistic claim.

According to Shen (1990:25), after her experimental analysis:

> However, at the starting point the F0 values of all questions are significantly higher than those of statements, regardless of the register at the ending point. Even the three categories of questions which have low endings also have higher values at the

> starting point: Interrogative intonation begins at a higher register, usually with a mid-high key. My conclusion is consistent with those of De Francis (1963), Ho (1977), and J. Shen (1985).

Shen (1990:28) quotes from Dwight Bolinger after personal communication with him on the topic. In summary, he is of the opinion that if a large enough sample were collected and measured for overall pitch level, English would turn out to be the same as Mandarin Chinese, where questions as a class were signalled more frequently by higher pitch throughout than by terminals. He stated that there are certain classes of question, i.e. reprise and complementary, which nearly always have the rise; but other subclasses do not, and it is generally the pitch which is used as a signalizer.

Eady and Cooper (1986:409), in summarizing their findings on questions stated that questions generally have higher F0 values than the corresponding statement versions, with a particularly high value at the end of the sentence. This observation seems to be in agreement with previous authors' descriptions of F0 patterns in yes/no questions (Lieberman, 1967; O'Shaughnessy, 1979). In fact, they state that the major difference in the F0 toplines of questions and statements is that the former have a much higher F0 value on the last word of a sentence. However, it is true to say that this notion of intonation signifying queclaratives or question status does not have total support. Geluykens (1987:492), having experimentally researched queclaratives, states that

> If pragmatic queues (sic) fail to make the utterance question-prone, rising intonation may, but need not, be used to turn such an utterance into a queclarative.

Similarly Fries (1964) in his investigation of 2,561 yes/no questions in English , found that in percentage terms 61.7% of these questions had a falling intonation pattern and only 38.3% a rising intonation pattern. Fries (1964:250) remarks that

> The quantitative information derived from the corpus examined for this study does not support the much repeated assertion that yes/no questions 'regularly', 'usually', 'characteristically' have a rising intonation pattern.

Contributions to the topic of queclaratives in the Bantu languages are limited and those which are available are predominantly impressionistic in nature. These impressionistic descriptions incorporate views such as those expressed by Khumalo, who states (1981:92) that

> In statement intonation the tone of the final syllable of the phonological phrase is lowered. ...this lowered final tone does not occur in question intonation.

Other scholars who have commented impressionistically on the use of intonation as a disambiguating factor in the Bantu languages include Cole (1955) for Zulu, Snyman et al. (1991) for Tswana, Steyn and Louwrens (1991) for Northern Sotho, and Poulos (1990) for Venda.

According to Snyman et al. (1991), a statement is converted into an interrogative by raising the tone at the end of a sentence. Khoali (1994), writing on Tswana, says that intonation is also used to mark yes/no questions in Setswana and that there is no reordering of words when intonation is used. Doke and Mofokeng (1957:433-434) say that for Southern Sotho:

> The tone sequences of the various syllables are the same in a question as in a statement. But in order to convert a statement into a question a higher tone register is generally used for the whole sentence; though this need not always be the case.

Experimental research on intonation and 'other' strategies that contribute to the process of encoding questioning within Bantu languages is limited and confined to contributions on Xhosa by Louw (1968), Theron (1991) , Roux (n.d.) and on Chichewa by Myers (1999).

Theron (1991:69,80) in her rather more extensive experimental work on statements and questions (involving the production of 60 statement/question pairs by one male and one female speaker of Xhosa), also lends support to the view that statements are characterized by tonal down-drift, while also identifying a measure of this feature in echo-questions as well.

Myers (1999:1984) concluded in experimental research on Chichewa that

> (..) Down-drift is less in questions than in statements.
> (..) The range of F0 values for high tones is higher in questions than in statements
> (..) The range of F0 values for high tones is higher in louder speech than in softer speech.

There are countless findings some cited above, which are both impressionistic and experimentally based, on the use of pitch and intonation as disambiguating factors within many languages. The comments and claims relating to Bantu languages, however, are predominantly impressionistic in nature, except for those cited above. These claims include the use of intonation over the entire utterance (register), while most propose that the intonation on the utterance (whether final or penultimate syllable) is the disambiguating factor. These claims will be considered after the acoustic analysis of this corpus of data.

2.6.2 Calculation of pitch

The CSL offers two pitch calculation algorithms. Both operate on the speech signal in the time domain. The *Pitch* command finds the average pitch in a frame by locating positive and negative peaks, while the *Impulse* command sequentially locates individual pitch periods in the signal.

The CSL *Pitch* command produces fairly accurate pitch estimates in the steady-state parts of voiced speech, but performance degenerates during voiced/unvoiced transitions and in low-energy regions. Another major disadvantage is that pitch estimation errors cannot be corrected by hand. Using the CSL *Impulse* command one can edit pitch markers, but this is a time-consuming and error-prone task since the initial pitch estimates are often erroneous or noisy.

For the reasons stated above, the pitch was calculated outside the CSL in a general purpose mathematical package called MATLAB. This pitch tracking algorithm[3] adaptively low-pass filters the speech signal to emphasize pitch periods. Pitch period markers are then

---

[3] Developed by Mr J.A.N. Louw, RUEPUS, 1997.

inserted at the negative-to-positive zero crossings of the filtered signal. Some refinement is still necessary to make the voiced/unvoiced decision and to minimize pitch doubling errors. At this time these errors were corrected manually in MATLAB by deleting pitch markers. The algorithm is able to track rapid changes in pitch accurately.

The procedure for the calculation of the pitch is outlined below:

1. The autocorrelation method is used to calculate a rough pitch estimate for non-overlapping blocks of 50 ms.
2. Calculate the coefficients for second order bandpass filters centered around the pitch estimates from step 1.
3. Adaptively filter the speech signal with the filters from step 2. This yields a low pass filtered signal that emphasizes the pitch periods.
4. Pitch period markers are then inserted at the negative-to-positive zero crossings of the filtered signal.

The pitch markers were calculated for all words and saved with the speech signals. After pitch errors were corrected the data were saved on CD-ROM. This comprised the final data set used for further analysis.

A representation of the layout of the screen for the editing of pitch markers in MATLAB is represented in Figure 2.2 below.

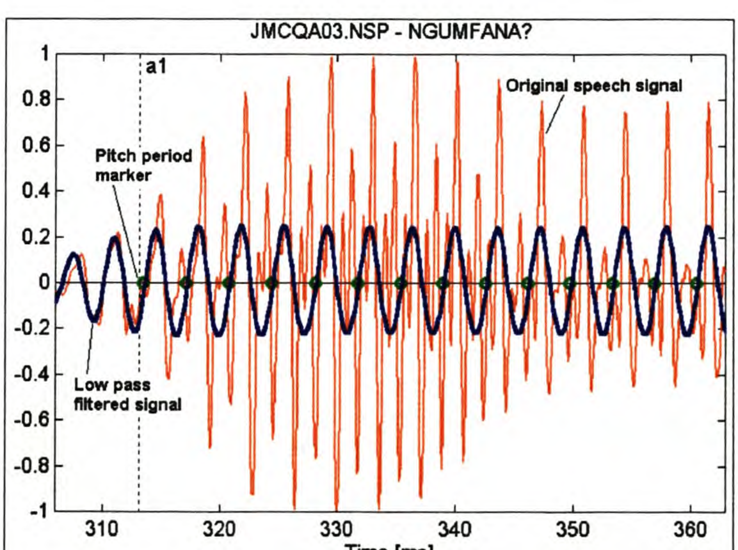


Figure 2.2 Screen layout for editing pitch markers in MATLAB.

The larger signal (red) in Figure 2.2 is the original speech signal, while the smaller signal (blue) is the low-pass filtered version thereof. Pitch period markers are shown as circles (green) on the horizontal axis. The user can manually remove one or several pitch markers at a time. The resulting pitch curve is displayed in another window (Figure 2.3).

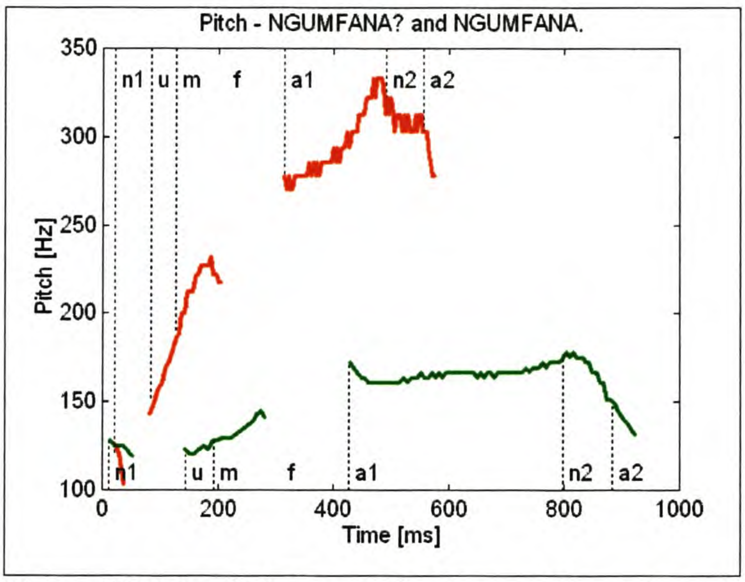


Figure 2.3 The pitch contour for the word /ngumfana/ for the queclarative (red) and the statement (green) forms for one speaker as displayed in MATLAB.

Pitch was calculated only on the vowels and the syllabic /m/ (such as the /m/ in /ngumfana/). In order to simplify the analysis, it was decided to evaluate only the pitch on the vowel of the typical CV syllable structure and not on the consonant part. Each vowel (or syllabic /m/), located by means of the tags inserted during the tagging process in preparation for analysis, yielded several pitch measurements. These measurements were first smoothed with a third-order median filter. The average was then computed and taken as the pitch feature for that phoneme. For example, the following pitch features were calculated for the word /ngumfana/:

(2.17) p_u p_m p_{a1} p_{a2}

The average pitch of the word ($p_{ngumfana}$) was also calculated. All pitch measurements are in Hz. Table 2.17 provides as an example of the pitch features and values calculated for the word /ngumfana/. These calculations were, however, done for each analyzed parameter for the entire corpus of data used in this research project.

Table 2.17 Median vowel (and syllabic /m/) pitch and median average pitch for the utterance /ngumfana/.

	P_u	p_m	p_{a1}	p_{a2}	$p_{ngumfana}$
Queclarative	140.9 Hz	170.5 Hz	257.6 Hz	198.3 Hz	229.7 Hz
Statement	116.5 Hz	135 Hz	158.8 Hz	140.1 Hz	148.6 Hz

To assist in the visual realization of the shape of the pitch contour, it was necessary to calculate the *sign* of the *slope* between successive pitch measurements. In this context a positive sign means that the median pitch of a particular vowel is lower than that of the following vowel. A negative slope implies that the median pitch is higher on the preceding vowel. For example, the following pitch slope signs were calculated for the question form of the utterance /ngumfana/:

(2.18) + + –

From these the following conclusions can be drawn:

(2.19) $p_u < p_m$ is implied by the first + sign
$p_m < p_{a1}$ is implied by the second + sign
$p_{a1} > p_{a2}$ is implied by the – sign
$p_u < p_m < p_{a1}$ is implied by the first two + signs

Note that the slope signs are used to compare two *adjacent* vowels. Only when two or more signs are identical can the comparison be extended to vowels further than the immediate predecessor or successor of a particular vowel. This was indeed the case in the fourth conclusion in 2.19. Table 2.18 below provides the pitch slope signs for the utterance /ngumfana/ as a queclarative and statement.

Table 2.18 Pitch slope signs for the utterance /ngumfana/.

	Pitch slope signs
Queclarative	+ + –
Statement	+ + –

In Figure 2.4 below the median pitch values for the vowels and the syllabic /m/ are shown as circles (blue) superimposed on the actual pitch contour (red). The pitch slopes appear as straight lines (green) connecting the median pitch values.

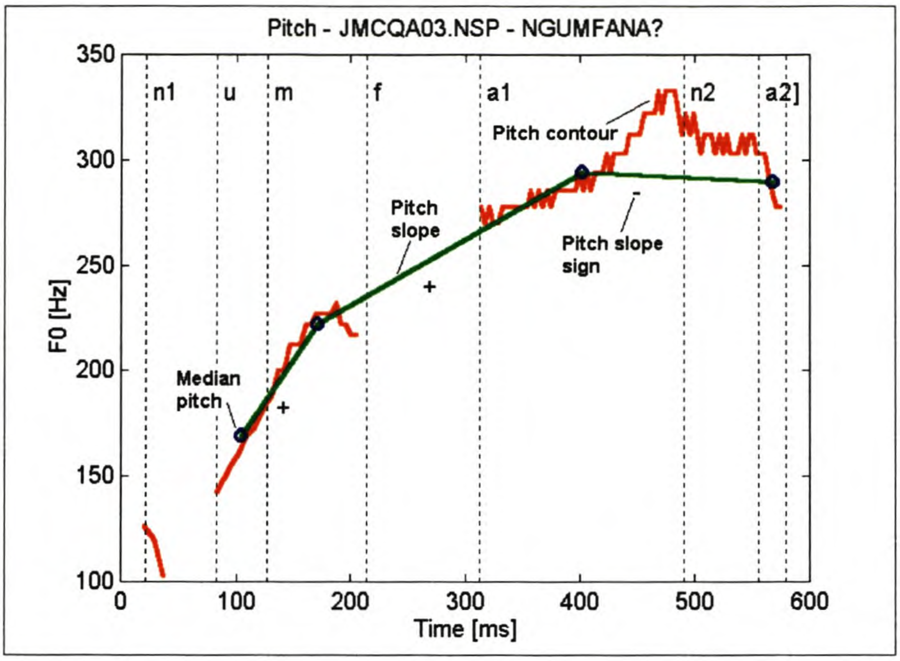


Figure 2.4 Pitch slope signs for the utterance /ngumfana/.

Patterns in the pitch slope signs were identified in the same manner as in the case of the significance of duration and pitch features. In this case the symbol *s* represents a '+' sign or a '–' sign. The symbol *s* represents a template that accepts a defined set of values. The pattern '*s* –' may be any one of the following:

(2.20) **+** –
– –

The pattern '*s* –' allows any sign ('+' or '–') in the first position, but restricts the second sign to be the '–' symbol. The pattern '*s s* –' includes the following possibilities:

(2.21) **– –** –
– + –
+ – –
+ + –

Patterns with more than one template character (i.e. '*s*') require systematic enumeration of all possibilities as in 2.21. A practical example of implementing these concepts is the following: the patterns '–', '*s* –' and '*s s* –' all represent cases where the pitch on the final syllable is lower than that of the penultimate syllable, regardless of the pitch contour of the rest of the utterance or the number of syllables in the utterance.

2.6.3 Results

The results of the statistical analysis on pitch are presented below and have been derived from the numerical data. Due to the quantity of data from which these results are derived, the complete set of tables representing the results on this parameter of each word have been included as Tables A.5, A.6 and A.7 in Appendix A.

Table 2.19 Average pitch measurements over all words.

Parameter	Queclaratives	Statements	Percentage change
Average word pitch	227.9 Hz	165.6 Hz	$\frac{227.9-165.6}{165.6}100\% = 37.6\%$

Table 2.20 Patterns observed from the Wilcoxon and $\bar{\mu}/\sigma$ analyses of vowel pitch features (derived from Table A.7).

Observation	Pattern	Occurrences
All vowels are significant.	$\mathbf{S}_a\ \mathbf{S}_b\ (1 \le a,b \le 2)$	1 / 1 (100%)
	$\mathbf{S}_a\ \mathbf{S}_b\ \mathbf{S}_c\ (1 \le a,b,c \le 3)$	26 / 27 (96.3%)
	$\mathbf{S}_a\ \mathbf{S}_b\ \mathbf{S}_c\ \mathbf{S}_d\ (1 \le a,b,c,d \le 4)$	9 / 11 (81.8%)
		Total: 36 / 39 (92.3%)
Penultimate vowel pitch is the *most* significant pitch feature.	$\mathbf{S}_1\ \mathbf{S}_2$	0 / 1 (0%)
	$\mathbf{S}_a\ \mathbf{S}_1\ \mathbf{S}_b\ (a,b \neq 1)$ $(0 \le a,b \le 3)$	15 / 27 (55.6%)
	$\mathbf{S}_a\ \mathbf{S}_b\ \mathbf{S}_1\ \mathbf{S}_c\ (a,b,c \neq 1)$ $(0 \le a,b,c \le 4)$	11 / 11 (100%)
		Total: 26 / 39 (66.7%)
Penultimate vowel pitch is the second *most* significant pitch feature while the pitch of the *final* vowel is most significant.	$\mathbf{S}_2\ \mathbf{S}_1$	1 / 1 (100%)
	$\mathbf{S}_a\ \mathbf{S}_2\ \mathbf{S}_1\ (a=0 \text{ or } a=3)$	11 / 27 (40.7%)
	$\mathbf{S}_a\ \mathbf{S}_b\ \mathbf{S}_2\ \mathbf{S}_1\ (0 \le a,b \le 4)$	0 / 11 (0%)
		Total: 12 / 39 (30.8%)
Other patterns on these words	$\mathbf{S}_a\ \mathbf{S}_2\ \mathbf{S}_1\ \mathbf{S}_b\ (0 \le a,b \le 4)$	10 / 11 (90.9%)

Note that the symbol s in Tables 2.21 and 2.22 below is a template that can take on the value of a '+' sign or a '–' sign.

Table 2.21 Patterns observed from the analysis of queclarative pitch slope signs (derived from Table A.6).

Observation	Patterns	Occurrences
Down-drift on the final syllable	$-\ ,\ s-\ ,\ s\,s-$	26 / 39 (66.7%)
Predominantly rising contour	$+\ ,\ ++\ ,\ ++s$	15 / 39 (38.5%)
Predominantly falling contour	$-\ ,\ --\ ,\ s--$	6 / 39 (15.4%)

Table 2.22 Patterns observed from the analysis of statement pitch slope signs (derived from Table A.6).

Observation	Patterns	Occurrences
Down-drift on the final syllable	$-$, $s-$, $ss-$	35 / 39 (89.7%)
Predominantly rising contour	$+$, $++$, $++s$	9 / 39 (23.1%)
Predominantly falling contour	$-$, $--$, $s--$	15 / 39 (38.5%)

The interpretations of the results of the acoustic and statistical analyses of pitch on the vowels (or syllabic /m/ where appropriate) of this corpus are presented below.

1. The average pitch of the word was found to be significantly higher for all queclaratives than for all statements. The queclarative pitch was on average 37.9% higher than that of the statements (Table 2.19). Therefore one may deduce from these results that the **register** for queclaratives is higher than for statements.

2. Every significant vowel of every queclarative had a **higher average pitch** than the corresponding vowel of the statement according to the positive sign of the $\overline{\mu}/\sigma$ (Table A.7). Similarly therefore the same deduction may be made, namely, that queclaratives have a higher overall register than statements.

3. For 92.3% of the words **all** vowels (and the syllabic /m/ where appropriate) were significant according to the Wilcoxon test (Table 2.20). Of the four insignificant vowels three were final vowels (Table A.7).

4. The **pitch** on the vowel of the penultimate syllable was the most significant pitch feature for 66.7% of the words. For 30.8% of the words, the pitch on the penultimate vowel was the *second* most important pitch feature, while the final vowel was the most significant (Table 2.20).

5. The order of significance of syllables obtained from analyzing **duration** features agreed with the order calculated using **pitch** features in 66.7% of the words (Tables A.4 and A.7).

6. **Down-drift** on the final vowel, i.e. where the pitch of the final vowel is lower than that of the penultimate vowel occurred for 89.7% of the statements and for 66.7% of the queclaratives (Tables 2.21 and 2.22).

7. For statements, a **rising contour** with down-drift on the final vowel was observed for 23.1% of the words, while a predominantly **falling contour** occurred in 38.5% of the cases (Table 2.22).

8. 38.5% of the queclaratives showed a predominantly **rising contour** and a **falling contour** in 15.4% of the cases (Table 2.21).
9. For the 27 three-syllable words the **pitch on the penultimate vowel** was the most significant pitch-differentiating feature in 55.6% of the cases, while the final vowel was most significant for 40.7% of the words (Table 2.20). The first vowel was always significant, but of lower significance.
10. For the 11 four-syllable words the **pitch on the penultimate vowels** was always the most significant disambiguating feature (Table 2.20).
11. The **sign test** returned the same significant vowels as the **Wilcoxon test** for 74.4% of the words, while in 25.6% of the cases it ranked a vowel of secondary importance as being insignificant (Table A.7).
12. The **Wilcoxon test** returned some equal significance values for more than one vowel in 89.7% of the cases. The $\overline{\mu}/\sigma$ can be used in these instances to rank vowels in order of significance. There was agreement between the Wilcoxon test and the $\overline{\mu}/\sigma$ in 92.3% of the cases in the general, as opposed to the finite, in the order of significance (Table A.7).

2.6.4 Discussion

In the introduction to the acoustic analysis of pitch, there were many references cited from scholars claiming the significance of pitch and intonation in the disambiguation process. In languages generally, many of these claims are supported by experimental data and perceptual testing. However, to date, to the author's knowledge, within the Bantu languages most claims are purely impressionistic. Very limited work in Xhosa has an experimental base and is still not supported by perceptual testing of data while experimental research has also been done on Chichewa (Myers, 1999).

Crystal (1985) and Ultan (1978) both claim that intonation is used to distinguish between declarative utterances and their question counterpart; furthermore, Ultan (1978) claims that questions all shared one feature, namely a raised register. Ohala (1983) also states that high or rising tone signals questions and low or falling pitch signals statements. From this acoustic analysis on this corpus of data, it is evident that these claims may also be said to be true for Xhosa.

The fact that the pitch of the whole word, and that of every significant vowel, was higher for queclaratives than for statements implies that the overall register for questions is higher than for statements. This finding supports the impressionistic claim made by De Francis (1963), corroborated by experimental research by Ho (1977) and J. Shen (1985) and Shen (1990) on Mandarin Chinese. These results for Xhosa similarly support Bolinger's comments to Shen (in her personal communication with him), stating that English would probably also show a higher overall pitch level throughout for questions as opposed to being signalled by terminals. Support also comes from Studdert-Kennedy and Hadding (1973), who found in their research in English, both experimentally and perceptually based, that, generally, questions were spoken with an overall high F0 compared to statements. The observations made by Eady and Cooper (1986) and previous authors Lieberman (1967) and O'Shaughnessy (1979) also support the finding that the questions have generally higher F0 values than the corresponding statement versions.

Within Bantu languages claims have been made relating to the use of intonation as a disambiguating factor between statements and questions, whether over the utterance as a whole or confined to specific areas, for example raising of tone on the final phonological syllable (Khumalo, 1981). Other claims have been documented on the use of pitch for disambiguation in languages such as Zulu, Xhosa, Tswana, Northern Sotho, Venda and Sothern Sotho. These results on this corpus of data would support these claims as a clear differential was evidenced in the average pitch levels between every statement and its corresponding queclarative counterpart. With regard to the evidence from this corpus, claims made concerning the raised overall pitch (register) in Bantu languages by Doke and Mofokeng (1957), Louw (1968), Riordan (1969), Theron (1991) and Myers (1999) may certainly be supported as, in the present research every queclarative had a higher pitch level (register) than its statement counterpart.

Apart from the raised register, however, it also appears that some specific relationship may exist between the pitch realization on the final two syllables. The pitch on the penultimate vowel appears to be the most significant feature in distinguishing between statements and queclaratives in 66.7% of the cases. The pitch on the final vowel was the most significant pitch feature in 30.8% of the cases, with the penultimate vowel being the second most significant disambiguating feature.

Evidence of down-drift on the final vowel was revealed in both statement (89.7%) and queclarative (66.7%) signals, while the raised register remained consistent throughout for queclaratives. This could be attributed to final lowering or declination, indicating finality or phrasing, which refutes Khumalo's (1981) claim that this final lowered tone only occurs in statements. Theron (1991) also lends support to the impressionistic claim that statements are characterized by tonal down-drift, in spite of her experimental findings that this also occurs for questions, yet less frequently. However, Myers (1999) concluded after his research in Chichewa that down-drift is less in questions than in statements which was also true for this corpus of data on Xhosa queclaratives. These findings do not support statements made by Snyman et al. (1991), who argue that a statement is converted into a question by raising the tone at the end of a sentence. While most queclaratives had a rising pitch contour, and statements a falling pitch contour, a number of statements had a rising contour and some queclaratives had a falling contour, which reinforces Geluykens (1987) and Fries (1964) that rising and falling contours occur in both statements and questions. This evidence suggests that categorical statements to the effect that "*question intonations always have a rising pitch contour as opposed to falling contours in statements*" cannot be supported.

2.7 Loudness

2.7.1 Introduction

The loudness or amplitude of an utterance has been postulated as being a possible factor contributing to differentiating between statements and questions.

Hirschberg and Ward (1992:245), in their study on the interpretation of the rise-fall-rise intonation contour in English, noticed consistent differences in the productions conveying uncertainty and incredulity in amplitude, duration, pitch range and spectral features. It was found (Ward and Hirschberg, 1988) that tokens of incredulity reading were in fact shorter and louder than those of uncertainty; and it was therefore hypothesized that one or more of these factors might account for differences in interpretation.. However Hirschberg and Ward (1992:250), in summary stated:

> However, amplitude and duration appeared to play no significant role in subjects' interpretation of the contour.

Jun and Oh (1996), in their study on three types of /wh/-phrases in Korean, found that the three types of questions have different peak amplitudes at two points in the /wh/-phrase. It was found that for all but one speaker the peak amplitudes of both the /wh/-word and the question particle were greatest in incredulity questions. It was also observed that incredulity questions were distinguishable by a higher amplitude.

According to Roux (n.d.), in his research on this topic of interrogativity in Xhosa:

> Based on this data (of which similar variations are attested), it may be argued that resetting of the declining intonation curve in the latter part of the articulation and raising of intensity levels at that point, changed a statement into a question.

It seems, therefore, that the fluctuation in amplitude is concomitant with fluctuation in F0 levels. Pierrehumbert (1979), in her experiment on the perception of fundamental frequency, concluded that amplitude, which normally accompanies fundamental frequency declination, may have an important role in the perception of phrasing. Once again the implication of this statement is that the amplitude level is in unison with the level of the fundamental frequency.

Roux (n.d.), in his experimental research on the prosodics of declarative and question sentences in Xhosa, found that after an initial comparable down-drift on the noun, rephrasing took place on the verb in which both amplitude and pitch were very prominent. The point is that although the question displays down-drift, it may be perceived as level or even rising due to the raised amplitude values. This assumption is implicitly supported by Pierrehumbert (1979:366) when she demonstrates that an increase in amplitude may in fact lead to the impression of higher pitch levels.

Myers (1999) found in Chichewa that the range of F0 values for high tones is higher for louder speech than for softer speech. These experiments infer that amplitude levels may be closely associated with pitch and that amplitude may be confused with pitch levels and be interpreted as raised pitch. By implication, therefore, the amplitude or loudness may contribute to the differentiation between a statement and a question.

2.7.2 Calculation of loudness

The loudness of a word was calculated in 20 ms non-overlapping windows. Both *relative* and *absolute* loudness levels were calculated. The absolute loudness measurements are inaccurate to a certain extent, since input levels varied somewhat between speakers. This is an effect of the different input methods that were used for the different speakers; some speakers were recorded on tape first while others were digitized directly. It is also difficult to select a fixed, optimum input level, even for a single speaker, because the clicks in Xhosa tend to cause clipping with higher input levels. The loudness contour of a word was normalized by dividing the samples with the maximum sample value for that word to yield a relative loudness contour. These relative loudness features are dimensionless and indicate the relative loudness of each phoneme within a word. The absolute loudness measurements, however, were not normalized.

The absolute loudness E in the kth window of length N samples was calculated as

$$(2.22)\quad E_k = \frac{1}{N}\sum_{n=1}^{N} x^2(n)$$

where $x(n)$ is the speech signal. The relative loudness e was given by

$$(2.23)\quad e_k = \frac{E_k}{E_{\max}}$$

where

$$(2.24)\quad E_{\max} = \max(E_k) \qquad \text{for all } k\,.$$

The windows were 20 ms in duration and non-overlapping.

Each phoneme, located by means of the tags, yields several loudness measurements, of which the average is taken as the loudness feature for that phoneme. For example the following loudness features were calculated for the word /ngumfana/:

(2.25) relative loudness: e_{n1} e_u e_m e_f e_{a1} e_{n2} e_{a2}
absolute loudness: E_{n1} E_u E_m E_f E_{a1} E_{n2} E_{a2}

The mean absolute loudness for this utterance ($E_{ngumfana}$) was defined as the average of the absolute, nonnormalized loudness measurements for the whole word.

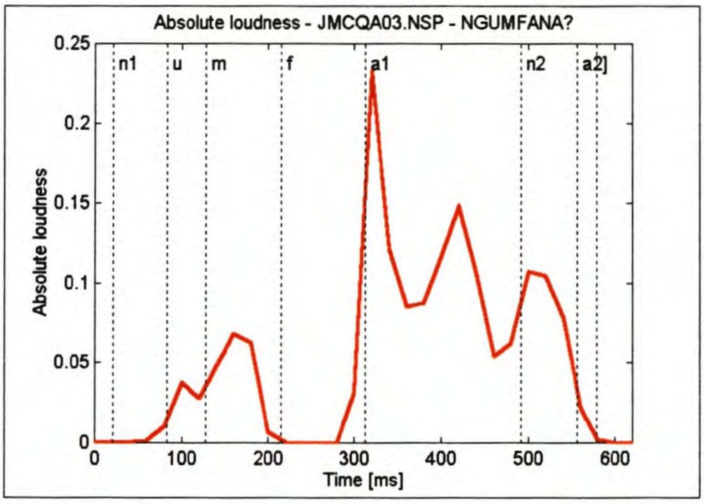


Figure 2.5 Absolute loudness contour for the utterance /ngumfana/.

2.7.3 Results

The following interpretations of the results of the acoustic and statistical analyses on loudness on this corpus of Xhosa copulative queclaratives are based on Tables A.8, A.9 and A.10 in Appendix A.

1. Based on the mean absolute loudness, the queclaratives were found to be **louder** than the statements (Table 2.23 below and Table A.9 in Appendix A).
2. **No consistent pattern** of significance could be found on the phonemes for either the relative or the absolute loudness features (Tables A.8 and A.10 respectively).

Table 2.23 Average absolute loudness measurements over all words.

Parameter	Queclaratives	Statements
Average absolute word loudness	0.0243263	0.0142447

2.7.4 Discussion

Although queclaratives in this corpus of Xhosa data were generally produced more loudly than the statements, neither the mean nor the relative loudness of each phoneme within a word was found to be significant in any examples. Therefore, due to the inconsistencies revealed in the statistical analysis, no claims can be made concerning the order of significance (if any) of the phoneme or syllable loudness features. However as queclaratives were generally produced louder than statements and their F0 values were higher than those of the statements, these results concur with those documented by Myers (1999) for Chichewa.

In conclusion further investigation of this feature as a contributory factor in differentiating between statements and questions under more controlled conditions is certainly viable and should be undertaken.

2.8 Results and observations

The following results were obtained, from these, observations are made from the acoustic and statistical analyses on the parameters duration, pitch and loudness on this corpus of data:

1. The duration of the word and that of every syllable was longer for statements than for queclaratives. Therefore the tempo in syllables per second was higher for queclaratives. This conclusion therefore supports claims made by Khumalo (1981), Louw (1968) and Theron (1991).
2. The duration of the penultimate syllable was the most significant syllable duration feature for 92.3% of the words, although other syllables may play a secondary role in disambiguation (59.0%).
3. The pitch of the whole word and that of every significant vowel was higher for queclaratives than for statements. This, therefore, by implication, means that the overall register for queclaratives is higher than for statements, once again supporting claims made by Louw (1968), Khumalo (1981), Riordan (1969) and Theron (1991).
4. The pitch on the penultimate vowel was the most significant pitch feature for 66.7% of the words. The pitch on the final vowel was the most significant pitch feature, with the penultimate vowel second most significant in 30.8% of the cases.
5. The evidence of down-drift was revealed in both statement (89.7%) and queclarative (66.7%) signals on the final vowel, while the raised register remained consistent

throughout for queclaratives. This could be attributed to final lowering or declination indicating finality or phrasing, which refutes Khumalo's (1981) claim that the final lowered tone only occurs in statements. Theron (1991) also lends support to this impressionistic claim that statements are characterized by tonal down-drift, in spite of her findings that this occurs for questions, yet less frequently.

6. While most queclaratives had a rising pitch contour and statements a falling pitch contour, evidence of the opposite trends was found in this corpus. This result concurs with findings documented by Geluykens (1987) and Fries (1964).
7. The queclaratives were louder than the statements. Nonetheless, neither the absolute nor the relative loudness of each phoneme within a word was found to be significant in any of the examples. Therefore, due to the inconsistencies revealed in the statistical analysis, no claims can be made on the order of significance (if any) of the phoneme or syllable loudness features.

The results of the acoustic analyses, which have been statistically validated, have elucidated very definite areas of significant acoustic differences between queclaratives and statements. These differences were found to be in duration, specifically on the penultimate vowel, pitch on the penultimate vowel and the overall raised pitch of queclaratives as opposed to statements.

These findings on this corpus of Xhosa copulative queclaratives and statement pairs, therefore, support the impressionistic claims to date that duration, tempo and register are indeed the parameters that differentiate statements from questions.

However, bearing in mind that there is no direct one-to-one relationship between the acoustic signal and its perception, it is evident that perceptual testing of the parameters is essential in validating research of this nature.

2.9 Chapter summary

This chapter provides:

1. the aims of the acoustic analysis followed by the methods used in the preparation of the data for analysis. This includes a description of the corpus, the process of recording, tagging, and then the statistical analysis.
2. the statistical results on the calculations of each parameter analyzed, namely duration, pitch and loudness on each utterance and each syllable of that utterance for the entire corpus of data.
3. a summary of the results received on each parameter analyzed and the proposed perceptual tests on the areas found to be acoustically significant.

CHAPTER 3

PERCEPTUAL EXPERIMENTS

The intellect pierces the form, overleaps the wall, detects intrinsic likeness between remote things and reduces all things into a few principles.

Ralph Waldo Emerson, Intellect 1841 (in Borden and Harris, 1984:166)

3.1 Introduction

Chapter 3 is a sequel to the acoustic analysis of statements and queclaratives in Xhosa. It includes a description and the results of the perceptual experiments compiled and conducted on selected data already acoustically analyzed. The corpus of data for this analysis was selected from one of the 11 mother tongue speakers whose data was used in the original acoustic analysis.

The focus therefore, of this chapter on Xhosa queclaratives, will be on the two perceptual experiments and the compilation and administration of each individual perceptual test on the relevant acoustic parameters. This research, by it's very nature, proposes a move away from impressionistic claims and interpretations of data to a more holistic approach lending authenticity and validity to research in this field of study within the Bantu languages.

Strange (1995:4) in the publication on speech perception and linguistic experience states that

> Although research on human perception is as old as psychology itself, empirical study of the perception of speech is of relatively recent origin.

Early researchers in this field soon discovered that there was no simple correspondence between segments of the acoustic signal on the one hand and perceived units on the other.

Therefore speech perception provides an example of the omnipresent problems related to perception that being of perceptual constancy (Strange 1995).

> The problem of perceptual constancy arises because there is no one-to-one correspondence between phonemes as perceived and the acoustic patterns generated by speech gestures that constitute the stimuli for speech perception. (Strange 1995:5)

By implication therefore the acoustic results of a study do not necessarily predict that features acoustically relevant are necessarily perceptually relevant. This fact endorses the motivation for a study of this nature including both an acoustic and perceptual investigation to produce authentic validated results.

According to t'Hart and Collier (1990:1) the code of language exploits numerous devices in order to convey the meaning of the message. It also controls other vocal features such as loudness, tempo, rhythm, pitch, and voice quality. While these features do not shape the phonetic identity of the utterance they constitute a truly segmental or prosodic tier in the sound pattern. They (t'Hart and Collier 1990:1) also state that;

> The prosody of an utterance adds an expressive dimension to the communication process: by modifying the prosodic features the speaker can supplement his utterance with elements of meaning that are not explicitly contained in its lexical and syntactic make-up. This added meaning must be taken in the broad sense of 'communicatively relevant information'

Significantly queclaratives which were selected and used for this investigation do not contain anything explicitly as a component either lexical or syntactic which may contain an element of meaning differentiating a queclarative from a statements. According to Jones et al. (1998:20)

> The nature of the queclarative structure provides a means to investigate how listeners differentiate between questions and statements at perceptual level. In particular the perceptual significance of the prosodic features pitch and duration was examined in this project.

In this instance it is necessary to consider what feature it is which provides this added meaning which is communicatively relevant in differentiating between queclaratives and statements. This therefore implies the possible role played by prosodics in the differentiation process.

When one moves to suprasegmental information, according to Mehler (1996:152)

> The picture seems to be quite unequivocal: acquisition takes place at a very early age. Jusczyk, Freiderici, Wessels, Svenkerud, and Jusczyk (1993) show that six-month olds prefer English to Norwegian words.
>
> Moreover as we show below, infants under two months of age have already extracted some properties of the prosodic representation characteristic of their native language.

During the past forty years a considerable amount of research has been undertaken and devoted to the mechanisms involved in the perception of spoken language. However a predominant percentage of this research has focused on the perception of specific speech sounds i.e. speech segments or phonemes which constitute syllables and words. A much lesser component of this research has been devoted to the processing of the cues required for the perception of prosodic (suprasegmental) aspects of spoken language.

According to House (1990:1)

> An orientation towards the segment as the object of perception research has also formed the basis for many current models and theories of speech perception (see Liberman and Studdert-Kennedy 1978, Picket 1980, Stevens and Blumstein 1981, Liberman and Mattingly 1985, Fowler 1986, and Repp 1988 for reviews and examples)

However, in spite of this bias in speech perception the perception of prosody has long been recognized as important and necessary for the perception of spoken language (Fry 1958, Lehiste1970, Svensson 1974, Nooteboom 1978).

According to House (1990:9) of the three acoustic correlates of prosody fundamental frequency, intensity and duration

> fundamental frequency is generally recognized as supplying listeners with the greatest amount of prosodic information on many different levels simultaneously (Bolinger 1958, Fry 1958, Bruce 1985)

These changes in fundamental frequency which are the result of controlled laryngeal movements made by the speaker assist the listener in the perception of many different linguistic categories which include relative syllable and word importance (stress, focus and emphasis respectively) and language specific information at both the word and phrase level (Ohala 1978).

According to Ohala (1983:15)

> The frequency code explains the similarities in cross-language and cross-cultural use of the pitch of the voice to mark questions versus non-questions, to signal different social attitudes (dominance, submission, assertiveness, politeness), and to refer to things small and large using sound symbolic vocabulary.

The precursor to perception or the interpretation of a message is the pronouncement or production of the message itself through speech. In order for the process of communication to take place therefore i.e. the transference of thoughts or feelings from the mind of a speaker to that of the listener the message has to be embodied in a linguistic framework. This process of communication which is described by Denes and Pinson (1973) as the 'speech chain' includes both the physiological process of message production and the retrieval thereof together with the process of the decoding of the sent message. According to Fraser (1992:3)

> Speech communication is seen as a process of the message transfer, in which the speaker converts a meaning (the message) into sound (articulatory gestures with acoustic consequences) which is transferred to the ear of the hearer. The hearer receives the sound and matches it against meanings, similar to those of the speaker, stored as part of his or her own linguistic knowledge.

This description above which is given within the IP (information-processing) framework of the communication process and includes speech perception may be loosely classified as psycholinguistics or as a branch of cognitive science. It is this process of information processing which includes human cognition where symbols or sounds are processed and ultimately transformed using stored knowledge to produce and arrive at an intelligent output or message. Succinctly therefore, this process of communication, is a progression from a speech level to a message level.

In the literature on speech perception the emphasis has been on the phoneme or syllable as the basic units of speech perception. According to Nooteboom, Brokx and de Rooij (1978:75)

> More recent proposals, in the line of the information processing approach, while still based on these old assumptions, now acknowledge the potential contribution of 'speech prosody' to perceptual segmentation (e.g. Pisoni and Sawush, 1975; Massaro, 1975; Massaro and Cohen, 1975)

It has been claimed that the pitch on the first part and on the final part of an utterance influences the judgement of listeners as to the question or statement identity of the utterance. Studdert-Kennedy and Hadding (1973:312) in their research on intonation contours using the utterance 'November' state that

> ...the single most powerful cue for question/statement judgements in this experiment was the terminal glide. Listeners evidently prefer, and presumably expect, a question to end with a rise, a statement with a fall (-----). However, earlier sections of the contour may also enter into the decision and, if sufficiently marked, override an incompatible, but weak terminal glide.

Before this perceptual analysis was contemplated an extensive acoustic analysis was performed on a corpus consisting of three copulative queclaratives from 13 noun classes of Xhosa from 11 informants. The results of these acoustic analyses were used as guidelines in designing a series of perception tests using a subset of the corpus that was acoustically analyzed. Details of these acoustic and statistical analyses of the full corpus are given in Chapter 2 of this study.

This perceptual analysis was comprised of two perceptual experiments designed and administered in which 64 and 63 Xhosa mother tongue listeners participated respectively. The results of the first experiment were used to refine and perfect the compilation and the administration of the tests presented in the second experiment.

The most important results of the acoustic analysis applicable to the design of the perception tests were as follows:

1. The **duration** of the **penultimate syllable** was the most significant syllable duration feature. The penultimate syllable was also shorter for queclaratives than for statements.
2. The **pitch** on the **penultimate vowel** was the most significant pitch feature. The pitch was higher on the penultimate vowel of the queclarative form than that of the statement form.
3. The overall **register** (pitch) was higher for queclaratives than for statements.
4. Although the queclaratives were found to be louder than the statements, no consistent pattern of significance of syllables was observed.
5. Down-drift was found to be present in both queclarative and statement forms.

The parameters selected therefore, were duration on the penultimate syllable and pitch on the penultimate vowel. A gating test was also compiled and included in both perceptual experiments in an attempt to determine at what point in the utterance a listener was able to identify questions and similarly statements (Cutler and Otake, 1999; Grosjean, 1996).

This chapter comprises detailed proceedings of two perceptual experiments, (namely Experiments A and B). **Experiment A**, the first perceptual experiment included three perception tests, namely **Test A1** compiled on duration on the penultimate syllable, **Test A2** on pitch on the penultimate vowel and **Test A3**, a gating test. The second experiment, **Experiment B**, was based on the results of the first experiment and comprised of four perception tests: **Test B1** compiled on duration on the penultimate syllable, **Test B2** pitch on the penultimate vowel, **Test B3** pitch on the first vowel and **Test B4** a gating test.

The stimuli for each of these experiments were manipulated signals from the recorded original data. The table that follows below is a summarized version of Experiments A and B. Information provided relates to the duration, number of stimuli, number of participants and number of responses for each individual test undertaken.

Table 3.1 The perception tests comprising Experiment A.

Experiment A					
Test	Description	Duration [minutes]	Number of stimuli	Number of participants	Number of responses
A1	Duration on the penultimate syllable	11:58	130	64	8320
A2	Pitch on the penultimate vowel	04:10	46	63	2898
A3	Gating test	16:27	202	61	12322
Total		***32:35***	***378***		***23540***

Table 3.2 The perception tests comprising Experiment B.

Experiment B					
Test	Description	Duration [minutes]	Number of stimuli	Number of participants	Number of responses
B1	Duration on the penultimate syllable	12:35	143	63	9009
B2	Pitch on the penultimate vowel	04:21	48	63	3024
B3	Pitch on the first vowel	06:44	78	63	4914
B4	Gating test	13:56	176	63	11088
Total		***37:36***	***445***		***28035***

The chapter begins with an introduction to the concept of perceptual testing and in so doing also justifies its inclusion in an experimental approach to language research. The aims of the experiments will then be presented and will be followed with a detailed account of automated perception testing using a multimedia computer laboratory. Then the procedure followed in the administration of these tests will be given followed by the methods employed in the analysis of the perception test responses, by using the chi square test. A description of each experiment under each of the relevant headings then follows. These will include an introduction, a description of the compilation of the stimuli, the methods of calculation of stimuli, and the results of the test on each specific parameter perceptually tested. Each experiment will be discussed individually under the relevant headings, within each parameter. The chapter will end with a summary and evaluation of the results obtained on each parameter and then consideration and comments on the responses observed in the light of possible universals in the perception of speech across languages will be presented.

3.2 Aims

As indicated in the introduction, the acoustic analysis revealed some of the strategies Xhosa speakers appear to employ in **encoding** differences between questions and statements through the acoustic speech signal. These perceptual experiments were designed to determine how Xhosa listeners **decode** the acoustic speech signal in differentiating between statements and queclaratives. The two experiments consisted of seven perception tests, (three in Experiment A and four in Experiment B) each test compiled with a specific aim in mind:

1. To determine to what extent the **duration** of the **penultimate syllable** is perceptually relevant in differentiation (Experiment A and B).
2. To determine to what extent the **pitch** on the **penultimate vowel** is perceptually relevant in differentiation (Experiment A and B).
3. To determine, through the implementation of a **gating test** (Otake and Cutler, 1999; Grosjean, 1996), the minimum temporal information subjects need in order to distinguish between queclaratives and statements (Experiment A and B).
4. To determine to what extent the **pitch** on the **first vowel** is perceptually relevant in differentiation if only the first syllables are presented to subjects. (Experiment B only).
5. To determine to what extent the results obtained from each of the parameters perceptually tested comply with human universals in speech perception.

3.3 Automated perception testing

The perception experiments were conducted using custom written software[1]. In the first experiment a total of 378 stimuli were presented to 64 listeners and in the second experiment 445 stimuli were presented to 63 listeners. This resulted in 51575 individual

---

[1] Developed by Mr J.A.N. Louw, RUEPUS, 1997.

responses. Clearly this number of responses cannot be transferred by hand from paper into computer readable form for further analysis. Since a system to do this automatically was not readily available, software was developed to (1) present the stimuli to listeners and (2) to collect their responses automatically. Additional advantages to this approach was that the listeners' response times could be measured and that a maximum limit could be imposed on the time available to respond.

The queclarative and statement stimuli were presented together in one group. Within a group the order of presentation of stimuli was randomized. The groups corresponded to the word classes. In the perception test all the queclarative and statement stimuli from the first group were presented, followed by the second group, etc.

Table 3.3 Examples of randomized queclarative and statement stimuli presented in the first two groups (classes) of Perception Test B1.

Group 1		**Group 2**	
Stimulus Number	Stimulus Filename	Stimulus Number	Stimulus Filename
1	JMCVSA3A	12	JMCVSB3A
2	JMCVSA3B	13	JMCVQB3D
3	JMCVQA3A	14	JMCVQB3C
4	JMCVQA3D	15	JMCVQB3B
5	JMCVSA3C	16	JMCVSB3D
6	JMCVSA3	17	JMCVQB3E
7	JMCVQA3	18	JMCVQB3
8	JMCVQA3B	19	JMCVSB3C
9	JMCVSA3D	20	JMCVSB3B
10	JMCVQA3E	21	JMCVSB3
11	JMCVQA3C	22	JMCVQB3A

The perception test program runs under Windows 95 and consists of a dialogue box with two buttons, identifying one for questions and one for statements in both English and Xhosa. These appeared as 'Question/Ngumbuzo' and 'Statement/Yinxelo' (Figure 3.1).

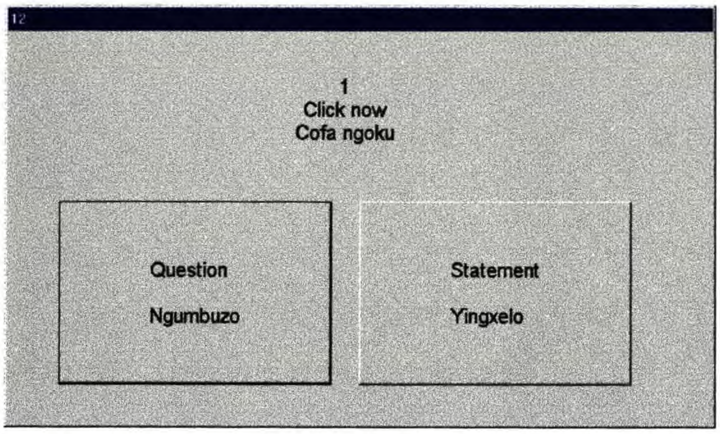


Figure 3.1 Screen layout of the perception test program prompting the subject to respond to a stimulus by clicking on either of the two buttons.

The perception test program performed the following actions:

1. A list of stimuli file names for the particular perception test was loaded.
2. Although the program was executed from a single location on the network, it was evoked with a unique number in the command line for each workstation. This number was used to create a unique file for each workstation for storing the responses.
3. Initially the buttons contained no text. The subjects were instructed on-screen to click with the left mouse button on either one of the two buttons to start the test when they were ready. The screen layout at that stage is shown in Figure 3.2.

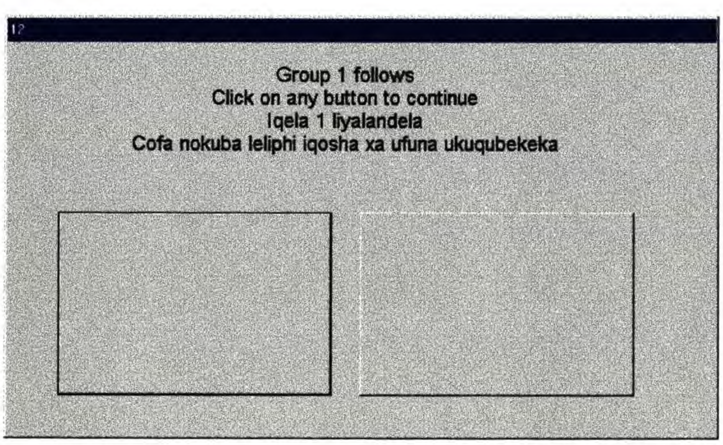


Figure 3.2 Screen layout prompting the listener to respond to start a batch of tests.

4. The stimulus was then played over the headphones. While the stimulus was being played, the buttons could not be clicked to prevent listeners from responding before the entire stimulus had been presented. In this phase the instruction on the screen was presented in English and Xhosa and read 'Listen/Mamela' (Figure 3.3).

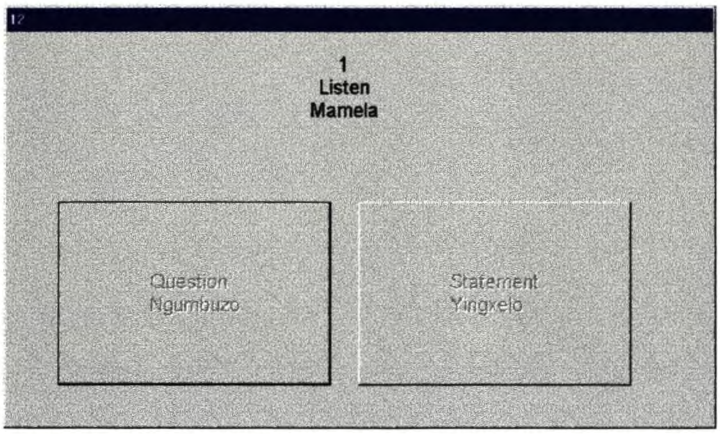


Figure 3.3 Screen layout while the program is playing the stimulus.

5. The buttons shown in Figure 3.3 above were then enabled and the instruction presented in both English and Xhosa changed to 'Click now/Cofa ngoku'. The listener then had 3 seconds to click on the 'Question' or the 'Statement' button depending on what the subject perceived the stimulus to be. If the 'Question' button was clicked a '1' was stored, while a '2' was stored if the 'Statement' button was clicked. If the subject did not respond within the 3 second interval that response was taken as Undecided and a '0' stored. The appearance of the screen at that point is shown in Figure 3.1. The stimulus number and the response time in milliseconds was also stored.
6. Steps 4 and 5 were repeated until all stimuli in a group had been played. Then step 3 was performed in order to commence with the next group of stimuli. Note that there was no time limit while the program was waiting for the listener to start the next group. This allowed slower participants to maintain composure, avoiding confusion, and allowed progress at the individuals' own rate within a group of stimuli.

After the perceptual testing the response files were processed. The processing entailed the following:

1. The number of Question, Statement and Undecided responses were counted for each stimulus.
2. The mean response time for each stimulus was calculated from the response times of the individual listeners.

3. The stimuli were sorted by filename so that successive manipulations followed each other.
4. The responses were then statistically analyzed.

Tables 3.4 and 3.5 below show extracts of the processed response data for the queclarative and statement forms of the word /ngumfana/.

Table 3.4 Example of processed Undecided, Question and Statement responses for the queclarative form of the word /ngumfana/ in Perception Test B1.

Original queclarative form					
Filename	Stimulus Number	Percentage Undecided	Percentage Question	Percentage Statement	Mean Response Time
JMCVQA3	7	3.2 %	90.5 %	6.3 %	990.3 ms
JMCVQA3A	3	1.6 %	95.2 %	3.2 %	878.8 ms
JMCVQA3B	8	4.8 %	49.2 %	46 %	1144.7 ms
JMCVQA3C	11	0 %	30.2 %	69.8 %	749 ms
JMCVQA3D	4	1.6 %	36.5 %	61.9 %	1102.3 ms
JMCVQA3E	10	1.6 %	33.3 %	65.1 %	978.4 ms

Table 3.5 Example of processed Undecided, Question and Statement responses for the statement form of the word /ngumfana/ in Perception Test B1.

Original statement form					
Filename	Stimulus Number	Percentage Undecided	Percentage Question	Percentage Statement	Mean Response Time
JMCVSA3	6	0 %	7.9 %	92.1 %	809.8 ms
JMCVSA3A	1	4.8 %	9.5 %	85.7 %	1028.8 ms
JMCVSA3B	2	1.6 %	3.2 %	95.2 %	693.5 ms
JMCVSA3C	5	0 %	34.9 %	65.1 %	911.7 ms
JMCVSA3D	9	4.8 %	31.7 %	63.5 %	1119.8 ms

These responses in percentages were also represented graphically as can be observed in Figures 3.4 and 3.5. This representation allowed the researcher to identify crossover points (Figure 3.4) where perceptions changed from question to statement and vice versa. It was also possible to observe a change in perception from one form to another by the convergence of a graph (Figure 3.5).

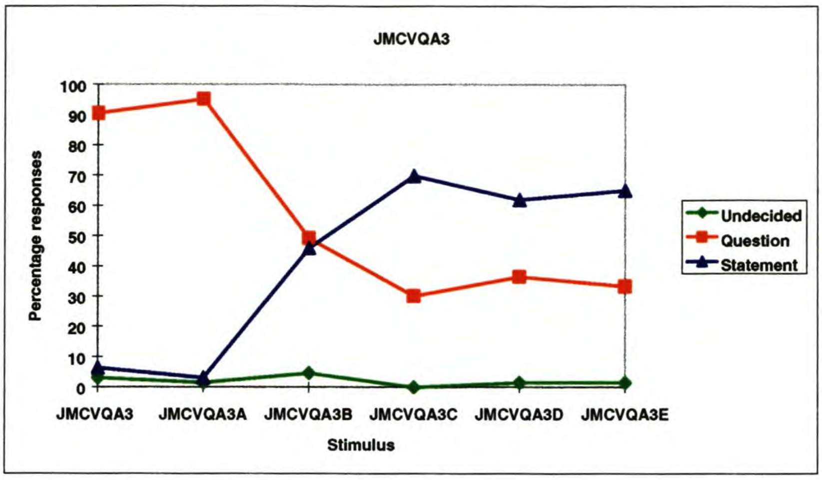


Figure 3.4 Example of a graph of the percentage Question, Statement and Undecided responses for the queclarative form of the word /ngumfana/ in Perception Test B1.

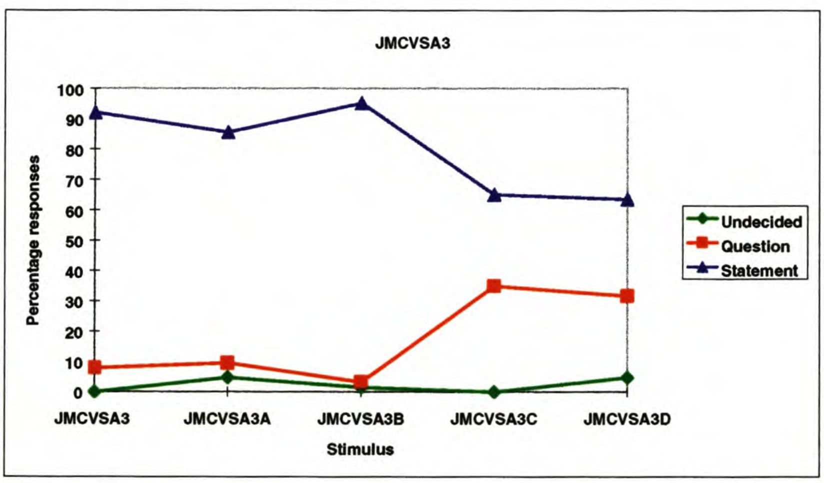


Figure 3.5 Example of a graph of the percentage Question, Statement and Undecided responses for the statement form of the word /ngumfana/ in Perception Test B1.

3.4 Administration of perception tests

The participants in the perception tests were Xhosa mother-tongue grade 11 and 12 pupils with no known hearing defects from Khayamandi High School in Stellenbosch whose ages varied between 15 and 28 years. Two perceptual experiments were conducted and 64 and 63 pupils participated in each experiment respectively.

The perception tests were administered on multimedia computers in the language laboratory of the University of Stellenbosch. Twenty pupils were accommodated at each testing session and the experiment did not exceed 45 minutes from the introductory lecture to its conclusion. The actual time spent directly by the participants in responding to stimuli was less than 38 minutes overall.

Before commencement of each of the testing sessions a questionnaire was completed by each of the participating pupils. An introductory lecture using transparencies for greater clarity was then presented, in both English and Xhosa. This provided the pupils with an insight into the project and instructions on the procedure to be followed in responding to the stimuli.

A practice test was compiled which consisted of examples of each of the perception tests in order to familiarise pupils with the manner of presentation of the data and to practise interacting with the computer using a mouse. These practice tests were administered 3 times to each sitting of respondents before commencement of the perception tests. The results of these tests were not recorded.

3.5 Analysis of perception test responses

The *chi square test* (Kreyszig, 1988) was used to formally evaluate the *statistical significance* of a majority response of one class and the significance of the changes in the number of responses.

The chi square test was employed in two variations: (1) the chi square test with one degree of freedom is applied to every stimulus in isolation to test whether the particular majority of Question or Statement responses is significant and (2) the chi square test with two degrees of freedom is used to classify the change in the number of responses from one stimulus to the next as being significant or not. These two tests will be described briefly below.

Consider the number of responses obtained for a stimulus numbered *i*. The number of Question, Statement and Undecided responses are $n_{i,Q}$, $n_{i,S}$ and $n_{i,U}$ respectively. The *observed frequency* for Questions, $f_{i,Q}$, is defined as:

(3.1) $$f_{i,Q} = n_{i,Q} + {}^{n_{i,U}}\!/_{2}$$

Similarly the observed frequency of the Statement responses $f_{i,S}$ is defined as:

$$(3.2) \quad f_{i,S} = n_{i,S} + {n_{i,U}}/{2}$$

Note that the Undecided responses are divided evenly between the Question and Statement responses in order to use the chi square test as indicated. It may be argued that Undecided responses could equally probably be Question or Statement responses. Therefore the Undecided responses are distributed evenly between the two main classes.

The chi square value χ_i^2 for the ith stimulus in the *chi square test with one degree of freedom* (Kreyszig, 1988) is calculated as shown below:

$$(3.3) \quad \chi_i^2 = \frac{(f_{i,Q} - e_i)^2}{e_i} + \frac{(f_{i,S} - e_i)^2}{e_i}$$

where e_i is the *expected frequency*:

$$(3.4) \quad e_i = \frac{n_{i,Q} + n_{i,S} + n_{i,U}}{2}$$

If χ_i^2 is greater than some threshold, it means that the particular majority, be it Question or Statement, is significant. The example below illustrates these concepts for one stimulus.

(3.5)

Filename	Stimulus Number	$n_{i,Q}$	$n_{i,S}$	$n_{i,U}$	$f_{i,Q}$	$f_{i,S}$	e_i	χ_i^2
JMCVQA3	7	57	4	2	58	5	31.5	44.59

In the *chi square test with two degrees of freedom* the chi square value χ_i^2 for the ith stimulus, given the previous stimulus (i.e. $i-1$) is given by:

$$(3.6) \quad \chi_i^2 = \frac{(f_{i-1,Q} - e_{i-1,Q})^2}{e_{i-1,Q}} + \frac{(f_{i-1,S} - e_{i-1,S})^2}{e_{i-1,S}} + \frac{(f_{i,Q} - e_{i,Q})^2}{e_{i,Q}} + \frac{(f_{i,S} - e_{i,S})^2}{e_{i,S}}$$

The expected frequency of the Question responses $e_{i,Q}$ is defined as:

$$(3.7) \quad e_{i,Q} = \frac{(f_{i-1,Q} + f_{i,Q})(f_{i,Q} + f_{i,S})}{(f_{i-1,Q} + f_{i-1,S}) + (f_{i,Q} + f_{i,S})} = \frac{(f_{i-1,Q} + f_{i,Q})(f_{i,Q} + f_{i,S})}{(f_{i,Q} + f_{i,S}) + (f_{i,Q} + f_{i,S})} = \frac{f_{i-1,Q} + f_{i,Q}}{2}$$

For Statements the expected frequency $e_{i,S}$ follows as:

$$(3.8) \quad e_{i,S} = \frac{f_{i-1,S} + f_{i,S}}{2}$$

Note that $e_{i-1,Q} = e_{i,Q}$ and $e_{i-1,S} = e_{i,S}$. If χ_i^2 is greater than some threshold, it means that the change in the number of Question and Statement responses ($n_{i-1,Q} \rightarrow n_{i,Q}$ and $n_{i-1,S} \rightarrow n_{i,S}$) from the previous stimulus ($i-1$) to the current stimulus (i) is significant.

(3.9)

Filename	Stimulus Number	$n_{i-1,Q}$	$n_{i-1,S}$	$n_{i-1,U}$	$f_{i-1,Q}$	$f_{i-1,S}$			
JMCVQA3	7	57	4	2	58	5			
Filename	Stimulus Number	$n_{i,Q}$	$n_{i,S}$	$n_{i,U}$	$f_{i,Q}$	$f_{i,S}$	$e_{i,Q}$	$e_{i,S}$	χ_i^2
JMCVQA3A	3	60	2	1	60.5	2.5	59.25	3.75	0.886

In all the perception tests conducted, excluding the gating test, original queclaratives were modified to emulate statements and statements were modified to emulate queclaratives. For the purpose of interpreting the results of these tests a *crossover point* was defined as a change in the number of responses of one class (e.g. Question) from a majority to a minority. Since the number of Undecided responses were low and fairly constant, this means that the number of responses of the other class (e.g. Statement) must have changed from a minority to a majority. An example of this type of crossover point is shown in Figure 3.6.

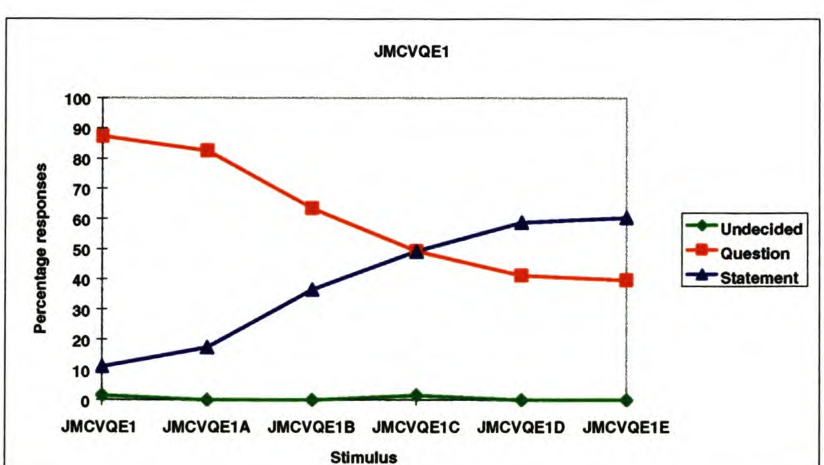


Figure 3.6 Example of a graph of the percentage Question, Statement and Undecided responses for a given set of stimuli showing a *crossover point*.

The presence of a crossover point was regarded as a *significant change in perception*. If the number of responses of both Question and Statement classes changed, but not enough so that the two curves crossed for the particular set of stimuli presented to subjects, the case was labelled as an instance where *some change in perception* occurred. Figure 3.7 shows an example.

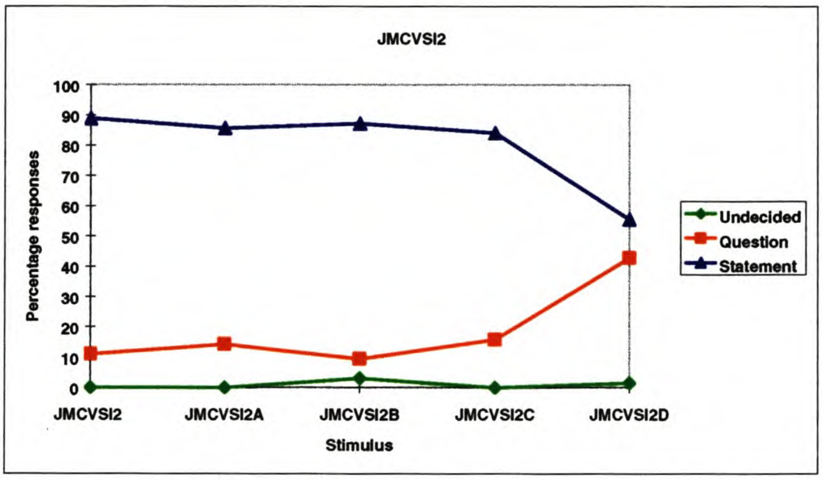


Figure 3.7 Example of a graph of the percentage Question, Statement and Undecided responses for a given set of stimuli showing *some change in perception* and a *possible crossover point.*

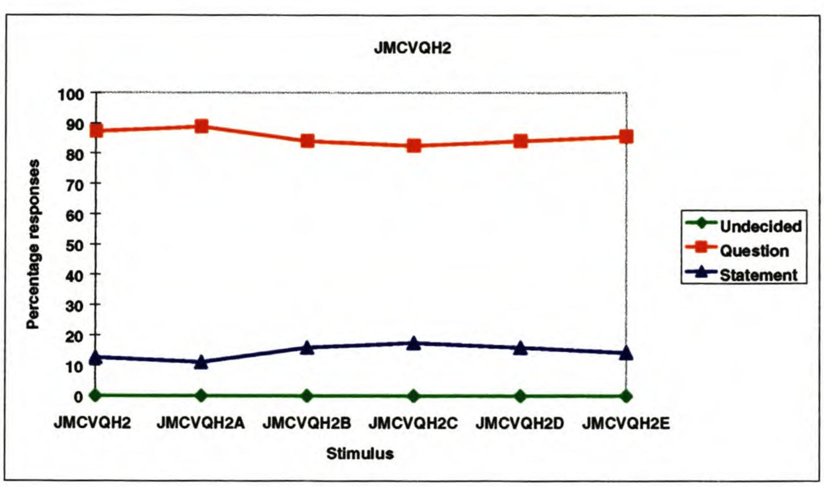


Figure 3.8 Example of a graph of the percentage Question, Statement and Undecided responses for a given set of stimuli showing *no significant change in perception.*

A crossover point would be *possible* after presentation of more stimuli that have been manipulated to a larger degree. A status of *no change in perception* would be assigned to cases where the number of responses of either class remained practically constant for all stimuli (Figure 3.8).

A graph was also made of the mean response time for each set of stimuli. In general decreasing response times would indicate that the subjects' confidence in their decisions increased. This is illustrated in the following two figures which show a decreasing trend in the mean response times (Figure 3.9) as gates of increasing duration (Figure 3.10) were presented to subjects. Only the increasing or decreasing trend of the mean response time was noted and no further statistical analyses were performed on this aspect.

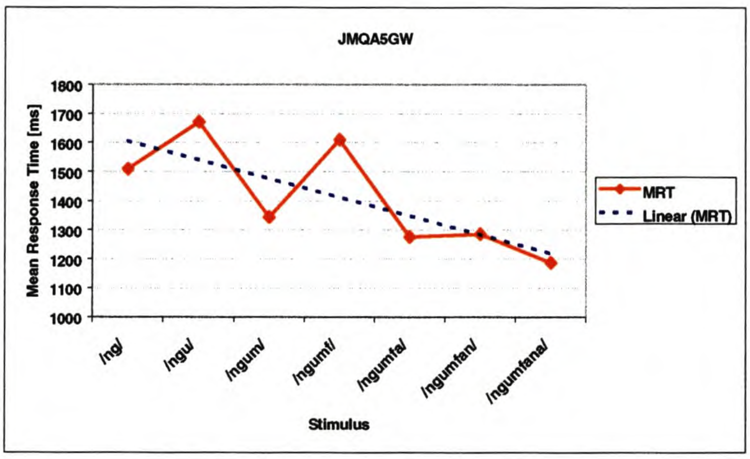


Figure 3.9 Mean response time for successive gates in the gating test of the queclarative /ngumfana/ showing a superimposed trend line (blue).

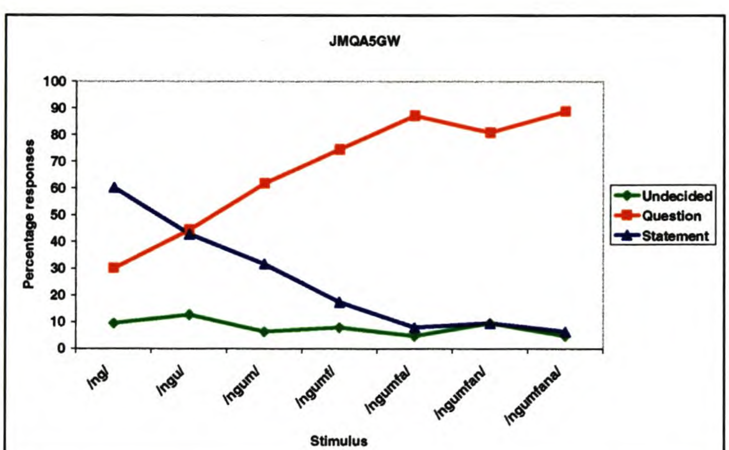


Figure 3.10 Responses for the queclarative /ngumfana/ in the gating test showing the initial misidentification and subsequent increase in confidence as segments increased in duration.

3.6 Perception experiments

Two perception experiments (A and B) were conducted each of which consisted of Perception Tests A1, A2, A3 and B1, B2, B3, B4 which will be discussed below under the relevant headings.

The results from the first experiment were used to refine the second experiment and in some instances the presentation of stimuli differed. However all the details of each test will be included for purposes of clarity under the relevant headings. The duration and pitch properties of stimuli are listed in Tables B.1 to B.7 in Appendix B due to the amount of numerical data involved.

The number of respondents for this perceptual study, consisting of two experiments, was 127 in total.

3.6.1 Experiment A, Test A1: Duration of penultimate syllable

Introduction

The use of duration as a possible feature in the process of disambiguation has been investigated by many researchers for many different languages. Wales and Taylor (1987:207-208) in their investigation of Australian English state that

> These results also show that DA (auxiliary duration) was a significant predictor for just over two thirds of the auxiliary classes, adding around 10% to the explained variance in each case. Hence, the important features do not appear to be limited to the terminal region of the sentence. Note that although sentences were uttered such that DS [duration of sentences final syllable, CJJ] of questions tended to be longer than that of the statements, this variable did not significantly predict question/statement judgements.

In 1988 Ward and Hirschberg observed that there were consistent prosodic differences between utterances produced to convey an uncertainty as opposed to an incredulity reading in the parameters namely duration, amplitude, pitch range and spectral characteristics.

Hirschberg and Ward (1992) in their research on the final rise-fall-rise intonation contour in English also investigated 'incredulity' as opposed to 'uncertainty' readings of a contour considering the features duration, amplitude, pitch range and spectral characteristics. They state that in their 1988 research

> Tokens of the latter [meaning incredulity, CJJ] tended to be shorter and louder than tokens of the former [meaning uncertainty CJJ]...

Similarly in the 1992 investigation they state that:

> ...uncertainty is somewhat longer than the same uttered to convey incredulity,...

In their final conclusions Hirschberg and Ward (1992:250) also found that

> However, amplitude and duration appeared to play no significant role in subjects interpretation of the contour.

Muira and Hara (1995:291) in their research in Osaka Japanese emphasize the many studies both acoustic and perceptual which have been conducted in identification of the factors contributing to disambiguating syntactically ambiguous sentences as also did Streeter (1978), Berkovits (1981), Scott (1982) and Price, Ostendorf, Shattuck-Hufnagel and Fong (1991). All these studies have verified the importance of F0 and segmental duration in perceptual disambiguation. In their investigation Muira and Hara (1995:301) also found that manipulation of duration on the sentence final segment in Yes/No questions was a factor for questions to be perceived as rhetorical. This again verifies the possible use of duration in the disambiguation of utterances.

Within the Bantu languages claims relating to the use of duration as a disambiguating factor are once again predominantly impressionistic in nature as expressed by Doke and Mofokeng (1957), Lanham (1963) and Riordan (1969) to name but a few.

Theron (1991:89) in her research on the phonetic and phonological features of statements and questions in Xhosa indicates a distinct shortening of sentence duration in questions as opposed to statements. The implication of a difference in duration in the process of disambiguation implies that the rate of production or the tempo is simultaneously affected. This aspect was also noted by Khumalo (1981:92), and also largely supported by Theron's experimental findings in 1991. As the possible use of duration as a disambiguating factor has been documented in many studies and was also shown in this acoustic analysis of Xhosa data to be significant, the following perceptual test on this parameter was designed and administered to Xhosa mother tongue listeners. The aim of this test therefore was to determine whether the acoustically significant duration feature on the penultimate syllable was indeed perceptually relevant.

Method used in compilation of stimuli

The corpus used in the compilation of this test are listed in Table 3.6 below.

Table 3.6 Corpus used in Perception Test A1.

Word	Identifier	Word	Identifier
ngumfana	A03	zizifo	H02
ngabafana	B03	yintloko	I02
ngumthi	C01	ziinkomo	J03
yimililo	D02	lulwimi	K02
lilitye	E01	bubuhlanti	L02
ngamafu	F03	kukutya	M01
sisilo	G03		

The manipulated signals presented for perceptual testing totalled 130 in this test. The original unmodified utterance and three manipulations of the statements and similarly three of the queclaratives were presented for differentiation. One manipulated stimulus (version 'B') was presented twice. Stimuli were presented in groups of ten sound files comprising the five question and five statement stimuli for each word. These ten stimuli were randomized as shown in Table 3.6. The next group would consist of the stimuli for the following word.

Using the Analysis/Synthesis Lab (ASL) program of Kay Elemetrics the penultimate syllables of queclaratives were lengthened in steps of equal size. The step size depended upon the difference between the penultimate syllable duration of the queclarative and the statement. This was done to determine the amount of duration change necessary for subjects to perceive queclarative stimuli as statements. Using the same principle the penultimate syllables of statements were shortened to emulate the penultimate syllable duration of queclaratives. The manipulation was done only on the penultimate vowel as there were instances of voiceless consonants preceding these vowels which made the manipulation of the syllable itself (i.e. CV) not feasible.

Table 3.7 Stimuli presented for the word /ngumfana/ in Perception Test A1.

Randomized		**Sorted**			
		Original question form		**Original statement form**	
Filename	Stimulus Number	Filename	Stimulus Number	Filename	Stimulus Number
JMCVQA3B	1	JMCVQA3	7	JMCVSA3	3
JMCVSA3B	2	JMCVQA3A	6	JMCVSA3A	5
JMCVSA3	3	JMCVQA3B	1	JMCVSA3B	2
JMCVSA3B	4	JMCVQA3B	9	JMCVSA3B	4
JMCVSA3A	5	JMCVQA3C	10	JMCVSA3C	8
JMCVQA3A	6				
JMCVQA3	7				
JMCVSA3C	8				
JMCVQA3B	9				
JMCVQA3C	10				

The following example will illustrate how the step size, Δd_{Qa1}, was calculated for the word /ngumfana/ in Perception Test A1. The number of steps, *n*, was chosen as 3. This choice will be discussed later. The same procedure was followed for each queclarative/statement pair in the corpus. Suppose the duration of the penultimate vowel ('a1') of the queclarative and the statements forms is d_{Qa1} and d_{Sa1} respectively. Then the step size follows as:

(3.10) $$\Delta d_{Qa1} = \frac{d_{Sa1} - d_{Qa1}}{n} = \frac{370.2 - 177.7}{3} = 64.2 \text{ ms}$$

The duration of the penultimate vowels of the *k*th stimulus, d^k_{Qa1}, was obtained as shown below:

(3.11) $$d^k_{Qa1} = d_{Qa1} + k\Delta d_{Qa1} \quad \mathrm{k} = 0,1,2,3$$

Therefore, the first stimulus (k = 0) had the penultimate vowel duration of the original queclarative form: $d^0_{Qa1} = d_{Qa1}$. The last stimulus (k = 3) had a penultimate vowel of duration $d^3_{Qa1} = d_{Qa1} + 3\Delta d_{Qa1}$. Table 3.8 shows the numerical values for the word /ngumfana/ that were calculated and measured for the penultimate vowel duration of the queclarative stimuli in Perception Test A1.

Table 3.8 Duration changes for the queclarative form of the word /ngumfana/ in perception Test A1.

Stimulus file	k	$k\Delta d_{Qa1}$	d_{Qa1}^{k}
JMCVQA3	0	0 ms	177.7 ms
JMCVQA3A	1	64.2 ms	241.9 ms
JMCVQA3B	2	128.4 ms	306.1 ms
JMCVQA3C	3	192.6 ms	370.3 ms

The choice for the number of steps, n, will now be discussed using the queclarative form of the word /ngumfana/ as an example. As shown in Figure 3.11 the original value of the duration of the penultimate vowel of the queclarative form, d_{Qa1} , is the starting point. The hypothesis experimentally evaluated in Test A1 was that subjects could be convinced that a stimulus was a statement by lengthening the penultimate vowel duration of an original queclarative form to equal the length of the statement form. Therefore, the target value for an original queclarative form was taken as the value of the statement form. However, taking only one measurement (i.e. one target value) would seriously impair the scientific credibility of the experiment. For this reason a number of intermediate steps were included. By presenting a range of stimuli which were manipulated to an increasingly greater degree, it would be possible to investigate subjects' responses to the whole range of manipulations and to infer the amount of manipulation required to alter subjects' perceptions. A constraint on the maximum number of intermediate steps is the duration of the entire perception experiment. It was decided to restrict the duration of Experiment A to 30 minutes in order not to exhaust the concentration level of subjects to the point where their responses would be meaningless. With the choice of $n = 3$ the actual duration of the perception test was no more than 33 minutes. There were two intermediate points between the original value and the target value.

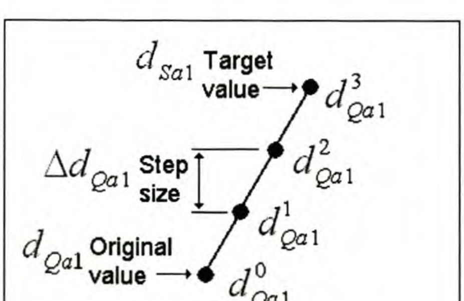


Figure 3.11 Calculation of duration changes for the queclarative form of the word /ngumfana/ in Perception Test A1.

The procedure outlined above for queclarative stimuli was repeated for statement stimuli with the only change that successive stimuli were shortened instead of lengthened. Thus the step size for the statement form, Δd_{Sa1} , is just the negative of the step size for the queclarative form:

$$(3.12)\quad \Delta d_{Sa1} = \frac{d_{Qa1} - d_{Sa1}}{n} = \frac{177.7 - 370.2}{3} = -64.2 \text{ ms}$$

Similar to the queclarative form, the calculation and example numerical values of stimuli vowel duration for the statement form of the word /ngumfana/ appear in 3.13 and Table 3.9 below.

$$(3.13)\quad d_{Sa1}^{k} = d_{Sa1} + k\Delta d_{Sa1} \quad \text{k} = 0,1,2,3$$

Table 3.9 Duration changes for the statement form of the word /ngumfana/ in Perception Test A1.

Stimulus file	k	$k\Delta d_{Sa1}$	d_{Sa1}^{k}
JMCVSA3	0	0 ms	370.2 ms
JMCVSA3A	1	64.2 ms	306 ms
JMCVSA3B	2	128.4 ms	241.8 ms
JMCVSA3C	3	192.6 ms	177.6 ms

The lengthening of the queclaratives was achieved in the ASL program by duplicating the pitch period at the centre of the penultimate vowel until the desired target length had been reached. After the addition of pitch periods the signal was resynthesized and saved which then constituted a stimulus for presentation. The centres of the penultimate vowels were tagged manually in preparation for this test. For the word /ngumafana/ the centre of the penultimate vowel ('a1') was indicated by the addition of the tag '+a1' (Figure 3.12). Figure 3.13 shows the speech signal after successive lengthening of the penultimate vowel duration using the described procedure.

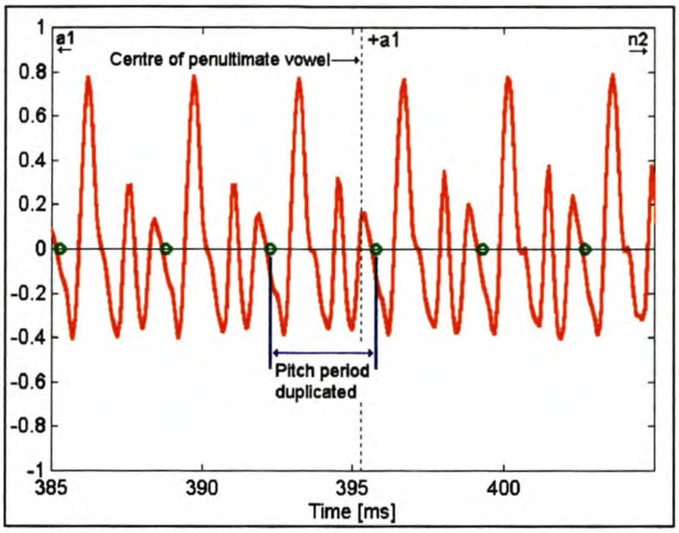


Figure 3.12 Detail of queclarative vowel centre.

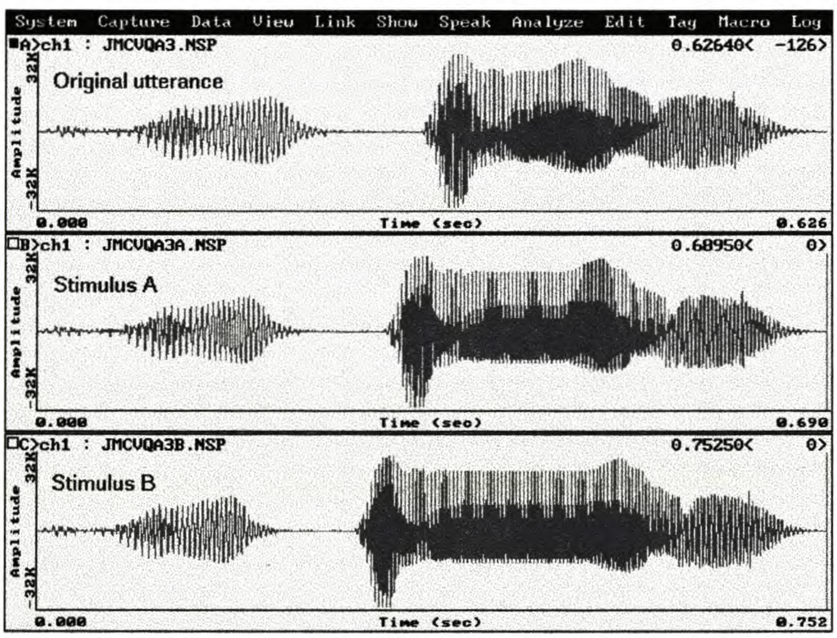


Figure 3.13 Examples of lengthening of the penultimate vowel duration of a queclarative form.

In the shortening of original statement signals, a number of pitch periods were cut out in equal proportions on either side of the tagged centre (e.g. '+a1') to achieve the queclarative target length. Figure 3.14 shows the deletion of two pitch periods. After the deletion of pitch periods with the ASL program the signal was resynthesized and saved. Figure 3.15 illustrates the Numeric Editor in the ASL program that was used to duplicate or delete pitch periods.

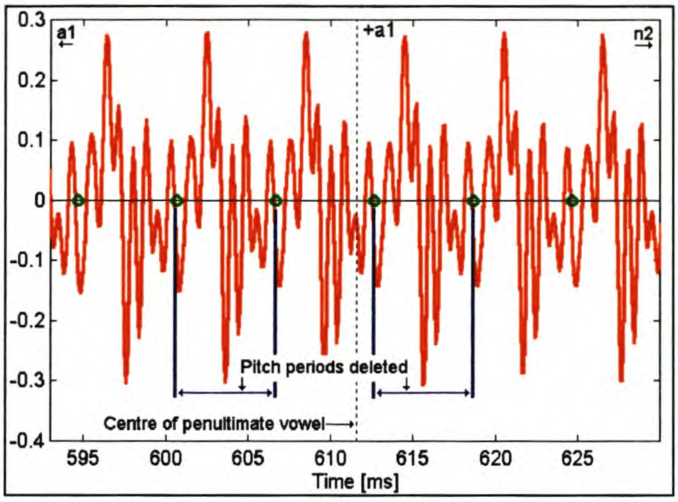
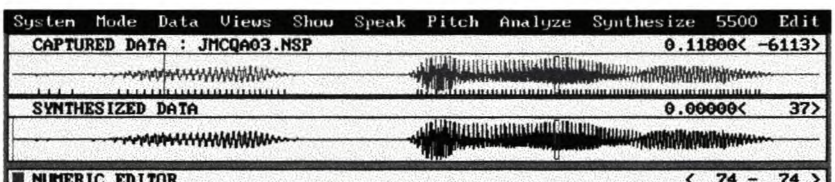


Figure 3.14 Detail of statement vowel centre.



#	M	RES	PK	F0	LEN	F1	B1	F2	B2	F3	B3	F4	B4
64	0	64	25385	286	35	506	198	821	163	1392	67	2606	110
65	0	65	25525	286	35	528	228	806	197	1392	77	2611	121
66	0	66	25234	286	35	527	219	819	202	1395	78	2619	87
67	0	67	25078	286	35	540	241	819	222	1396	87	2631	112
68	0	68	25419	294	34	547	263	812	223	1400	82	2602	109
69	0	69	25853	286	35	575	248	830	260	1409	110	2616	91
70	0	70	26103	294	34	560	263	830	220	1403	99	2598	122
71	0	71	27224	286	35	610	276	817	267	1400	132	2619	149
72	0	72	28412	294	34	631	388	787	217	1394	166	2616	224
73	0	73	28851	294	34	631	376	801	211	1398	179	2594	194
74	0	74	29494	294	34	606	311	848	196	1423	202	2577	209
75	0	75	30114	303	33	555	236	896	145	1464	219	2574	161
76	0	76	30539	294	34	534	182	910	105	1496	220	2597	174
77	0	77	30766	303	33	504	144	905	76	1453	143	2619	115
78	0	78	29475	303	33	519	153	923	77	1524	297	2669	168
79	0	79	27404	303	33	522	148	934	71	1681	354	2758	226
80	0	80	26756	303	33	506	137	931	67	1635	268	2712	228
81	0	81	26285	156	64	467	144	933	54	1714	340	2583	244
82	0	82	24889	156	64	450	140	949	68	1633	292	2606	188
83	0	83	20557	161	62	425	116	997	104	1596	262	2787	206
84	0	84	18424	161	62	394	87	1003	72	1586	179	2951	155
85	0	85	17449	164	61	376	79	1013	77	1626	181	2966	95
86	0	86	16193	167	60	370	72	1027	129	1607	185	3021	67

Figure 3.15 Screen layout of the ASL program used for manipulation of the duration of the penultimate vowel.

Results

The results shown in Tables 3.10 and 3.11 below have been derived from the actual responses listed in Table B.9 in Appendix B.

Table 3.10 Results of the manipulation of the duration of the penultimate syllable in Perception Test A1.

	Manipulation and number of noun classes to which observation applies	
Observation	Duration of penultimate syllable of a question form increased	Duration of penultimate syllable of a statement form decreased
•Crossover point observed •Significant change in perception	3 (23.1%) QA03, QB03, QF03	0 (0%)
•Crossover point improbable •Some change in perception	4 (30.8%) QD02, QE01, QI02, QL02	5 (38.5%) SA03, SB03, SE01, SF03, SL02
•Crossover point improbable •No change in perception	6 (46.2%) QC01, QG03, QH02, QJ03, QK02, QM01	8 (61.5%) SC01, SD02, SG03, SH02, SI02, SJ03, SK02, SM01

Table 3.11 Detail of crossover points achieved in Perception Test A1.

	Original	**Stimulus A**	**Stimulus B**	**Stimulus B**	**Stimulus C**
Queclarative – ngumfana (QA03)					
d_{a1}	177.7 ms	240.7 ms	303.7 ms	303.7 ms	370.2 ms
% change in d_{a1}		35.5%	70.9%	70.9%	108.3%
% queclarative	95%	96. 7%	36. 7%	58. 3%	35%
Comment	smQ	smQ	mS, cop	mQ	smS
Queclarative – ngabafana (QB03)					
d_{a3}	127.7 ms	190.4 ms	253.1 ms	253.1 ms	319.1 ms
% change in d_{a3}		49.1%	98.2%	98.2%	149.9%
% queclarative	93.3%	93.3%	81.7%	78.3%	25%
Comment	smQ	smQ	smQ	smQ	smS, cop
Queclarative – ngamafu (QF03)					
d_a	176.9 ms	247.7 ms	318.5 ms	318.5 ms	389.3 ms
% change in d_a		40%	80%	80%	120.1%
% queclarative	91.7%	68.3%	40%	51.7%	50%
Comment	smQ	smQ	mS, cop	mQ	mQ

Note that stimulus B appears twice in Table 3.11 because that stimulus was repeated in the perception test. The following abbreviations are used in Table 3.11:

smQ significant majority Queclarative

smS significant majority Statement

mQ majority Queclarative

mS majority Statement

cop crossover point

The following results were obtained from this test:

1. Crossover points and a significant change in perception was achieved for 23.1% of the original queclarative forms that were lengthened to emulate statements (Table 3.11).
2. The shortening of original statements did not result in any crossover points or definite changes in perception (Table 3.11).
3. For 34.6% (9/26) of all queclarative and statement forms some change in subjects' perception could be observed, although not enough to result in a crossover point.
4. Overall, for 53.9% (7/13) of queclaratives and 38.5% (5/13) of statements a change in perception, be it large or small, was achieved (Table 3.10).
5. In 53.9% (14/26) of the cases no change in perception was achieved (Table 3.10).

3.6.2 Experiment A, Test A2: Pitch on the penultimate vowel

Introduction

As can be seen in the introduction to this perceptual analysis where many examples have been cited regarding the use of pitch as a disambiguating factor this tenet has been postulated by many researchers of many different languages of the world. Many such studies concern the English language and show uniformity in their results of fundamental frequency being an important factor in disambiguating syntactically ambiguous sentences.

Pioneering studies by Hadding-Koch and Studdert-Kennedy (1963, 1964) and by Gårding and Abramson (1965) did research on the perceptual differentiation of questions and statements in American English and Swedish. These early contributions showed that perceptually it was not only the final rise or fall which differentiated between these forms but more or other parts of the intonation contour which also played a role in this process. Thorsen (1980) using both manipulated speech stimuli in her study of Danish questions and statements found that listeners make use of the general course of the entire F0 contour in the differentiation process. Muira and Hara (1995) in their study of the production and perception of rhetorical questions in Osaka Japanese also found that lowering of the sentence initial F0 and sentence final raising in the 'wh' question was a dominant factor for the question to be perceived as rhetorical.

According to Eady and Cooper (1986:409) in their research entitled *Speech intonation and focus location in matched statements and questions*

> Thus, regardless of whether a sentence is a question or a statement and irrespective of any focused words, the initial F0 peak seems to be fairly invariant for the sentences examined here.

Eady and Cooper (1986:403) state that early descriptions of intonation in English show that F0 (or voice pitch) has a generally rising pattern in yes-no questions, in contrast to the generally falling pattern of statements (Pike,1945) and that the shape of the contour at the end of the sentence is said to be a very important factor. Lieberman too (1967) also makes claims to this effect and states that this differentiation process takes place in the final 150-200ms of phonation. Wales and Taylor (1987:209) in their study on Australian English state that

> There it was found that F0 peaks associated with sentence focus (accent) could serve as predictors of question/statement judgements, although not as powerful predictors as the peaks occurring at the end of the sentence.
>
> The present results suggest that coupled or sequential F0 sweeps serve as important cues to intonation-based question/statement differentiation, with other cues playing less important roles.

Clearly from the above comments Wales and Taylor (1987) found that F0 serves as a very important cue to intonation based questions and that other cues are certainly of secondary importance in the disambiguation between statements and questions. They also made an interesting comment when suggesting that intonation or prosodic features may carry intrinsic meaning of their own and that they may serve as attention markers that attract the listeners attention. This proposal receives some support in research which has shown that particular subgroups of English speakers use the cues of the terminal F0 contour to convey different meanings (Allan 1984, Guy and Vonwiller, 1984).

Within the Bantu languages similar claims on the use of pitch as a disambiguating factor have been documented. The numerous contributions however remain impressionistic in nature. These impressionistic claims relating to pitch as a factor in disambiguation have been made for languages such as Sesotho, Zulu, Tswana, Northern Sotho, Venda, and Xhosa.

Nkabinde (1999:378-379) commenting on interrogative and declarative sentences in Zulu makes the following claims

> The interrogative sentence has a quick tempo, a high pitch and no down-stepping.
>
> This type of sentence [referring to declaratives CJJ] is characterized by a cadent intonation with unchecked length on the penultimate syllable of the only word or the last word in a sentence. It has a low pitch...

Mathibela (1989:15) says that

> Tswana intonational pattern questions end with a rising intonation.

In Xhosa however the limited experimental work done by Louw (1968) and Theron (1991) also coincide with these impressionistic notions in spite of them lacking in perceptual testing of data.

As this prosodic feature of pitch has been widely documented, after experimental research including perceptual testing, as a contributing factor in the process of disambiguation in so many languages of the world it would seem appropriate that this should also be done for the Bantu languages. As the acoustic analysis on this corpus of data shows that the pitch is acoustically significant specifically on the penultimate vowel, this first perceptual test was designed to test the validity of these findings from a perceptual perspective to see if in fact if this was the case for Xhosa.

In the acoustic analysis the overall pitch of the questions was found to be higher than that of the statements, with the pitch difference on the penultimate vowel being the most significant. The aim of this test was to determine whether listeners' perceptions of the identity (i.e. question or statement) could be shifted if the pitch on the penultimate vowels of queclaratives were reduced towards the level of statements. For the statements the pitch on the penultimate vowel would be raised to approximate the level of the pitch on the penultimate vowel of the queclarative.

Method used in the compilation of stimuli

The corpus used in the compilation of this test are listed in Table 3.12 below.

Table 3.12 Corpus used in Perception Test A2.

Word	Identifier
ngumfana	A03
lilitye	E01
yintloko	I02
lulwimi	K02
bubuhlanti	L02

Five word pairs (statement/queclarative) from the original corpus were selected for manipulation of pitch on the penultimate vowel and 46 stimuli were presented in this test. The selection was based on the requirement that the queclarative and statement forms had to have a fairly stable pitch contour on the penultimate vowel and the consonants preceding and following it.

The pitch on the penultimate vowels of queclaratives was reduced in three steps of approximately 30 Hz towards the level of statements. Similarly for the statements the pitch on the penultimate vowels was raised in three steps. Software[2] was developed to manipulate the pitch contour of a speech file by drawing the desired pitch contour with the mouse on a graph with several superimposed pitch contours. Figures 3.16 and 3.17 below show the lowering and raising of pitch on the penultimate vowels of queclaratives and statements respectively.

The methods that were used to implement the manipulation of the pitch contour was that of interpolation and deletion of samples. Interpolation was done to insert a number of samples into each pitch period. This causes the duration between voice impulses to

---

[2] Developed by Mr J.A.N. Louw, RUEPUS, 1997.

increase and thus results in a lower pitch. By deleting samples from each pitch period the voice impulses are moved closer to each other resulting in a raising of the pitch. For relatively small changes the degradation in speech quality is tolerable, but as the amount of manipulation increases so does the intelligibility decrease. It was found that this effect particularly limited the number of steps that could be synthesized by raising the pitch of statements. For some statements less steps were included in the perception test. For this reason stimuli were not repeated for statements as was the case for queclaratives. The greatest degree of manipulation that could be done while maintaining speech intelligibility was determined to be approximately 90 Hz. This value was found by repeated synthesis and listening by the author. The 'C' stimuli shown in Figures 3.16 and 3.17 thus represent the largest change in pitch. The 'A' and 'B' stimuli was therefore spaced evenly between the original pitch contour and the 'C' stimulus contour, resulting in the 30 Hz step size mentioned earlier.

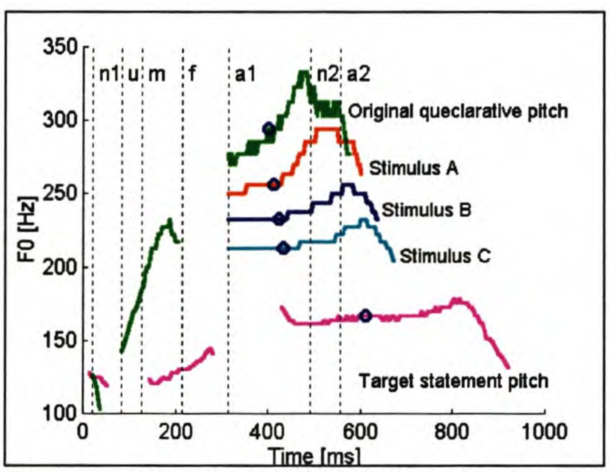


Figure 3.16 Example of lowering the pitch on the penultimate vowel of the queclarative of the utterance /ngumfana/ in Perception Test A2.

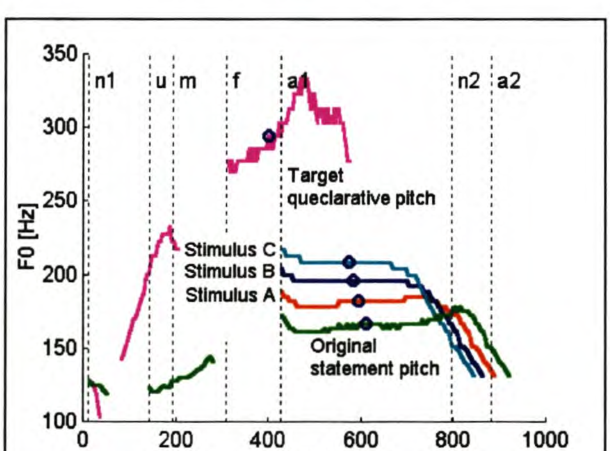


Figure 3.17 Example of the raising of the pitch on the penultimate vowel of the statement form of the utterance /ngumfana/ in Perception Test A2.

Of the five word pairs used in this test, one word pair, /yintloko/, was identified that was suitable for the manipulation of the pitch on the first vowel and also on the final vowel. These extra stimuli were presented as part of Perception Test A2 (manipulation of pitch on penultimate vowel). Figure 3.18 below shows the lowering of the pitch on the first vowel of the queclarative /yintloko/, while Figure 3.19 shows the lowering of the pitch on the final vowel.

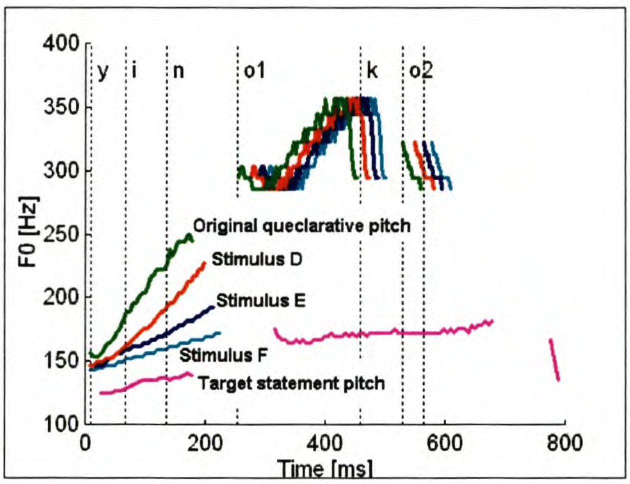


Figure 3.18 Example of the lowering of the pitch on the first vowel of the queclarative of the utterance /yintloko/ in Perception Test A2.

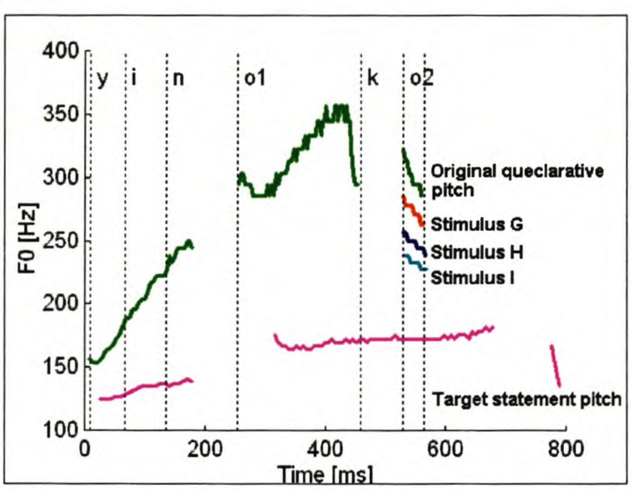


Figure 3.19 Example of lowering of the pitch on the final vowel of the queclarative of the utterance /yintloko/ in Perception Test A2.

Results

The following results were obtained:

1. For 4 of the 5 queclarative forms (A03, E01, I02 and L02) the number of queclarative responses decreased as the pitch on the penultimate vowel was lowered (Table B.10).
2. For 4 of the 5 statements (A03, E01, I02 and K02) the number of Statement responses decreased as the pitch on the penultimate vowel was raised (Table B.10).
3. For this very limited test on the manipulation of the pitch on the first and the final vowel of one utterance (/yintloko/, I02) no significant change in perception was observed for either queclaratives or statements (Tables B.11 and B.12).

The results from this perception test did reveal some change in subjects' perception, albeit not enough to result in any crossover points. It was felt however that more extreme manipulations would assist in identifying a crossover point indicating at what position in the utterance perceptions actually changed from one form to the other.

3.6.3 Experiment A, Test A3: Gating test

Introduction

As may be observed from the introduction to the perceptual test on pitch on the penultimate vowel (Test A2) numerous references have been cited acknowledging the use of pitch in the disambiguation process for many languages of the world. It has also been postulated that different segments of the pitch contour are employed in this disambiguation process. It is apparent that researchers after experimental investigation and perceptual testing of data found in some instances that the entire contour is used Jun and Oh (1996) Korean, Hirschberg and Ward (1992) for English and Thorsen(1980) for Danish. Eduard Hermann (1942) who surveyed 175 languages, and identified a high interrogative pitch in some form or other everywhere pointed out that high pitch does not only imply a rising terminal but can also mean a relatively high overall pitch level. Hadding-Koch (1961) also indicated that a high pitch level was also evident in Swedish questions confirmed by perceptual testing. Hadding-Koch and Studdert-Kennedy (1964) after a further investigation found that listeners made use of the entire F0 contour in identifying questions and statements. After their research in 1971 it was found that the terminal glide was not the unique clue to distinguishing statements from questions but that there are other cues which play a decisive role.

In other studies it was found that specific sections of the contour are employed in this process and predominantly mentioned here is the terminal segment of the utterances. Eady and Cooper(1986) stated that the shape of the contour at the end of the sentence is said to be very important while Lieberman (1967) said that the differentiation process takes place in the final 150-200 ms of phonation.

The use of pitch manipulation in the expression of incredulity has also been documented and found to be valid yet not unique to Kipare (Herman 1996:186). Hirschberg and Ward (1992) show for English that the primary factor distinguishing 'incredulity' from 'uncertainty' is pitch range. Jun and Oh (1994) show that for Korean although there are idiosyncracies between speakers in their expression of incredulity, larger pitch range is one of the factors manipulated to distinguish between incredulity and 'wh' questions. Shen (1990:25) in her research on basic intonation patterns in Mandarin states that

> However, at the starting point the F0 values of all questions are significantly higher than those of statements, regardless of the register at the end point. Interrogative intonation begins at a higher register, usually with a mid-high key.

These conclusions reached by Shen (1990) are consistent with those of De Francis (1963), Ho (1977) and Shen (1985).

For the Bantu languages the final rise has often been nominated as the disambiguating factor. Khumalo (1981:91-92) differentiates between statements and questions by saying that

> 'Statement intonation is unmarked for key (i.e. it is at "average" speaking key), while question intonation is at a much higher key'.

> 'Statement intonation is unmarked for tempo (i.e. it is at "average tempo), while question intonation is at a much faster tempo.'

> In statement intonation the tone of the final syllable of the phonological phrase is lowered. (…) this lowered final tone does not occur in question intonation.

According to Louw (1968:88) the questions are produced within a higher register and at a quicker tempo than statements. This seems to support the impressionistic claims without quantifying the research to any great degree. Khumalo also concurs with Louw with regard to his claims relating to raised register and quicker tempo of questions.

Riordan (1969:16) on the issue of tonal register and movement stated that a question

> …begins on a slightly higher absolute pitch and maintains it throughout.

The most recent research on the process of disambiguation relating to Xhosa has been undertaken by Theron (1991) and Roux (n.d.). Theron's results (1991:16) support this view in general, claiming that an approximate tonal register for a male is 115-200 Hz for statements and between 170-260 Hz for questions. Roux (n.d.) in his research concerning the prosodics of declarative and echo questions in Xhosa and specifically referring to two examples states that

> Based on this data (of which similar variations are attested), it may be argued that resetting of the declining intonation curve in the latter part of the articulation and raising intensity levels at that point, changed a statement into a question.

The many claims point to pitch as being a very significant feature in the differentiation process. Also from the references cited it is apparent that there are differing views as to *where* i.e. at what point within the utterance pitch operates to bring about this process of disambiguation. The question may be posed: is it pitch on the penultimate syllable, in the terminal section of the utterance, over the entire contour or at the very beginning of the utterance which brings about changes in perception?

The acoustic analysis on this corpus of data showed that the pitch on the penultimate vowel was the most significant and it also showed that the register was consistently higher for questions than for statements. Pitch on the penultimate vowel has been perceptually investigated in Test A2 and therefore the significance of the difference in register for these two forms is now investigated using a gating test.

This gating test was designed and compiled with the aim of determining at what point in the utterance people perceived the stimuli presented to them as different forms, namely functioning either as a question or statement. For this purpose Test A3 was conducted, consisting of a total of 202 stimuli which were derived from the original corpus.

The stimuli for the gating test (Cutler and Otake, 1999) were compiled by presenting the word in gates of increasingly longer speech segments. According to Grosjean (1996:602)

> Recognized as a good paradigm when used together with other tasks. It can certainly tell us something about the final outcome of word recognition. Whether it can also do so about intermediate levels (if these exist) remains an empirical issue.

Method used in compilation of stimuli

The corpus used in the compilation of this test are listed in Table 3.13 below.

Table 3.13 Corpus used in Perception Test A3.

Word	Identifier	Word	Identifier
ngumfana	A03	zizifo	H02
ngabafana	B03	yintloko	I02
ngumthi	C01	ziinkomo	J03
yimililo	D02	lulwimi	K02
lilitye	E01	bubuhlanti	L02
ngamafu	F03	kukutya	M01
sisilo	G03		

This gating test was performed to determine how early in the word subjects make the decision as to the question or statement functionally and therefore by implication understand how significant the differences in register were in differentiating between these two forms.

For this test the first gate was taken from the beginning of the word into the first phoneme of the second syllable. Every subsequent gate was taken from the beginning of the word in to the next and every phoneme to the end of the word. For example the word /ngumfana/ consisted of the following gates: /ngum/, /ngumf/, /ngumfa/, /ngumfan/, /ngumfana/ (cut into the centre of the terminal vowel 'a') and /ngumfana/ (the complete word). Figure 3.20 illustrates how gates were cut from the original speech signal, while Table 3.14 shows the duration of gates for the queclarative and statement forms of the word /ngumfana/.

Filename	JMQA5G1W	JMQA5G2W	JMQA5G3W	JMQA5G4W	JMQA5G5W	JMQA5G6W
Gate	ngum	ngumf	ngumfa	ngumfan	ngumfana	ngumfana

Figure 3.20 Example of gates for the word /ngumfana/ consisting of increasingly longer speech segments cut in the centres of phonemes.

Table 3.14 Example of duration of gates for the word /ngumfana/ as presented in Perception Test A3.

	d_{ngum}	d_{ngumf}	d_{ngumfa}	$d_{ngumfan}$	$d_{ngumfana}$
Queclarative	152.1 ms	242.5 ms	381.1 ms	503.4 ms	545.5 ms
Statement	238.8 ms	355.9 ms	599.4 ms	827.2 ms	891.2 ms

Within this test there were repetitions of gates (stimuli) within each word. Two randomly selected gates per word were presented twice to subjects (Table 3.15). However there were never repetitions of the first and last two gates in any of the groups.

Table 3.15 Stimuli presented for the word /ngumfana/ in Perception Test A3.

Randomized		**Sorted**			
		Original question form		**Original statement form**	
Filename	Stimulus Number	Filename	Stimulus Number	Filename	Stimulus Number
JMQA5G5W	1	JMQA5G1W	11	JMSA5G1W	7
JMQA5G4W	2	JMQA5G2W	14	JMSA5G2W	15
JMSA5G5W	3	JMQA5G2W	16	JMSA5G2W	8
JMSA5G4W	4	JMQA5G3W	10	JMSA5G3W	13
JMQA5G3W	5	JMQA5G3W	5	JMSA5G3W	6
JMSA5G3W	6	JMQA5G4W	2	JMSA5G4W	4
JMSA5G1W	7	JMQA5G5W	1	JMSA5G5W	3
JMSA5G2W	8	JMQA5G6W	12	JMSA5G6W	9
JMSA5G6W	9				
JMQA5G3W	10				
JMQA5G1W	11				
JMQA5G6W	12				
JMSA5G3W	13				
JMQA5G2W	14				
JMSA5G2W	15				
JMQA5G2W	16				

The speech segments were cut from the original signal files using the Computerized Speech Laboratory (CSL) program (Figure 3.21). After being cut in the CSL program each gate was then windowed with a raised cosine windowing function to reduce the amplitude at the point of cutting of the signal to suppress any illusionary clicks. Figure 3.22 below illustrates the effect.

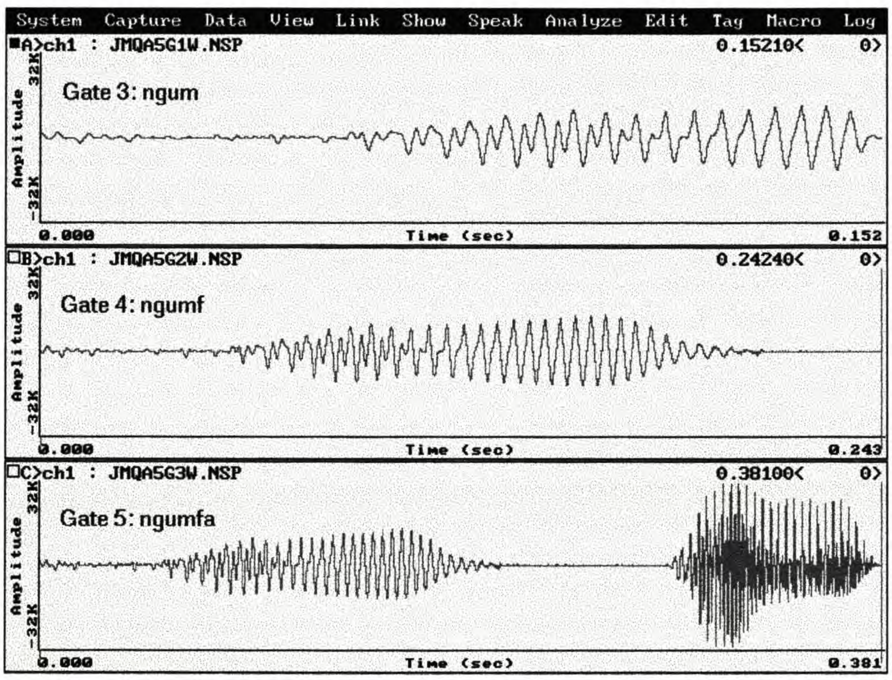


Figure 3.21 Speech signals of gates for the word /ngumfana/ consisting of increasingly longer speech segments.

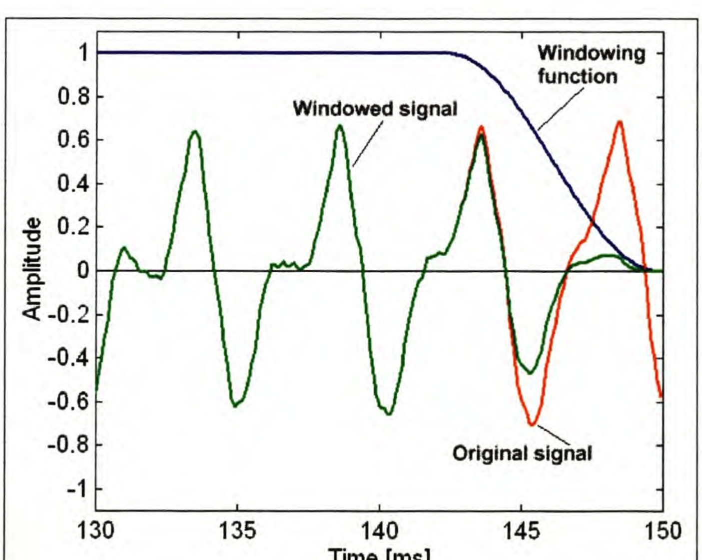


Figure 3.22 The effect of windowing a signal (red) with a raised cosine windowing function (blue) is that the amplitude of the windowed signal (green) is gradually reduced to zero.

Results

The results shown in Table 3.16 below have been derived from the actual responses listed in Table B.13.

Table 3.16 Results of the gating test (Perception Test A3).

	Number of classes to which observation applies	
Minimum temporal information necessary for a majority correct decision	Queclaratives	Statements
First gate (first syllable and part of second consonant)	13 (100%) QA03, QB03, QC01, QD02, QE01, QF03, QG03, QH02, QI02, QJ03, QK02, QL02, QM01	12 (92.3%) SA03, SB03, SC01, SD02, SE01, SF03, SG03, SH02, SI02, SJ03, SK02, SM01
Second gate (first syllable and part of second vowel)	0 (0%)	1 (7.7%) SL02

The following results were obtained in this test:

1. The average percentage of participants that correctly identified the stimuli at the first gate was found to be 92.3%. The first gate comprised the first syllable and part of the second consonant of the utterance (i.e. CVC).
2. All queclaratives and 92.3% (12/13) of statements were correctly identified at the first gate by a majority of subjects. Only one statement form (/bubuhlanti/, SL02) required a longer segment (/bubu/) to be correctly identified (Table 3.16).

3.6.4 Experiment B, Test B1: Duration of penultimate syllable

Introduction

The aim of this test (as for Test A1) was to determine whether duration on the penultimate syllable was perceptually relevant. While the results of Test A1 showed that subjects' perceptions were changed in both directions i.e., from question to statement and vice versa in the manipulations presented as stimuli it was clear that by increasing the degree of manipulation on these signals stronger evidence of perceptual changes may be extracted. Hence, this test focussed on increased degrees of manipulation.

Method used in the compilation of stimuli

The corpus used in the compilation of this test are listed in Table 3.17 below.

Table 3.17 Corpus used in Perception Test B1.

Word	Identifier	Word	Identifier
ngumfana	A03	zizifo	H02
ngabafana	B03	yintloko	I02
ngumthi	C01	ziinkomo	J03
yimililo	D02	lulwimi	K02
lilitye	E01	bubuhlanti	L02
ngamafu	F03	kukutya	M01
sisilo	G03		

The number of stimuli presented for perceptual testing totalled 143 in this test. Five manipulations of the queclarative and four of the statement penultimate syllables were presented to listeners. Within this test there were no repetitions of stimuli included in the randomized sets for each group which comprised the queclarative and its statement counterpart for each individual class.

Table 3.18 Stimuli presented for the word /ngumfana/ in Perception Test B1.

Randomized		Sorted			
		Original question form		Original statement form	
Filename	Stimulus Number	Filename	Stimulus Number	Filename	Stimulus Number
JMCVSA3A	1	JMCVQA3	7	JMCVSA3	6
JMCVSA3B	2	JMCVQA3A	3	JMCVSA3A	1
JMCVQA3A	3	JMCVQA3B	8	JMCVSA3B	2
JMCVQA3D	4	JMCVQA3C	11	JMCVSA3C	5
JMCVSA3C	5	JMCVQA3D	4	JMCVSA3D	9
JMCVSA3	6	JMCVQA3E	10		
JMCVQA3	7				
JMCVQA3B	8				
JMCVSA3D	9				
JMCVQA3E	10				
JMCVQA3C	11				

The method used here for the compilation of stimuli was exactly the same as that used for Test A1 using the ASL program. However, two further manipulations were included for the queclarative. Only one additional manipulation was done for the statement forms as the penultimate syllables would otherwise be too short to be comprehensible.

Figure 3.23 shows how the range of duration values tested was expanded in Experiment B beyond the target values used in Experiment A.

The duration of the penultimate vowels of the *k*th stimulus for queclarative and statement forms, d_{Qa1}^{k} and d_{Sa1}^{k} respectively, were calculated as follows:

(3.14) $d_{Qa1}^{k} = d_{Qa1} + k\Delta d_{Qa1}$ k = 0,1,2,3,4,5

(3.15) $d_{Sa1}^{k} = d_{Sa1} + k\Delta d_{Sa1}$ k = 0,1,2,3,4

The numerical values of these parameters are exemplified in Tables 3.19 and 3.20 below.

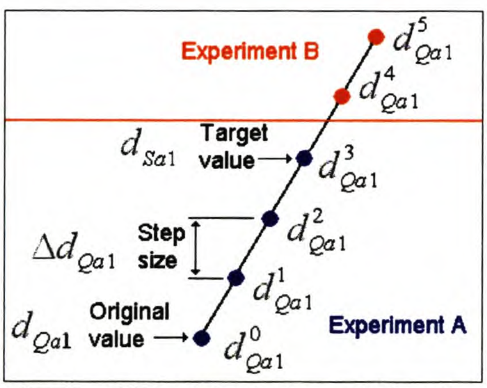


Figure 3.23 Calculation of duration changes for the queclarative form of the word /ngumfana/ in Perception Test B1.

Table 3.19 Duration changes for the queclarative form of the word /ngumfana/ in Perception Test B1.

Stimulus file	k	$k\Delta d_{Qa1}$	d_{Qa1}^{k}
JMCVQA3	0	0 ms	177.7 ms
JMCVQA3A	1	64.2 ms	241.9 ms
JMCVQA3B	2	128.4 ms	306.1 ms
JMCVQA3C	3	192.6 ms	370.3 ms
JMCVQA3D	4	256.8 ms	434.5 ms
JMCVQA3E	5	321 ms	498.7 ms

Table 3.20 Duration changes for the statement form of the word /ngumfana/ in Perception Test B1.

Stimulus file	k	$k\Delta d_{Sa1}$	d_{Sa1}^{k}
JMCVSA3	0	0 ms	370.2 ms
JMCVSA3A	1	64.2 ms	306 ms
JMCVSA3B	2	128.4 ms	241.8 ms
JMCVSA3C	3	192.6 ms	177.6 ms
JMCVSA3D	4	256.8 ms	113.4 ms

Results

The results shown in Tables 3.21 and 3.22 below have been derived from the actual responses listed in Table B.14.

Table 3.21 Results of the manipulation of the duration of the penultimate syllable in Perception Test B1.

	Manipulation and number of noun classes to which observation applies			
Observation	Duration of penultimate syllable of a question form increased		Duration of penultimate syllable of a statement form decreased	
Experiment	Experiment B	A	Experiment B	A
•Crossover point observed •Significant change in perception	7 (53.9%) QA03, QB03, QE01, QF03, QG03, QI02, QL02	3	0 (0%)	0
•Crossover point possible after further manipulation •Some change in perception	2 (15.4%) QC01, QD02	0	5 (38.5%) SA03, SB03, SF03, SG03, SI02	0
•Crossover point improbable •Some change in perception	3 (23.1%) QJ03, QK02, QM01	4	3 (23.1%) SD02, SE01, SM01	5
•Crossover point improbable •No change in perception	1 (7.7%) QH02	6	5 (38.5%) SC01, SH02, SJ03, SK02, SL02	8

Table 3.22 Detail of crossover points achieved in Perception Test B2.

	Original	**Stimulus A**	**Stimulus B**	**Stimulus C**	**Stimulus D**	**Stimulus E**
Queclarative – **ngumfana** (QA03)						
d_{a1}	177.7 ms	240.7 ms	303.7 ms	370.2 ms	433.2 ms	496.2 ms
% change in d_{a1}		35.5%	70.9%	108.3%	143.8%	179.2%
% queclarative	90.5%	95.2%	49.2%	30.2%	36.5%	33.3%
Comment	smQ	smQ	mQ	smS, cop	smS	smS
Queclarative – **ngabafana** (QB03)						
d_{a3}	127.7 ms	190.4 ms	253.1 ms	319.1 ms	378.5 ms	441.2 ms
% change in d_{a3}		49.1%	98.2%	149.9%	196.4%	245.5%
% queclarative	79.4%	84.1%	60.3%	49.2%	57.1%	36.5%
Comment	smQ	smQ	mQ	mQ	mQ	mS, cop
Queclarative – **lilitye** (QE01)						
d_{i2}	158.9 ms	221.6 ms	290 ms	358.4 ms	421.1 ms	489.5 ms
% change in d_{i2}		39.5%	82.5%	125.6%	165%	208.1%
% queclarative	87.3%	82.5%	63.5%	49.2%	41.3%	39.7%
Comment	smQ	smQ	smQ	Q=S, cop	mS	mS

Table 3.22 Detail of crossover points achieved in Perception Test B2 (continued).

	Original	**Stimulus A**	**Stimulus B**	**Stimulus C**	**Stimulus D**	**Stimulus E**
Queclarative – **ngamafu** (QF03)						
d_{a2}	176.9 ms	247.7 ms	318.5 ms	389.3 ms	460.1 ms	530.9 ms
% change in d_{a2}		40%	80%	120.1%	160.1%	200.1%
% queclarative	79.4%	77.8%	49.2%	54%	30.2%	41.3%
Comment_1	smQ	smQ	Q=S	mQ	smS, cop	mS
Queclarative - **sisilo** (QG03)						
d_{i2}	136.3 ms	196.3 ms	250.3 ms	304.3 ms	362 ms	423 ms
% change in d_{i2}		44%	83.6%	123.3%	165.6%	210.3%
% queclarative	93.7%	88.9%	47.6%	61.9%	46%	54%
Comment	smQ	smQ	mS	smQ	mS, cop	mQ
Queclarative - **yintloko** (QI02)						
d_{o1}	205.9 ms	267.9 ms	329.9 ms	391.9 ms	453.9 ms	515.9 ms
% change in d_{o1}		30.1%	60.2%	90.3%	120.4%	150.6%
% queclarative	85.7%	92.1%	47.6%	42.9%	55.6%	33.3%
Comment	smQ	smQ	mS, cop	mS	mQ	smS
Queclarative - **bubuhlanti** (QL02)						
d_a	178.4 ms	226.5 ms	270.9 ms	319 ms	363.4 ms	407.8 ms
% change in d_a		27%	51.8%	78.8%	103.7%	128.6%
% queclarative	82.5%	84.1%	79.4%	61.9%	61.9%	47.6%
Comment	smQ	smQ	smQ	mQ	mQ	mS, cop

The following abbreviations are used:

smQ significant majority Queclarative

smS significant majority Statement

mQ majority Queclarative

mS majority Statement

cop crossover point

From the tables above the following observations have been made:

1. Crossover points and a significant change in perception was achieved for 53.9% (7/13) of the original queclarative forms that were lengthened to emulate statements. The number of crossover points increased in Experiment B compared to Experiment A (Table 3.21).
2. The shortening of original statements did not result in any crossover points or definite changes in perception. The same result was obtained in Experiment A (Table 3.21).
3. For 26.9% (7/26) of all queclarative and statement forms some change in subjects' perception could be observed, although not enough to result in a crossover point. In

these cases the possibility existed of achieving a crossover point after presentation of more manipulations. For another 23.1% (6/26) some change in perception was evident, but the occurrence of crossover points would be fairly improbable due to the small change in perception over the range of stimuli presented (Table 3.21).

4. Overall, for 92.3% (12/13) of queclaratives and 61.5% (8/13) of statements a change in perception, be it large or small, was achieved (Table 3.21).
5. In 23.1% (6/26) no change in perception was achieved. This figure is lower than the 53.9% in Experiment A, indicating that the larger range of stimuli presented in Experiment B evoked more meaningful perceptual responses (Table 3.21).
6. The average percentage of undecided responses for all stimuli was 0.9%. The highest number of undecided responses was 4.8%.

Discussion

In a number of instances in both Experiment A and B no significant change in perception was achieved. This may be attributed to the following factors:

1. The method of increasing the duration in the ASL program led to audible unnaturalness in the manipulated signal when the factor of increase was large. This limits the amount by which a queclarative form could be lengthened while maintaining the speech quality.
2. When a statement form was to be shortened, the limiting factor was the actual duration of the penultimate syllable. This syllable cannot be shortened to the point that it is no longer audible.
3. If the duration was changed in the ASL program the duration *and* the pitch was changed. The average pitch, however, would not change that much and in effect *only* the duration was changed perceptually. In contrast, in the tests where the pitch was manipulated *both* the pitch and the duration changed due to methods employed. If pitch was raised the duration decreased implicitly. Therefore, if the pitch of a statement form was raised, the duration decreased. Thus *two* parameters were adjusted, both in the 'direction' of a statement. Similarly, the duration of a question form would increase if the pitch was lowered. This means that the tests where the pitch was manipulated may have yielded slightly stronger changes in perception compared to the tests where the duration was modified. To modify only pitch, one needs to correct the duration after the pitch manipulation.

3.6.5 Experiment B, Test B2: Pitch on the penultimate vowel

Introduction

The aim of this test was to determine whether listeners perception of the identification of statements and questions could be shifted if the pitch on the penultimate vowels of queclaratives was reduced towards the level of the statement and vice versa. The subjects' reactions to these stimuli were expected to show at what point of manipulation queclaratives were perceived as statements and statements as queclaratives.

Method used in compilation of stimuli

The corpus used in the compilation of this test are listed in Table 3.23 below. In total 48 stimuli were presented to subjects.

Table 3.23 Corpus used in Perception Test B2.

Word	Identifier
ngumfana	A03
lilitye	E01
ngamafu	F03
yintloko	I02
lulwimi	K02
bubuhlanti	L02

In this test there were no repetitions of stimuli. Repetition of stimuli in Test A2 made it more difficult to analyze the results due to inconsistent responses from subjects and for this reason the same procedure was not followed.

The table below represents the mean pitch values on the original and manipulated penultimate vowel of the word /ngumfana/ as presented to subjects.

Table 3.24 Pitch manipulation on the penultimate vowel of /ngumfana/ as presented in the Perception Test B2.

	Original Pitch	Stimulus D Pitch	Stimulus E Pitch	Stimulus F Pitch	Stimulus G Pitch
Question	297 Hz	252 Hz	219 Hz	193 Hz	168 Hz
% change		-15%	-26%	-35%	-43%
Statement	168 Hz	198 Hz	223 Hz	251 Hz	273 Hz
% change		18%	33%	49%	63%

The pitch on the penultimate vowels of questions was reduced in four steps of approximately 30 Hz towards the level of statements. Similarly for the statements the pitch on the penultimate vowels was raised in four steps. The quality of the speech after manipulation with the interpolation/deletion method used in the first experiment proved to be a limiting factor. From the results of that test it was clear that a better pitch manipulation algorithm was required in order to change the pitch of a speech segment to a greater degree while maintaining the quality. The *Time Domain Pitch Synchronous Overlap-Add* (TD-PSOLA) technique (Moulines and Charpentier, 1990) was therefore used to modify the pitch contour of stimuli. In order to perform PSOLA the pitch periods must be marked in the original speech file. The process to modify the pitch contour of a voiced segment firstly requires that a set of short term signals be obtained by windowing each pitch period with a Hanning window centered around pitch marks. These short term signals are then shifted in time so that the delays between successive short term signals correspond to the pitch periods of the desired pitch contour. Finally these overlapping short term signals are then added to yield the manipulated speech signal. The TD-PSOLA process is illustrated in Figure 3.24 below.

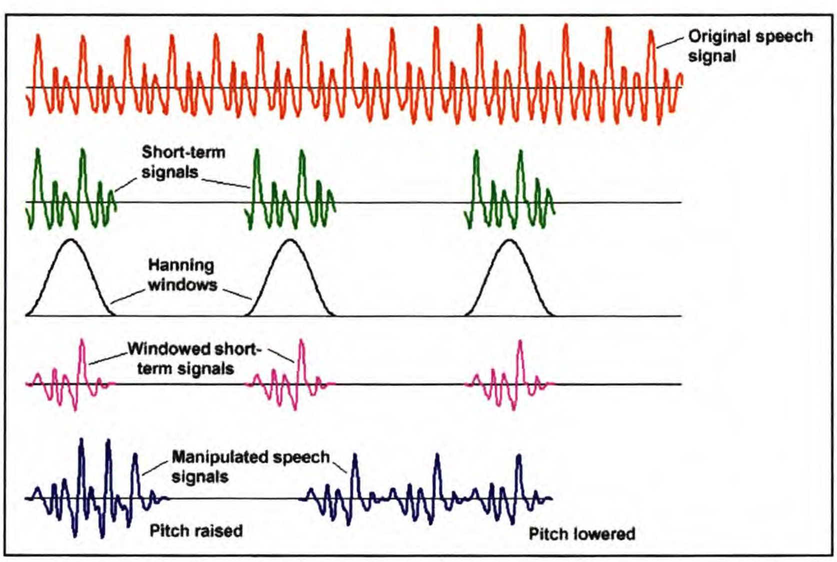


Figure 3.24 The TD-PSOLA process of manipulation of the pitch of a speech segment.

The specification[3] of the modified pitch contour was done in MATLAB. The pitch contour of the original signal and that of the target signal were plotted simultaneously on the screen. The pitch contours of all the manipulations could be shown. The researcher had to redraw that part of the original pitch contour that had to be modified with the mouse. The TD-PSOLA algorithm was then used to modify the pitch of the

[3] Developed by Mr J.A.N. Louw, RUEPUS, 1997.

original signal accordingly. Note that in Figure 3.25 below it was not possible to change only the pitch on the penultimate vowel but that the pitch on the final syllable was modified as well to maintain continuity of the pitch contour.

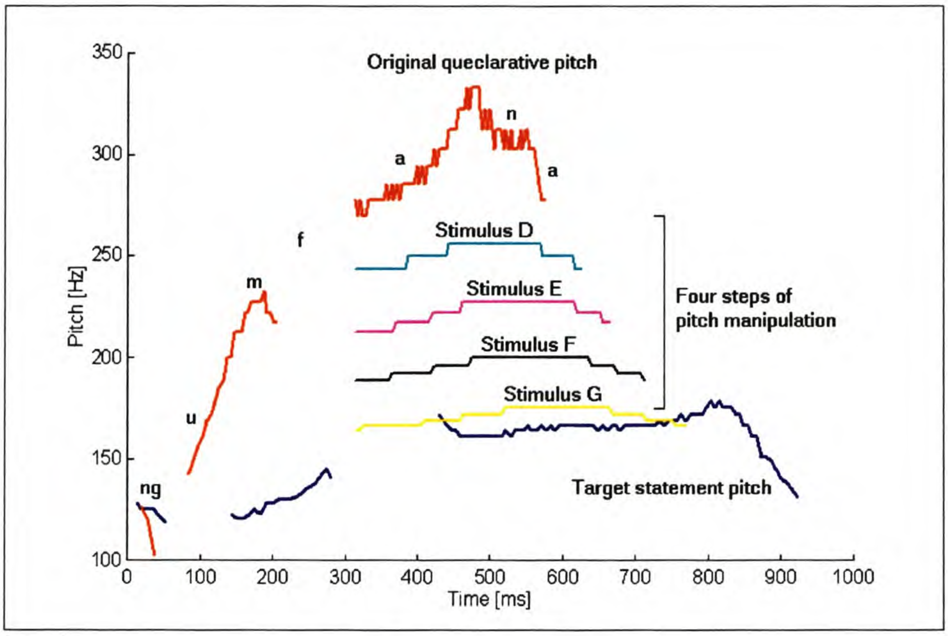


Figure 3.25 Example of lowering of the pitch on the penultimate vowel of the question form of the utterance /ngumfana/ in Perception Test B2.

Results

The results shown in Tables 3.25 and 3.26 below have been derived from the actual responses listed in Table B.15.

Table 3.25 Results of the manipulation of the pitch on the penultimate vowel in Perception Test B2.

	Manipulation and number of classes to which observation applies	
Observation	Pitch on penultimate vowel of a question form decreased	Pitch on penultimate vowel of a statement form increased
•Cross over point observed •Significant change in perception	2 (33.3%) QA03, QL02	2 (33.3%) SA03, SL02
•Cross over point possible after further manipulation •Some change in perception	4 (66.7%) QE01, QF03, QI02, QK02	4 (66.7%) SE01, SF03, SI02, SK02

Table 3.26 Detail of crossover points achieved in Perception Test B2.

	Original	**Stimulus D**	**Stimulus E**	**Stimulus F**	**Stimulus G**
Queclarative – **ngumfana** (QA03)					
p_{a1}	297 Hz	252 Hz	219 Hz	193 Hz	168 Hz
% change in p_{a1}		-15%	-26%	-35%	-43%
% queclarative		77.8%	65.1%	38.1%	30.2%
Comment		smQ	smQ	mS, cop	smS
Statement – **ngumfana** (SA03)					
p_{a1}	168 Hz	198 Hz	223 Hz	251 Hz	273 Hz
% change in p_{a1}		18%	33%	49%	63%
% statement		88.9%	76.2%	46%	22.2%
Comment		smS	smS	mQ, cop	smQ
Queclarative – **bubuhlanti** (QL02)					
p_a	285 Hz	270 Hz	241 Hz	219 Hz	193 Hz
% change in p_a		-5%	-15%	-23%	-32%
% queclarative		85.7%	93.7%	58.7%	41.3%
Comment		smQ	smQ	mQ, sc	mS, cop
Statement – **bubuhlanti** (SL02)					
p_a	151 Hz	182 Hz	208 Hz	233 Hz	255 Hz
% change in p_a		21%	38%	54%	69%
% statement		84.1%	73%	49.2%	41.3%
Comment		smS	smS	Q=S, cop	mQ

The following abbreviations are used:

smQ significant majority Queclarative

smS significant majority Statement

mQ majority Queclarative

mS majority Statement

cop crossover point

The following results were obtained when the pitch on the penultimate vowel was manipulated:

1. A definite crossover point was observed for 33.3% (2/6) of the noun classes in the case of the question form (Table 3.25). For the corresponding statement form this phenomenon was apparent in the same proportions. The smallest percentage pitch change required to achieve a crossover point was 32% (Table B.5).
2. For the other 4 noun classes (66.7%) changes in perception were evident, albeit not strong enough to result in crossover points (Table 3.25).
3. The mean response time increased as the amount of pitch manipulation increased (Table B.15).

4. The average percentage of undecided responses for all stimuli was 0.7%. The highest number of undecided responses was 4.8%.

Discussion

The manipulation of pitch on the penultimate vowel resulted in definite changes in perception for all words and crossover points were observed for 4 of the 12 words presented in the test (Table 3.25). It was therefore possible to shift subjects' perceptions of the identity of question and statement forms by lowering the pitch on the penultimate vowels of queclaratives and in the case of statements by raising the pitch. The magnitude of possible pitch changes was limited however, not only by constraints imposed by the manipulation method, but also in terms of naturalness. It was found in the test that the mean response time increased as the amount of pitch manipulation increased. This indicates that the subjects were more confused and took longer to respond when the pitch changes were larger. This as indicated could be due to the unnaturalness of the increased manipulation yet it may have also resulted from confusion arising from possible decisions which had already been made at the beginning of the utterance.

3.6.6 Experiment B, Test B3: Pitch on the first vowel

Introduction

This perception test was included and based on the early recognition of tokens in the gating test and the higher register of queclaratives. Due to the resounding results of the gating tests it is postulated that listeners use the pitch on the first vowel to differentiate between queclaratives and statements The aim of this test was to determine whether listeners' perceptions could be changed if only the pitch on the first vowel is raised in the case of statements and lowered in the case of queclaratives.

Cutler and Otake (1999) in their contribution conducted three experiments on the use of pitch-accent information in spoken-word recognition by Tokyo Japanese listeners. They found that the acoustic cues to pitch accent distinctions are available within a single syllable. In the second experiment, Cutler and Otake presented listeners with initial fragments of words and asked them to guess the words. It appeared that the listeners guesses had overwhelmingly the same initial accent structure as the presented word in spite of the listener only having heard the first consonant and a small portion of the following vowel. This meant that listeners could effectively exploit the cues to accent pattern available even in the first syllable of a word.

According to Otake and Cutler (1999:232)

> Together these results provide strong evidence that Tokyo Japanese listeners make early and effective use of pitch-accent information in recognising spoken words. The suprasegmental structure of Tokyo Japanese words is, like the segmental structure, of use in narrowing down the set of potential candidates of lexical recognition.

> Research in human spoken-word recognition has shown that listeners exploit relevant phonetic information as soon as it becomes available; any inter-word distinctions in the initial portions of words could therefore be of great use to listeners, enabling drastic reductions in the set of possible candidate words for recognition (Otake and Cutler, 1999:231).

From this research it becomes evident that word recognition using pitch accent information is made early in the word. It was also found that pitch-accent distinctions are available within a single syllable and that listeners' guesses were accurate in recognition in spite of the listeners only having heard the first consonant and a small portion of the following vowel. Considering these findings of this most recent research on pitch-accent and the early recognition of statement and question forms in the gating test the question posed is: could it not be that listeners also use pitch on the first syllable to distinguish between question and statement forms?

Method used in compilation of stimuli

The corpus used in the compilation of this test are listed in Table 3.27 below.

Table 3.27 Corpus used in Perception Test B3.

Word	Identifier	Word	Identifier	Word	Identifier	Word	Identifier
ngumfana	A03	lilitye	E01	yintloko	I02	kukutya	M01
ngabafana	B03	ngamafu	F03	ziinkomo	J03		
ngumthi	C01	sisilo	G03	lulwimi	K02		
yimililo	D02	zizifo	H02	bubuhlanti	L02		

The method used for the manipulation of pitch on the first vowels was the same as that used for manipulation of pitch on the penultimate vowel. Three manipulations were made on each queclarative and statement totalling 78 stimuli for presentation.

The step sizes of manipulation varied between 8 Hz and 30 Hz, depending on the pitch difference between the queclarative and the statement forms.

Table 3.28 Pitch manipulations on the first vowel of /ngumfana/ as presented in Perception Test B3.

	Original Pitch	Stimulus A Pitch	Stimulus B Pitch	Stimulus C Pitch
Question	160 Hz	148 Hz	139 Hz	130 Hz
% change		-8%	-13%	-19%
Statement	122 Hz	132 Hz	140 Hz	150 Hz
% change		8%	15%	23%

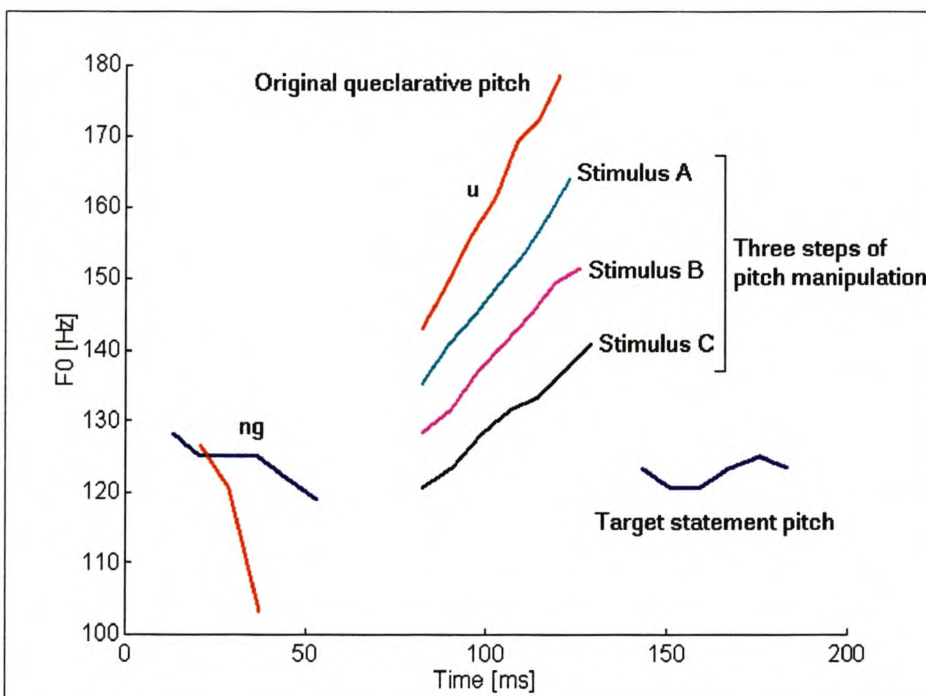


Figure 3.26 Example of lowering the pitch on the first vowel of the question form of the utterance /ngumfana/.

Results

The results shown in Tables 3.29 and 3.30 below have been derived from the actual responses listed in Table B.16.

Table 3.29 Results of the manipulation of the pitch on the first vowel in Perception Test B3.

Observation	Pitch of first vowel of a question form decreased	Pitch of first vowel of a statement form increased
• Significant change in perception • Cross over point observed	3 (23.1%) QA03, QF03, QI02	1 (7.7%) SK02
• Some change in perception • Cross over point possible after further manipulation	6 (46.2%) QB03, QC01, QH02, QJ03, QK02, QM01	7 (53.8%) SA03, SB03, SD02, SE01, SF03, SI02, SM01
• No change in perception	2 (15.4%) QE01, QL02	1 (7.7%) SG03
• Misidentified	2 (15.4%) QD02, QG03	4 (30.8%) SC01, SH02, SJ03, SL02

Table 3.30 Detail of crossover points achieved in Perception Test B3.

	Original	**Stimulus A**	**Stimulus B**	**Stimulus C**
Queclarative – **ngumfana** (QA03)				
p_{a1}	160 Hz	148 Hz	139 Hz	130 Hz
% change in p_{a1}		-8%	-13%	-19%
% queclarative		55.6%	38.1%	23.8%
Comment		smQ	mS, cop	smS
Statement – **ngamafu** (QF03)				
p_{a2}	189 Hz	172 Hz	164 Hz	154 Hz
% change in p_{a2}		-9%	-13%	-19%
% queclarative		58.7%	65.1%	42.9%
Comment		mQ	smQ	mS, cop
Queclarative – **yintloko** (QI02)				
p_{o1}	204 Hz	196 Hz	185 Hz	174 Hz
% change in p_{o1}		-4%	-9%	-15%
% queclarative		58.7%	58.7%	42.9%
Comment		mQ	mQ	mS, cop
Statement – **lulwimi** (SK02)				
p_{i1}	262 Hz	270 Hz	277 Hz	286 Hz
% change in p_{i1}		3%	6%	9%
% statement		50.8%	50.8%	39.7%
Comment		mS	mS	mQ, cop

The following abbreviations are used:

smQ significant majority Queclarative

smS significant majority Statement

mQ majority Queclarative

mS majority Statement

cop crossover point

The results that were obtained when the pitch on the first vowel was manipulated are as follows:

1. A crossover point and subsequent significant change in perception was achieved for three queclarative forms (23.1%) and one statement form (7.7%).
2. When the pitch on the first vowel of queclarative forms was lowered a perceptual change was achieved in 69.3% (9/13) of the cases (Table 3.29). The corresponding figure for statements was slightly lower, namely 61.6% (8/13).
3. In general the mean response time increased as the degree of pitch manipulation increased (Table B.16).
4. The average percentage of undecided responses for all stimuli was 6.9%. The highest number of undecided responses was 25.4%.
5. No significant trend in the perceptual responses of subjects was observed for 3 forms (Table 3.29).
6. Two queclarative forms and four statement forms were misidentified, meaning that a queclarative form was identified as a statement form and vice versa (Table 3.29). This is not unexpected given the short duration and manipulation of the stimuli. In all these cases all stimuli were incorrectly identified regardless of the amount of manipulation.

Discussion

These results reveal once again the importance of pitch in the decoding or disambiguation between statements and queclaratives. This test indicates that the decision on differentiating between statements and questions was made as early as the first syllable. Subjects' decisions on the queclarative or statement identity of stimuli could be changed or even reversed by raising or lowering the pitch on the first vowel. By implication therefore the pitch on this syllable already provides a significant disambiguating cue. A number of word groups were *misidentified*, meaning that a question form was identified as a statement form and vice versa. This is not unexpected given the short duration and manipulation of the stimuli. In these instances of misidentification all stimuli were incorrectly identified regardless of the amount of manipulation. Finally, it was found that the mean response time increased in general as the degree of pitch manipulation increased (Table B.16). This once again indicated that the subjects took longer to respond possibly due to the unnaturalness introduced by the pitch manipulation.

3.6.7 Experiment B, Test B4: Gating test

Introduction

The aim of this gating test was to determine at what point in the utterance listeners' perceived the stimulus as different forms namely either as a statement or as a question. In the first gating test (Test A3) gates derived from both statements and queclaratives were correctly identified by a high percentage of subjects (in excess of 85% and often over 90%). Very early recognition was also evident within the representation of the first two gates. However, a further gating test, Test B4, was necessary with even shorter gates to endorse the results from Test A3 and to increase the range of perceptual data obtained. The second gating test (B4) included two very short gates, while the repetition of gates and the final gate representing the entire word were excluded.

Method used in the compilation of stimuli

The corpus used in the compilation of this test are listed in Table 3.31 below.

Table 3.31 Corpus used in Perception Test B4.

Word	Identifier	Word	Identifier
ngumfana	A03	zizifo	H02
ngabafana	B03	yintloko	I02
ngumthi	C01	ziinkomo	J03
yimililo	D02	lulwimi	K02
lilitye	E01	bubuhlanti	L02
ngamafu	F03	kukutya	M01
sisilo	G03		

The methods used for the preparation of stimuli for Perception Test A3 were applied again in this test to obtain the extra two gates. The stimuli for the word /ngumfana/ for example consisted of the following gates: /ng/, /ngu/, /ngum/, /ngumf/, /ngumfa/, /ngumfan/ and eventually /ngumfana/. Table 3.32 shows the duration of these gates.

Table 3.32 Duration of gates for the word /ngumfana/ as presented in Test B4.

	d_{ng}	d_{ngu}	d_{ngum}	d_{ngumf}	d_{ngumfa}	$d_{ngumfan}$	$d_{ngumfana}$
Queclarative	78.2 ms	96.7 ms	152.1 ms	242.5 ms	381.1 ms	503.4 ms	545.5 ms
Statement	137.4 ms	162.4 ms	238.8 ms	355.9 ms	599.4 ms	827.2 ms	891.2 ms

Results

The results shown in Tables 3.33 and 3.34 below have been derived from the actual responses listed in Table B.17.

Table 3.33 Results of the gating test (Test B4).

	Number of classes to which observation applies	
Minimum temporal information necessary for a majority correct decision	Queclaratives	Statements
First gate (first consonant)	4 (30.8%) QJ03, QK02, QL02, QM01	11 (84.6%) SA03, SB03, SC01, SD02, SE01, SF03, SG03, SH02, SI02, SK02, SM01
Second gate (first consonant and part of first vowel)	8 (61.5%) QA03, QB03, QC01, QE01, QF03, QG03, QH02, QI02	0 (0%)
Longer gates (more than first syllable)	1 (7.7%) QD02	2 (15.4%) SJ03, SL02

Table 3.34 below shows the first gate for each word for which a majority correct response was obtained. The percentage correct responses is shown as well.

Table 3.34 First gate with a majority correct responses (Test B4).

Queclaratives							
	QA03	QB03	QC01	QD02	QE01	QF03	QG03
Gate	/ngu/	/nga/	/ngu/	/yim/	/li/	/nga/	/si/
% queclarative	44.4%	54%	76.2%	52.4%	55.6%	65.1%	55.6%
	QH02	QI02	QJ03	QK02	QL02	QM01	
Gate	/zi/	/yi/	/z/	/l/	/b/	/k/	
% queclarative	79.4%	68.3%	57.1%	65.1%	54%	74.6%	
Statements							
	SA03	SB03	SC01	SD02	SE01	SF03	SG03
Gate	/ng/	/ng/	/ng/	/y/	/l/	/ng/	/s/
% statement	69.8%	65.1%	69.8%	65.1%	50.8%	76.2%	71.4%
	SH02	SI02	SJ03	SK02	SL02	SM01	
Gate	/z/	/y/	/ziinko/	/l/	/bubu/	/b/	
% statement	81%	68.3%	85.7%	55.6%	81%	54%	

The following results were obtained from the gating test:

1. In 84.6% (11/13) of the cases the statements were correctly identified as statements by the majority of participants when only the first gate, i.e. the first consonant, was presented. Two instances required longer segments (more gates) for correct identification (Table 3.33).

2. The queclaratives were correctly identified on presentation of the first gate in 30.8% (4/13) of the cases (Table 3.33). 61.5% (8/13) of the queclaratives required the second gate, i.e. the first consonant and part of the first vowel, in order to make the correct decision (Table 3.33).
3. As the duration of the gates increased the mean response time decreased while the percentage of correct responses increased which indicates increased confidence (Table B.17).
4. The average percentage of undecided responses for all stimuli was 3.7%. The highest number of undecided responses was 19.1%.

Discussion

From the results of both the first and the second gating tests, it is evident that the differentiation between, and recognition of, both statements and queclaratives was achieved by providing a very limited amount of the speech signal, in fact this occurred as early as after the initial gate which was comprised of the first consonant. These results therefore suggest that participants' perceptions and linguistic decisions on these particular forms were made as from the first and second phonemes of the utterance. The results also showed that the mean response time decreased and that the number of correct responses increased as the gates grew longer (Table B.17). This indicated participants' growing confidence in their decisions on the longer gates. The stimuli were comprised of the original signals with no pitch or duration manipulations aside from the cutting of signals and the windowing at the cutting point. The first gates retained the pitch, duration, loudness and vowel quality characteristics of the first syllables from which they were cut. In the acoustic analysis of the data prior to the compilation of the perceptual test it was found that all queclaratives as opposed to statements had a raised register (Jones et al., 1998). The pitch on the first vowel was never the most significant disambiguating feature, however it was of secondary importance according to the statistics (see §2.6.3). Neither the duration nor the loudness on the first syllable was found to be consistently significant. In view of this one could postulate that indeed it was this pitch level (register) which could then be considered as a cue prompting the differentiation between statements and queclaratives.

Such a resounding result drew attention to the importance of pitch in the process of decoding the message and differentiating between statements and queclaratives. However, it must be conceded that this process of experimentation took place within a specific environment where listeners were expecting to hear either a statement or a question. This therefore does not imply that in a normal conversational setting this would also apply. However it does illustrate the sensitive nature of the auditory process amongst humans which may assist in interpretation or decoding of data.

3.7 Results and evaluation

A summary is now provided of the results from the statistical analysis of the listeners' responses and an evaluation of these with respect to research in other languages employing an experimental approach.

1. The results of the gating test showed very early recognition of queclarative and statement forms. In fact in 57.7% of instances only one phoneme was necessary for positive identification. These results therefore suggest that peoples' perceptions and linguistic decisions on these particular forms are made as from the first and second phonemes of the utterance. In the process of word recognition (Cutler and Otake, 1999) conducted three experiments on the use of pitch-accent information in spoken word recognition in Tokyo Japanese. It was found that listeners exploited accent pattern available even in the first syllable of a word. These results provide strong evidence that Tokyo Japanese listeners make early and effective use of pitch-accent information in recognizing spoken words. It seems therefore that in view of the early recognition of statements and questions, from the presentation of the first phoneme one could postulate that Xhosa listeners similarly make use of pitch in recognizing the different forms of questions and statements.

2. The manipulation of the pitch on the first vowel also showed very clear changes in the perception of people which once again occurred at the very beginning of the word. Both these perception tests therefore indicate that the decisions on differentiating statements from questions are made very early in the utterance and may both be related to the level of pitch on the first syllable of the utterance (Cutler

and Otake, 1999). Other studies which indicate this early differentiation using pitch as the cue include those of Shen (1990), De Fancis (1963), Shen (1985) and Ho (1977). Shen (1990:27) stated that the feature which differentiates assertive from interrogative intonation is the difference in register at the starting point. In Shen's personal communication with Bolinger (1990:28) he stated that for English the same might apply as for Mandarin should a large enough sample be collected in which it would turn out that questions may be signalled by being higher in pitch than by terminals. Jun and Oh (1996:39) stated that for example in English it has been claimed that incredulity readings have been associated with changes in parameters like duration and amplitude and a larger pitch range in production compared to uncertainty readings. However it was found that only large pitch range was perceptually salient in the perception of incredulity (Hirschberg and Ward, 1992). Jun and Oh (1996) also found that for all speakers, the incredulity showed a larger pitch range than that of the other two types of questions researched. All these results indicate that a difference in register is used in the disambiguation process, where some scholars have even identified this register difference at the starting point of the utterance (De Fancis (1963), Ho (1977), J Shen (1985) and Shen (1990). The results from the gating test and the manipulation of pitch on the initial vowel for this corpus of Xhosa data seem to coincide with the results from these other languages. The reduced mean response time and the increased number of correct responses as the gates increased in duration show peoples' growing confidence in their decisions as more of the utterance is presented. The stimuli were comprised of the original signals with no pitch or duration manipulations aside from the cutting of signals and the windowing at the cutting point. The first gates retain the pitch, duration, loudness and vowel quality characteristics of the first syllables from which they were cut. In the acoustic analysis of the data prior to the compilation of the perceptual test there was evidence of a consistently raised register on queclaratives as opposed to statements. The pitch on the first vowel proved to be a significant disambiguating factor, definitely not as important as the pitch on the penultimate vowel, but of some significance in the majority of the cases. Neither the duration nor the loudness on the first syllable was found to be consistently significant. In view of this one could postulate that indeed it was this pitch level (register) which could then be considered as a cue prompting the differentiation between statements and queclaratives.

3. The manipulation of pitch on the penultimate vowel also showed very definite changes in perception. It is possible to shift subjects' perceptions of the identity of queclarative and statement forms by lowering the pitch on the penultimate vowel of queclaratives and by raising the pitch in the case of statements. Studdert-Kennedy (1973) in research on the intonation contour said that the single most powerful cue for question/statement identity was the terminal glide. Muira and Hara (1995) in their study of rhetorical questions in Osaka Japanese found that the raising of the F0 sentence finally changed peoples' perceptions from /wh/ to rhetorical questions. Eady and Cooper (1986) also state that the end of the shape of the contour at the end of the sentence is very important in distinguishing between statement and question form. Lieberman (1967) also makes claims to the importance of the last 150-200ms of phonation as a factor in the differentiation process. Wales and Taylor (1987) felt that the peaks occurring at the end of the sentences are the most powerful predictors of question/statement judgement while Allan, Guy and Vonwiller (1984) felt that English speakers use the terminal F0 contour to convey different meanings. Within the Bantu languages Khumalo (1981) speaks of tone on the final syllable as being differentiating and Roux (n.d.) who proposes that resetting of the declining intonation contour in the latter part of the articulation (and raised intensity levels) may be argued as to the factors changing peoples perceptions from question to statement. It is evident therefore from the analysis of this corpus of Xhosa data that what has been found for languages in general experimentally and for the Bantu languages from an impressionistic observation together with limited experimental work (Xhosa) by those cited above, that perceptual changes may be realized from the pitch on the penultimate syllable or as stated by others the terminal pitch contour.

4. The perception test on the duration of the penultimate syllable also proved successful in that the manipulated signals definitely changed subjects' perceptions in both directions but especially from queclarative to statement. However in a number of instances there was no significant change in perception across the range of manipulations presented. This may be attributed to the following factors: firstly the method of increasing the duration in the ASL program leads to audible unnaturalness in the manipulated signal when the factor of increase is large. Secondly, if the duration is changed in the ASL program the duration *and* the pitch

contour are changed. However the average pitch will not change that much and in effect *only* the duration is changed perceptually. In the tests where the pitch was manipulated both the pitch and the duration changed due to methods employed. If pitch is raised the duration decreases implicitly. Therefore if the pitch of a statement form is raised, the duration decreases, thus *two* parameters are adjusted. Similarly the duration of a queclarative form would increase if the pitch were lowered. To modify only pitch, one needs to readjust the duration after the pitch manipulation. Wales and Taylor (1987) in their investigation on Australian English stated that duration as a variable did not significantly predict question/statement judgements. Hirschberg and Ward (1992) in their research on English found that amplitude and duration appeared to play no significant role in subjects interpretation of contours. However, Streeter (1978), Berkovits (1981), Scott (1982), Price, Ostendorf, Shattuck-Hufnagel and Fong (1991) verified the importance of segmental duration as one of the parameters in the process of disambiguation. Muira and Hara (1995) also found duration on the segmental final sentence to be a factor for questions to be perceived as rhetorical. Within the Bantu languages the limited experimental work in Xhosa by Theron (1991) showed definite acoustic durational differences between questions and statements. However, such results have not been perceptually investigated and therefore cannot be evaluated with the results from this corpus of data. It appears that while this durational parameter is relevant in the disambiguation process in this corpus of Xhosa data the fact that in some instances there were no significant change in perception allowing one to postulate that this cue to disambiguation is used more as an endorsement to decisions made earlier in the utterance as opposed to one which carries great weight in the disambiguation process.

Considering all the results from the perceptual tests, it becomes evident that many of the impressionistic claims relating to the importance of pitch and duration made by scholars to date are found to be acoustically and perceptually relevant in this corpus of Xhosa data in the disambiguation process between statements and queclaratives. Experimental research for other languages and the limited experimental research for Xhosa, not corroborated by perceptual testing, has also revealed the importance of these parameters in the disambiguation process. From the results of this acoustic and perceptual analysis it is evident too that statement and question forms are identified very early in the utterance in spite of the fact that pitch and duration on the penultimate syllable are also perceptually relevant. It would seem possible therefore to postulate that this cue of raised pitch on the initial syllable in the utterance and the raised pitch (register) of queclaratives as opposed to statements could be the most important cue in the disambiguation process certainly for this corpus of Xhosa data. It also seems possible therefore to postulate that listeners identify one form from the other early on in the utterance but that both pitch and duration on the penultimate syllable can change the listeners' perception of the decision made at the beginning of the utterance should there have been any confusion, for whatever reason, at the beginning of this utterance. Therefore both duration and pitch on the penultimate syllable could be said to serve as factors which reinforce perceptions made earlier, eliminating earlier confusions. It could also be postulated that certain parameters are more perceptually relevant than others i.e. to say that parameters are weighted in terms of their perceptual relevance. Specifically pitch plays a much more prominent role in the disambiguation process as for this corpus both the gating test and the pitch manipulations on the first vowel returned very positive results in peoples' perceptions of recognizing statements and queclaratives. This reinforces the importance of register as a cue to interrogativity. Listeners were able to correctly identify questions and statements from very short speech segments from which the penultimate syllables were absent altogether. If however the penultimate syllables are present, the duration and pitch on these syllables can either reinforce or override perceptual judgements made earlier in the utterance.

In the next chapter, the results of this acoustic and perceptual analysis on this corpus of Xhosa data will be considered and the application and implications of these results within the fields of linguistics, psycholinguistics and speech technology will be demonstrated.

3.8 Chapter summary

This chapter provides:

1. the aims of the perceptual analysis of data
2. a description of the automated perceptual testing using a multimedia computer laboratory
3. the procedures followed in the administration of the perceptual tests in a language laboratory
4. description of the use of the chi square test in the analysis of all the test responses
5. an explanation of Experiment A with subtests A1,(duration on the penultimate syllable, A2 (pitch on the penultimate vowel) and A3 (a gating test) conducted with 64 Xhosa mother tongue listeners.
6. An explanation of Experiment B with subtests B1 (duration on the penultimate syllable), B2 (pitch on the penultimate vowel), B3 (pitch on the first vowel) and B4 (a gating test) conducted with 63 Xhosa mother tongue listeners.
7. Results observations and an evaluation of the results received from Experiments A and B on all perceptually tested parameters.

CHAPTER 4

IMPLICATIONS FOR PSYCHOLINGUISTICS, LINGUISTICS AND HUMAN LANGUAGE TECHNOLOGY

So Socrates, you have made a discovery that false judgement resides, not in our perceptions among themselves, nor yet in our thoughts, but in the fitting together of perception and thought.

Plato's Theaetetus (in Massaro, Dominic.W. 1975:183)

4.1 Introduction

In this chapter the results obtained from the acoustic and perceptual analyses on the previously described corpus of Xhosa data will be assessed in relation to some of the implications for the disciplines of psycholinguistics, linguistics and human language technology.

This research, which has identified relevant perceptual cues in differentiating between statements and queclaratives in a language such as Xhosa, may, by inference, have psycholinguistic implications as it postulates on the devices used by the human brain in differentiating between statement and question forms.

Phonological rules, outputs and predictions specifically within Bantu languages, have been predominantly impressionistic in nature and many are lacking in any substantiation from validated authentic research. The approach used in this study, which is experimental in nature, provides statistically validated results on analyzed data and these results, therefore, have implications for the discipline of linguistics as they would lend authenticity to the phonological notions presented in this field within Bantu languages. This approach therefore also adds to the debate on the phonetics phonology interface which has been a much debated topic of discussion for some years relating to languages other than the Bantu languages. The concept of interface implies of course that there are two separate entities (phonetics and phonology in this instance) which in fact coexist within their own specific boundaries. However according to Ohala 1990:154

> ...phonology and phonetics are not mutually autonomous or independent whether as parts of the speech universe or as disciplines.

Human language technology, which involves primarily the conversion or transformation of language by a computer for the purposes of human-machine communication, is one that is vital for any community to remain part of the global information society. The results of research of this nature, which has identified essential components which differentiate between different forms, namely statements and questions, forms not only part of the speech recognition process but may also contribute to the synthesis of speech. Speech recognition is necessary for the development of language-enabled applications and, therefore, forms an essential and very prominent and intrinsic part of the process of human language technology. Implicitly, therefore, experimentally-based research of this nature has definite implications for human language technology.

4.2 Implications for psycholinguistics

The term 'neurolinguistics' encompasses the manner in which speakers and listeners process the production and perception of language to find out what real strategies and devices are used by the human neurological apparatus in achieving these operations. However, according to Laver (1994:52):

> The more familiar term for the study of such processes is **psycholinguistics** (Garman, 1990), reflecting what is sometimes a more abstract cognitive approach to the problem.

Psycholinguistic research, according to Laver (1994:52), within cognitive aspects of language, has tended to be rather specialized, with individual researchers investigating specific areas such as syntax, morphology, phonology, phonetics etc. However, there is now some evidence that cognitive research into speech production and perception is beginning to address the relations between the different aspects and

> ...fashioning the beginnings of an understanding of the complete cognitive chain from ideation to articulation (Garrett, 1988; Laver, 1989).
> (Laver, 1994:52)

The implications for psycholinguistics become apparent in a study of this nature, as this acoustic and perceptual investigation attempts to establish what devices Xhosa mother-tongue speakers and listeners use, in the production and perception of questions. Succinctly stated, what are the strategies used in question formation, i.e. other than the use of any lexical or syntactic markers, and which (if any) of these production strategies enable listeners to decode or perceive the message of questioning? According to Strange (1995:3):

> Speakers render their linguistic intentions into sequences of speech movements (articulatory gestures). These, in turn, generate complex acoustic signals that are picked up via the auditory system of the listener and "interpreted" so that the linguistic intention is recovered.

The field of psycholinguistics also embraces the notion of language acquisition. Language acquisition consists of acquiring the rules that govern and control the transformation of the structures of intention into linguistic surface structures. The communicative development of a child begins with the limbic stage (0-8 months), which includes reflexive emotional messages and then progresses to the gestural stage (8-12 months), the naming stage (12-15 months), focus (15-18 months), assertion (18-21 months) then the lexical stage and finally language (24 months) (Borden and Harris, 1984:267).

The process of language acquisition involves the stages of acquisition, the actual meaningful elements which are acquired, and the sequence of acquisition during this process. It would seem logical that language acquisition should be discussed with reference not only to these stages but also to the initial elements acquired and how they may relate to the results obtained from these experiments. Lecanuet (1998:340), in his contribution on foetal responses to auditory and speech stimuli, states that

> Several research hypotheses remain to be tested; but a large number of speech components - mostly, but not only, the prosodic ones – are transmitted to the amniotic milieu. Toward the end of gestation, the fetus is able to perform speech-relevant acoustic discriminations.

Jusczyk et al (1998:376), in their article on speech perception during the first year, also state that even 2-month-olds benefit from the organization that is present in sentential prosody and that, not only are they sensitive to prosodic markers, but that these markers may well play a role in what infants remember about utterances.

It is apparent that in the acquisition of language:

> During the babbling period before speech has developed, the infant reveals his sensitivity to intonation patterns of other speakers by mimicking them. (Borden and Harris, 1984:212)

In discussing language acquisition and the accompanying intonational signals, Lieberman (1967:41) says:

> Now the first "meaningful" element of speech behavior that can be observed in children actually occurs much earlier than from one to two years of age. In the very first months of life, during the babbling stage and indeed during very first minutes of life children employ "meaningful" intonation signals. The cries are at first meaningful, only in that they have a physiological reference. We believe that these signals, which appear to be innately determined, provide the basis for the linguistic function of intonation in adult speech.

As far back as 1936 Lewis noted stages in the development of language, where a child, at an early stage, discriminates in a broad sense between the different expressive patterns of intonation. Then follows a second stage, which includes both the phonetic and intonational form, where the intonational form dominates and finally when the phonetic form dominates in solving and soliciting specific responses. However, during this developmental process, while the intonation pattern may be considerably subordinated, it certainly does not vanish. Further, while children develop their language systems, they also increase their knowledge of semantics and at the same time discover the rules by which their language is governed.

> The rules are of three sorts: *syntactic* rules, which account for the structure of the sentences, including transformations of simple declarative sentences into questions or passives; morphological rules, which account for changes in meaning brought about by changing sounds (cat, cats) or by intonation (Yes,Yes?); and *phonological* rules, thought to account for the sounds of the speech stream (Borden and Harris, 1984:8).

According to Hörmann (1979:249):

> In communication with the child, the adult characteristically uses two speech forms which in this sense particularly facilitates understanding, namely question and command. While the normal conversation within the family typically contains 1 to 25 percent interrogative utterances, one finds up to 50 percent of such utterances addressed to the baby.

This developmental process, therefore, appears to begin *in utero* when the foetus is exposed to sounds that are transmitted via the amniotic fluid. It also seems apparent from research by DeCasper and Spence (1986) in Lecanuet (1998:344) that prenatal learning of some acoustic, probably prosodic features is suggested after testing 2- to 3-day-old newborns on speech sequence. Moon et al. (1993) also demonstrated in research that 2-day-old newborns preferred their mother's language to an unfamiliar language. Woodward (1992), who tested newborns who had been exposed from 34 weeks of gestation age to different types of music i.e. either classical or jazz, according to the mother's preference, using a conditioned sucking procedure, found a preference for the familiar music. Demonstration of a preference for the mother's language at such an early age and the preference for the familiar music seems to favor the interpretation of Mehler et al. (1986 and 1988) in terms of prenatal familiarization.

This process of development, as stated by Borden and Harris (1984), includes the changes in meaning, which include the transformation of declarative forms into question forms, using intonation as the strategy. From this statement it becomes apparent that this process for transformation using intonation is the device used by the child's neurological system in differentiating between the representation of a question as opposed to a statement and that it is learnt at a very early age.

It becomes evident from the literature on the subject that intonation may be used as an expressive and emotive device and takes on a linguistic reference at a very early stage in the developmental cycle of a child's life. It is also apparent that the frequent use of interrogative utterances (Hörmann 1979:249) addressed to the baby may allude to the preverbal existence of the concept and frequency of exposure to this form, perhaps, also leads to the familiarization (Mehler 1986, 1988) and the interpretation made by the infant by association also of the prosody with the question forms. In other words, it is not so much the meaning of the word that is significant in the decoding process, but verbally labelling the already available meaning or concept (Hörmann, 1979:249).

As stated above, the acquisition of language implies also the acquisition of intonation, which is inextricably linked to language in the transmission of different messages. Therefore, in view of the literature, which states that intonation is used by children at a very early stage of development as an expressive device with linguistic reference and the preponderate use of interrogative utterances, once again incorporating intonation with linguistic reference, directed at children during their early development, seems to indicate that intonation plays a very significant role in linguistic competence. In fact it, could be postulated that before the meanings of the words are learnt by the child other labels or devices (for example, intonation) are used and are given to an already available meaning, concept or idea.

According to Nooteboom et al. (1978) many investigations in speech research assumed that an understanding of how acoustic and phonetic features were extracted from the acoustic wave and combined into phonemes and syllables, allowed one to describe all language processing by using these underlying recognized phonemes and syllables as their input. This was the basic assumption underlying proposals such as the analysis by synthesis model (Halle and Stevens, 1964; Stevens and Halle, 1967) and the motor theory of speech perception (Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967). By contrast other proposals in the line of information-processing, while still based on these assumptions, now acknowledge the potential contribution of speech prosody to perceptual segmentation (e.g. Pisoni and Sawush, 1975; Massaro, 1975; Massaro and Cohen, 1975 and Nooteboom et al. 1978).

The importance of prosody in language acquisition in the perception of language and, more recently, in the line of information-processing has now been established. Taking into consideration the acquisition of language and intonation, the results of this study should be evaluated in ascertaining what possible devices were employed by listeners from a psycholinguistic perspective in the decoding or the perception of queclaratives, as opposed to statements.

The perceptual level of analysis is concerned specifically with the registration of the perceiver of sensory data of relevant types and according to Laver (1994:27):

> There are four perceptual domains available to the human auditory system which are all exploited in listening to speech. These are the domains of perceptual quality, duration, pitch and loudness.

In this study three of these parameters were investigated, namely duration, pitch and loudness and it was found that both duration and pitch were perceptually significant. While amplitude has also been nominated as a factor in the disambiguation process in other languages, for example Australian English (Wales and Taylor, 1987), English, (Hirschberg and Ward, 1992) and Korean (Jun and Oh, 1996), to name but a few, it was found that in this corpus of Xhosa data, no consistent patterns of significance emerged in the acoustic analysis and it was therefore decided to discard this parameter for further perceptual testing.

From this corpus of Xhosa data, the following perceptual factors play a significant role in the differentiation between statement and question forms:

1. Early recognition (derived from gating experiment) with minimal cues;
2. Variation of pitch on the first vowel;
3. Variation of pitch on the penultimate vowel;
4. Variation of duration on the penultimate syllable.

The early identification of these two different forms within this set of data seem, therefore, to coincide with impressionistic views that raised register is typical for question formation, but more importantly, that it is perceptually relevant in the identification of this question form.

According to Studdert-Kennedy and Hadding (1973:295)

> Previous studies, naturalistic and experimental, have suggested that listeners make use of an entire contour, not simply of the terminal glide, in judging an utterance (see Gårding and Abramson, 1965; Hadding-Koch, 1961; Hadding-Koch and Studdert-Kennedy, 1963, 1964, 1965).
>
> Hermann (1942) has pointed out the generality across languages, including Swedish, of a high pitch for questions (see also Hadding-Koch, 1961, especially pp. 119 ff.). Bolinger (1964), among others, has discussed the apparently "universal tendency" to use raised tone to indicate points of "interest" within utterances and also to indicate that more is to follow, as in questions (cf Hadding-Koch, 1965). The data of this experiment are consistent with these "universal tendencies".
> Studdert-Kennedy and Hadding (1973:309)

The results on this corpus of Xhosa data, which show the early recognition of the different forms from the first phoneme of the utterance and the evidence of a raised register for question as opposed to statement forms allow one to postulate that the early acquisition of intonation and the early use of intonation as a linguistic reference (relating to interrogative forms) could form the cornerstone in the perception of question identification. Certainly from this corpus of data it is evident that the perception of these two forms is made by using raised pitch from the first phoneme and an overall raised register. The variation of pitch on the penultimate syllable was also, in this corpus, perceptually relevant in changing listeners' perceptions from statement to question form and vice versa. The use of raised pitch from the first phoneme, then, could be attributed to the early acquisition of the intonational structures (subsequently stored in long-term memory), which provide imperative semantic value in interpretation and which contribute to the decoding of linguistic utterances in the communication process. Therefore the early exposure to and acquisition of the intonation structures and the preponderate use of and exposure to interrogative utterances addressed to babies, together with the familiarization factor as described by Mehler (1986,1988) are all factors that may contribute to the interpretation or decoding of question and statement forms.

The results from this corpus of data also coincide with the results of research in Mandarin Chinese done by Shen (1990), who found that assertive intonation was distinct from interrogative intonation, not at the ending point, but at the starting point, and that interrogative intonation also begins at a higher register. Her findings were also found to be consistent with those of De Francis (1963), Ho (1977) and J. Shen (1985).

With regard to the consistent evidence of a raised register for questions as opposed to statements within this corpus of data, the use of such a strategy seems also to be consistent with research by Hermann (1942), Bolinger (1964) Studdert-Kennedy and Hadding (1973), Eady and Cooper (1986), Hirschberg and Ward (1992), and Jun and Oh (1996). The use of an overall high F0 and a terminal rise have also been associated with other utterances, in which the speaker wants to draw the attention of the listener and according to Studdert-Kennedy and Hadding (1973:295)

> ...in listening tests the labels “question,” “surprise,” “interest” have been found to be interchangeable...

Wales and Taylor (1987:209) also suggested that intonational or prosodic features may carry no intrinsic meaning of their own, but simply serve as attention markers that attract listeners’ attention.

Another factor to be considered when acknowledging the early recognition of these two forms is the sentiments expressed by De Mori and Probst (1985:60), who say that, in the case of speech, listeners continuously develop and anticipate what will come next, based on what they have already heard. Furthermore, this anticipation will control what one will pick up next and what one will, on these bases, modify what one has already picked up. This account of perceptual activity has been borrowed from Neisser (1976). Therefore, the occurrence of the perceptually relevant parameters of duration and pitch on the penultimate syllable of the utterance could be interpreted as the parameters used as the reinforcing factors in the early identification process which occurred from the initial phoneme. This would be the case particularly if, for some reason, early identification did not take place or if there was any indecision on the initial cue used in the differentiation process or if the initial cue was not sufficiently marked (Studdert-Kennedy and Hadding, 1973:312).

As the prosodic feature of pitch has also been linked to emotive statements, it is important to note that the results from the perceptual testing of this corpus of Xhosa data include queclaratives comprised of one lexical item, and that the test itself in the differentiation process offered only two possible options on selection (forced choice test) i.e. "Is this a question or a statement?" The subjects, therefore, were not given the option of relating the stimuli to an emotive utterance: however, as stated above, the listening labels for question, surprise and interest have been found to be interchangeable (Studdert-Kennedy and Hadding, 1973:295).

Cutler and Otake (1999) conducted three experiments on the use of pitch-accent information in spoken-word recognition by Tokyo Japanese listeners and found that acoustic cues to pitch-accent were available within a single syllable.

Otake and Cutler (1999:231) state that:

> Research in human spoken-word recognition has shown that listeners exploit relevant phonetic information as soon as it becomes available; any inter-word distinctions in the initial portions of words could therefore be of great use to listeners, enabling drastic reductions in the set of possible candidate words for recognition.

In the second experiment Cutler and Otake (1999) presented listeners with initial fragments of words (within a gating experiment) and asked them to guess the words. The listeners' guesses overwhelmingly had the same initial accent structure as the presented word even when all that the listener had heard of the word was the first consonant and a portion of the vowel that followed. Thus, listeners could exploit the cues to accent pattern available to them even in the first syllable of a word. Sekiguchi and Nakajima (1999) conducted a priming study, in which they presented listeners with word fragments that were segmentally ambiguous but suprasegmentally unambiguous. They found that such fragments facilitated lexical decision responses to the words, which they matched suprasegmentally but not to those that were suprasegmentally mismatched.

According to Otake and Cutler (1999:232)

> Together these results provide strong evidence that Tokyo Japanese listeners make early and effective use of pitch-accent information in recognizing spoken words. The suprasegmental structure of Tokyo Japanese words is, like the segmental structure, of use in narrowing down the set of potential candidates for lexical recognition.

In the results from this corpus of Xhosa data the early recognition of statements and question forms on the initial phoneme is only differentiated by the pitch. In research on word recognition, (Otake and Cutler, 1999), it seems that listeners also used pitch/accent in word recognition. By implication, this type of suprasegmental information, which has been verified as used in the process of spoken-word recognition, may also then be postulated as being used in differentiating between two linguistic forms, namely statement and question in this instance. As pitch is the only information available to the listener in the form of a minimal cue on the initial gate in this corpus of data to the exclusion of any other cues it would seem logical to postulate that the use of pitch could well be the most significant disambiguating factor.

As this type of interrogative, namely the queclarative, may be classified as being a type of incredulity question it would have an emotive content. In this research it was found that in all examples the queclaratives showed a raised register as opposed to the statements, which coincides with the findings of expanded pitch range (raised register) for incredulity (Hirschberg and Ward, 1992; Herman, 1996; Jun and Oh, 1996). It was also found from this research and those cited above that the increased pitch range was perceptually relevant in differentiating between statements and questions.

The variation of duration on the penultimate syllable also proved to be perceptually relevant in the differentiation between statements and queclaratives for this corpus of Xhosa data. However, while Ward and Hirschberg (1988) and Hirschberg and Ward (1992), found that for English incredulity sentences are significantly shorter than their non-surprised counterparts they also found that duration appeared to play no significant role in subjects' interpretation of the contour. Jun and Oh (1996) found that duration was significant for all three types of questions investigated, but that yes/no questions were significantly shorter than /wh/- or incredulity questions, while Miura and Hara (1995) found that manipulation of duration on the sentence final segment was a factor for questions to be perceived as rhetorical. However, while this has been proven to be a perceptually significant disambiguating factor in some languages, it has also been refuted as being significant in interpretation of contours in others. Wales and Taylor (1987:208) found that, although the sentence-final syllables of questions tended to be longer than those of statements, this did not significantly predict question/statement status. Hirschberg and Ward (1992:250), however, found duration did not play any significant role in listeners' interpretation of the contour. It would seem important to note here that the manipulation of pitch (which was done and presented as stimuli to listeners) results in degeneration of the signal that is presented for perceptual testing and, according to Miura and Hara (1995), lengthening also causes unnaturalness to listeners. This, therefore, should be considered as a factor affecting the results of perception tests on these parameters, as the quality of stimuli degenerates as the manipulations are increased.

From a psycholinguistic perspective, the results from this corpus of Xhosa data showed that duration on the penultimate vowel, pitch on the penultimate vowel and overall raised pitch (register) for questions are all perceptually relevant and are devices used to differentiate between statement and question forms. However, in view of the early recognition of the different forms (i.e. from the initial phoneme or first gate) in the gating experiment, one may postulate that pitch plays the most significant role in these perceptual judgements. The other parameters of duration and pitch on the penultimate syllable were also perceptually significant, yet they appear at the end of the utterance. From this the postulation may be made that the perceptual cues could be weighted in their order of significance. Therefore these factors, pitch and duration on the penultimate syllable, are used here to reinforce decisions already made at the beginning of the utterance. Studdert-Kennedy and Hadding (1973), in their research on the perception of intonation contours, recall a central finding in their previous study, namely that of perceptual reciprocity among various sections of a contour, where listeners would trade a high F0 at one point in the utterance for a high F0 at another. For example, an utterance with a high F0 at peak or turning point required a smaller terminal rise to be heard as a question than an utterance with a relatively low F0 at turning point or peak. In view of this, an utterance is marked as a statement or a question by its apparent terminal glide and earlier sections of the signal are only important insofar as they alter perceptions of that glide and give rise to reciprocity effects. They state, further, that

> However, earlier sections of the contour may also enter into the decision and, if sufficiently marked, override an incompatible, but relatively weak, terminal glide. (Studdert-Kennedy, 1973:312)

It seems possible, therefore, that the raised pitch on the first phoneme was sufficiently marked in making judgments on statement and question forms from this corpus of data and that the perceptually relevant parameters of pitch and duration on the penultimate syllables served to reinforce these judgements made at the very beginning, on the first phoneme. One could also postulate that should, the signal on the first phoneme be less marked, due to some idiosyncrasy of the speaker or degeneration of the signal during the process of manipulation, then the listener would rely on the other two parameters of pitch and duration on the penultimate syllable to make perceptual judgements on the different forms. This would coincide with the postulations made by Studdert-Kennedy concerning the process of reciprocity in the perceptual interpretations.

The measurement of mean response time in the perceptual experiments yielded results that shed some light on the perceptual processes involved in the recognition of question forms. When subjects were presented with stimuli that were easily identifiable, the response times were shorter. As the difficulty of the recognition task increased, due to the increased degree of manipulation, so did the response times increase, indicating the growing uncertainty of subjects (Miura and Hara, 1995). Yet, for the majority of subjects, the response times were shorter than the maximum time allowed, because the number of undecided responses was low. On the other hand, response times were within reasonable bounds. If response times were too short, it would suggest that subjects responded (clicked) as fast as possible without really listening to stimuli.

The identification and the postulation that pitch plays a major role in the differentiation process is not only important linguistically and psycholinguistically, but also has major implications in research pertaining to automatic speech recognition and text to speech synthesis essential components of human language technology. The possible contribution to this area of research will be discussed in the next section, which deals specifically with these implications for human language technologies.

4.3 Implications for human language technology

Language is the natural human means of communication and is the most effective means of expression. It is used in many different ways and is therefore in fact fundamental to all facets of life. In general, language is used in direct communications between human beings and not in interactions with the systems services and appliances which are used on a daily basis. Therefore, one may say that in fact the use of language is restricted. However, a change is taking place which has the ability to enhance and revolutionize the value of language in every aspect of communication. This change is the result of developments in the field of Human Language Technologies (HLT). While this concept of HLT is well established as a field of research in Europe, it is, in South Africa, almost unknown. Not only is there a definite hiatus in knowledge and subsequently a miniscule research output, but there is also an alarming lack of awareness about the value of this field of research to society. The numerous international projects which are in progress in HLT bear witness to the recognition of this field of study. Projects listed on the international website[1] include ACCESS, automated call centre through speech understanding system; APPOLO, an open workbench for multilingual document creation and maintenance; CAVE, caller verification in banking and telecommunications; EAGLES, expert advisory group on language engineering standards, to name but a few.

---

[1] www.hltcentral.org

> The objective of HLT is to support business activities in a global context and to promote human-centred infostructure ensuring equal access and usage opportunities for all. By developing and demonstrating multilingual technologies and exemplary applications, HLT will provide features and functions that are critical to the realization of a truly user Information Society. [Online: www.hltcentral.org]

Speech technology has been actively studied since the 1950's and it holds intellectual challenges that diversify into a large spectrum of scientific disciplines, such as computer science, electrical engineering, biology, psychology, linguistics, statistics, philosophy, physics and mathematics.

This technology has as one of its aims the improvement of communication between humans and machines. It is a field in which our knowledge of language can be used to develop systems that recognize speech and writing, understand text well enough to select the precise information required, and translate between two different languages; then, on the other end of the spectrum, to generate speech as well as the printed word. The processes within this technology are numerous, but all are broadly concerned with inputting material into a computer, either by speaking/typing or writing, in order for the computer to recognize, understand, transform (or complete the required task) and then to generate and present the results to humans (in some form). Summarized, therefore, the technology consists of an input level, a process of transformation (computer-based) and finally an output level for human consumption (see Figure 4.1).

Figure 4.1 indicates clearly the phases involved in the processing of language. The input level shows that there are three types of input material: text, speech or image. The next progression in this process is the recognition of the input material, distinguishing separate words, and subsequently the validation of the material. The system then interprets or analyzes the information in order to generate a response. Example responses may be a translation of the input or an intelligent reply to a natural language query of a database. Finally the output level is reached, in which the results of the preceding processing are presented to humans either in the form of text, speech or image.

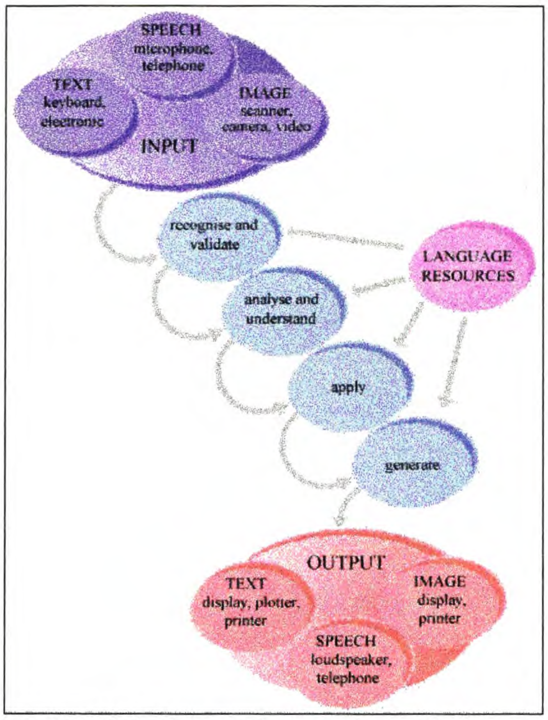


Figure 4.1 Processing performed on text, image and speech data from the input to the output phase. [Online[2]]

Research in this field facilitates the electronic handling of both spoken and written language and therefore by implication, because of the use of spoken language, empowers even the illiterate and semi-literate members of a community. Its many uses include the development of spell checkers, the retrieval and summary of text, automatic translation systems of both speech and text to name but a few. However, in order to use language in these various ways, enabled applications need to be developed. The field of language processing, whether in the form of written or of spoken language, may be better explained by referring to the schematic representation of the model of language engineering activities of the European Union as presented in Figure 4.2 below.

[2] www.linglink.lu/le/en/broch/harness.html

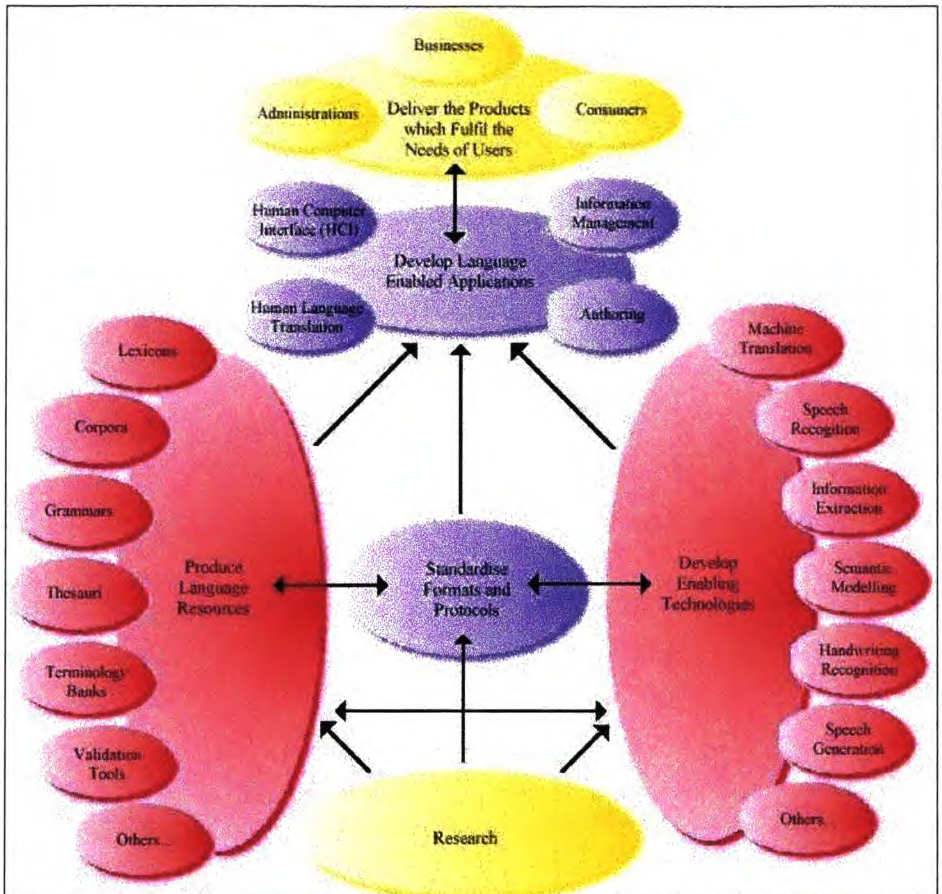


Figure 4.2 Schematic representation of language engineering [Online[3]]

In referring to Figure 4.2 it becomes clear that there are two major component activities involved in this process, both of which require a substantial volume of research. These components are:

1. the production of language resources
2. the developing of enabling technologies.

The products developed as a result of this extensive research include consumer products (e.g. automated telephone services, educational services and facilities for the physically impaired, see Roux 1999:handout), products for business and industry (automated multilingual translation services and automated multilingual systems with information on products from consumer items to insurance policies) and administrative support and security services (which include automated multilingual translation services, automated multilingual access to databases, automatic summarization of lengthy multilingual documents, and automatic identification of speakers for forensic purposes).

---

[3] www.linglink.lu/le/en/broch/harness.html

Within the South African context there are, according to Roux (2000:975), to date no real developments in this field of language and speech technology (LST). However, one such consumer durable item in the process of development is an automated multilingual voice based information retrieval system, designated for application in the hotel industry. This project, the working title of which is African Speech Technology (AST), is the first of its kind in South Africa and the associated research and development is conducted under the auspices of RUEPUS at the University of Stellenbosch. A typical architecture of this information retrieval system may be seen in Figure 4.3. In the development of this system it is apparent from the diagram that there are two extremely important components. Firstly, at the input level (after the request has been made) the process of automatic speech recognition (ASR) takes place and secondly, after the linguistic and information processing at the output level, text-to speech (TTS) is generated before the answer is supplied.

While this input-output approach does not represent anything new, the nature of the languages which could be used creates a vast number of challenges, especially with regard to design and database construction (Roux, 2000:976). These challenges include the multilinguality of the South African population, which presents complexities relating to home language and language use and code switching and code mixing (Roux, 2000:977). When one considers that South African English has five versions, one understands that the range of variation for English alone is extensive, without even beginning to consider the variations arising from the process of code-mixing and code switching, processes so prevalent in our society and more specifically within the African languages. Consideration should also be given to the social dialects used in South African society, some of which have very specific vocabulary. These would include '*tsotsitaal*', '*isihlonipho sabafazi*' and '*isikwetha*'.

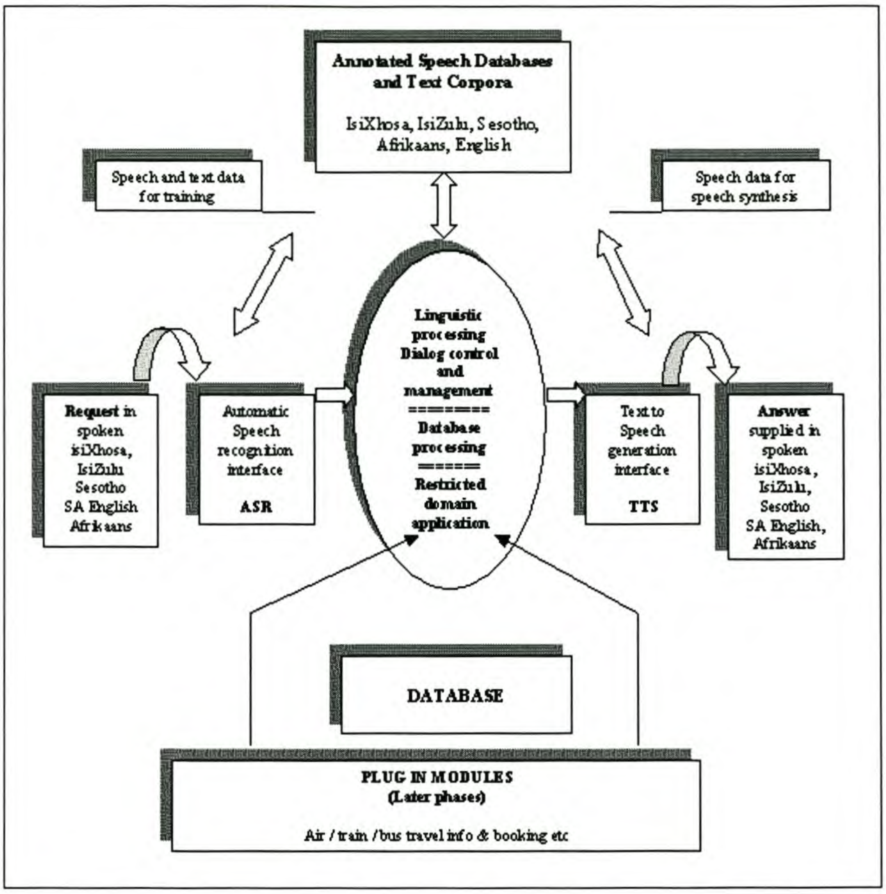


Figure 4.3 Macro-architecture of an automated multilingual voice based information retrieval and booking system as applicable to the African Speech Technology project. (Roux, Botha and du Preez, 2000).

Automatic speech recognition (ASR) is an essential component of language processing in the system shown in Figure 4.3. This includes the development of techniques and systems that enable computers to accept speech input in the form of an acoustic signal, and the subsequent identification of the units which make up the words. Speech recognition, therefore, may be described as the process of converting an acoustic signal, captured by a microphone or a telephone, into a set of words.

TTS technology is responsible for providing a natural sounding output to an interactive system. This task is relatively complicated as stated by Hertz (1997, online):

> But analyzing the input text is only half the battle. Once the system has determined the desired pronunciations of the input, it must generate the actual sounds that accurately produce these pronunciations. This task is complicated by the fact that perceptually-identical sounds, or phonemes, are acoustically different in different phonetic contexts. The precise duration and frequencies of

a sound depend on many factors – which segments precede and follow it, its position in the word, syllable or phrase, whether the syllable containing it is emphasized, whether the speech is fast or slow, whether the voice is that of a male or female, and so on.

This research, which has identified some significant components in the formulation of Xhosa queclaratives, provides useful information in the disambiguation phase of analysis which one can incorporate in the development of an ASR or TTS system functioning in a language such as Xhosa. This project has investigated queclaratives comprising one lexical item, and identified significant parameters in their formation. However the relevance of these findings would have to be investigated in the formation of questions using groups of words (forming queclaratives) and also questions derived by using lexical or syntactic markers.

From this corpus of data investigated and the parameters that were found to be both acoustically and perceptually significant, it would seem that, in order for the text module to be able to differentiate between a queclarative and a declarative, the following measurements would have to be considered:

1. the level of the pitch on the first syllable;
2. the overall level of pitch (register);
3. the level of pitch on the penultimate syllable;
4. the duration of penultimate syllable;

The following proposals on these parameters will now be presented, according to the number of syllables comprising each of the words investigated. All the values given in Tables 4.1 and 4.2 below are the percentage changes that should be applied to the duration and pitch values of a question form in order to make it sound like a statement. These changes may be used to synthesize statements by manipulating queclaratives, or vice versa. Appendix A contains the average numeric values for duration and pitch for the entire corpus. However, while these proposed values are related specifically to single word queclarative and statement counterparts, it is proposed that future research in this field should investigate and ascertain these values occurring in interrogatives comprised of phrases or sentences, and also those comprised using lexical and syntactic markers. Future experimental research on this topic may also investigate the naturalness of speech generated by these values.

Table 4.1 Average percentage phoneme duration changes from queclarative to statement forms.

Number of syllables	C_1	V_1	/m/	C_2	V_2	C_3	V_3	C_4	V_4
2 syllables	59.3%	47.2%		36.4%	4.4%				
3 syllables syllabic /m/	72.2%	27.3%	61.5%	47.7%	-6.5%				
3 syllables	21%	16.4%		34.2%	78.9%	29.4%	3%		
4 syllables syllabic /m/	16.9%	9.1%	24.4%	27.9%	77%	21.2%	-4.3%		
4 syllables	18.3%	11.8%		19.1%	9.3%	24.3%	81.8%	25.7%	-12.2%

As an example, consider the four syllable word /ngumfana/ that consists of the following tags: 'n1 u1 m f a1 n2 a2'. Given an existing queclarative form of the word, a statement form may be synthesized by assigning the following phoneme duration values as taken from Table 4.1 to the synthesized utterance:

(4.1)
$$
\begin{aligned}
d_{n1S} &= 16.9\% \text{ of } d_{n1Q} = \frac{100+16.9}{100} d_{n1Q} = 1.169 d_{n1Q} \\
d_{uS} &= 9.1\% \text{ of } d_{uQ} = \frac{100+9.1}{100} d_{uQ} = 1.091 d_{uQ} \\
d_{mS} &= 24.4\% \text{ of } d_{mQ} = \frac{100+24.4}{100} d_{mQ} = 1.244 d_{mQ} \\
d_{fS} &= 27.9\% \text{ of } d_{fQ} = \frac{100+27.9}{100} d_{fQ} = 1.279 d_{fQ} \\
d_{a1S} &= 77\% \text{ of } d_{a1Q} = \frac{100+77}{100} d_{a1Q} = 1.77 d_{a1Q} \\
d_{n2S} &= 21.2\% \text{ of } d_{n2Q} = \frac{100+21.2}{100} d_{n2Q} = 1.212 d_{n2Q} \\
d_{a2S} &= -4.3\% \text{ of } d_{a2Q} = \frac{100-4.3}{100} d_{a2Q} = 0.957 d_{a2Q}
\end{aligned}
$$

Table 4.2 Average percentage vowel pitch changes from queclarative to statement forms.

Number of syllables	V_1	/m/	V_2	V_3	V_4
2 syllables	-27.8%		-40.6%		
3 syllables, syllabic /m/	-21%	-26.5%	-39.5%		
3 syllables	-15.8%		-27.9%	-35.8%	
4 syllables, syllabic /m/	-15.7%	-20.6%	-36.9%	-17%	
4 syllables	-13.5%		-22.9%	-34.7%	-33.2%

Continuing the previous example, the statement form of the word /ngumfana/ may be synthesized by assigning the following vowel pitch values (Table 4.2) to the synthesized utterance:

(4.2)
$$p_{u\mathrm{S}} = -15.7\% \text{ of } p_{u\mathrm{Q}} = \frac{100-15.7}{100} p_{uQ} = 0.843 p_{uQ}$$
$$p_{m\mathrm{S}} = -20.6\% \text{ of } p_{m\mathrm{Q}} = \frac{100-20.6}{100} p_{mQ} = 0.794 p_{mQ}$$
$$p_{a1\mathrm{S}} = -36.9\% \text{ of } p_{a1\mathrm{Q}} = \frac{100-36.9}{100} p_{a1Q} = 0.631 p_{a1Q}$$
$$p_{a2\mathrm{S}} = -17\% \text{ of } p_{a2\mathrm{Q}} = \frac{100-17}{100} p_{a2Q} = 0.83 p_{a2Q}$$

The impact of language engineering has far-reaching implications in any society. However, for South Africans, the most critical implication is the necessity to become part of the global information society. The inability to recognize and acknowledge that this type of research is essential for participation within the global community will surely isolate this country and its people. This isolation, which will basically exclude South Africa from global markets and global communication, will also be manifested in commerce and industry and will show a ripple effect on every other aspect of society within the country. The domain of HLT entails harnessing the power of language and then using this power to its fullest potential in the development of the people and the country in its entirety.

It has been said that Europe has been an excellent region for the promotion of this type of research, owing to its heritage in cultural diversity. With this in mind, it would seem self-evident that South Africa would similarly benefit, since it enjoys a similar milieu. Speech and language processing allows for multilingual communication, while protecting and preserving a varied cultural heritage. It also promotes productivity and while often misconstrued as elitist research, serves to assist all people in communicative processes, even where literacy is at a low level.

Research of this nature requires funding and recognition and acknowledgement of its potential contribution to society as a whole. It therefore requires, in the South African context, an enormous national awareness campaign, not only about its existence but also, more importantly, of its practical applications to a country so rich in cultural diversity and multilingualism. It also requires cooperation in national research programmes, and transnational cooperation between research and industry.

4.4 Implications for linguistics

Over the period of the last few years there has been extensive debate on the role of experimental phonetic data and phonological analyses. This debate, much of which is to be found in contemporary literature, by inference also includes the existence of an interface between phonetics and phonology. The concept of an interface, while used predominantly within the discipline of computer science to refer to the relationship between a computer and ancillary equipment, implies that the two entities are separate and maintain their individual boundaries; yet, in so doing, do interface or communicate with one another. Many arguments and debates either supporting or refuting the existence or absence of such an interface have been presented (Fromkin, 1975; Keating, 1988; Ladefoged, 1988; Ohala, 1990). Some of these arguments propound the view that two entities coexist, retaining their individual boundaries, while others hold that the entities are highly integrated. Ohala (1990:153) emphasizes the point that a clear-cut distinction between phonetics and phonology is indeed problematic in quoting Halle (1954):

> ...It is my purpose to show that a sharp distinction between phonetics and phonemics cannot usefully be maintained; that phonetics actually make use of considerations which are strictly phonemic; and that a description of language at any level, from phonetics to stylistics, cannot be properly evaluated without considering its consequences on all other levels.

Clark and Yallop (1990), on the subject of phonetics and phonology, state that while it is true that certain aspects of speech are more amenable to quantitative measurement than others, it would be a mistake to conclude that such measurement, in itself, is sufficient to capture the truth about speech. However linguistic systems and mental organization are also open to the danger of investigators speculating about speakers' intuitions and insights. Therefore investigation of any aspect of speech must be empirical. While the techniques used for testing data empirically are different for phonetics and phonology they are and can be nonetheless empirical. In the final analysis Clark and Yallop (1990:4) state that

> It is not unreasonable, then, to say that phonology deals with the systems and structures of speech, while phonetics focuses more narrowly on articulation and acoustics. But the boundary need not be sharply drawn, nor should it be surreptitiously constructed on assumptions about the primacy of one kind of reality above others. In short, although we analyse speech by breaking it down into several aspects, we should not forget that the true reality is one of integration.

Roux (1978:105) quotes from Garnes (1973, 1974), who said that in abstract phonology phonetic facts are often taken for granted and that verification of these facts is largely ignored. Roux agreed with these sentiments, saying that this seemed to be quite true of many, if not most contemporary phonological descriptions. Roux (1978) investigated Tucker's (1929) and Kunene's (1961) impressionistic claims relating to labialization in Sesotho. After experimentally based research Roux (1978:115) found that very little objective evidence could be presented in support of Tucker while his data seemed to support the views of Kunene. This is further compounded by Roux's (1978) investigation into the phonological analysis by Ponelis (1974) which adheres to the basic views of Tucker (1929). In the conclusion to this, Roux (1978:128) states that

> It was shown that a phonological analysis based on unverified phonetic data tends to accommodate ad hoc, and unmotivated phonological rules. To a certain extent this finding does not come as any surprise, after all, how can a linguist expect to account for a set of data in a correct and credible manner, if the correctness of the data as such is not beyond any doubt?

Roux (1991:37), in his article on the integration of phonetics and phonology, investigated the pre-generative and generative descriptions of vowel raising and tonal metathesis in Zulu. He demonstrates how the perceptual confusion between vowel quality and pitch has ultimately culminated in the phonological rule for vowel raising. This means that the mid-low vowel with relatively high pitch was incorrectly interpreted in articulatory terms, assigning it a higher tongue position on the cardinal vowel chart.

Roux further demonstrated how phonological analysis may predict totally incorrect phonetic outputs, which are nevertheless accepted due to the practice of separating phonetics and phonology. Roux (1991:30) shows how Laughren, on the subject of tonal metathesis, generalized about an issue which resulted in pursuing phonological arguments at the expense of phonetic reality. Surprisingly, however, the analysis continues without the phonetic input to lend credibility to this study.

According to Roux (1991:49)

> Phonetics and phonology are merely flip sides of the same coin without any interface involved. Although it is the right of proponents of each domain to determine their own specific objectives, useful and credible explanations can only be expected from studies effectively utilizing both sides of the coin.

It becomes evident from this discussion that phonological analysis can only have credibility if it is based on sound, verified phonetic data and, therefore, the role of experimental phonology, which encompasses the use of an integrated approach, cannot be overstated.

With regard to this study, consideration should be given to the linguistic claims relating to the factors contributing to the differentiation between questions and statements. These claims, which have been predominantly impressionistic in nature, include claims relating to the parameters of duration and pitch as differentiating factors.

Claims made generally in linguistics include those by Lieberman (1967), who is of the opinion that it is possible to generalize about intonation to the extent that short declarative sentences usually end with a falling fundamental frequency contour. Other scholars who have made claims relating to the use of pitch as a disambiguating factor between statements and questions include the detailed instrumental analysis by Jones (1909), Chiba (1935), Fonagy (1958), Hadding-Koch (1961) and Abrahamson (1962). Studies show this for English, Spanish, French, Finnish, Hungarian, Italian, Thai, Japanese, Swedish and German. Impressionistic claims relating to the use of intonation as a feature differentiating statements from questions have been made concerning numerous languages by Fries (1964), Restan (1972) Comrie (1984), and Hinds (1984), to name but a few. Concerning Bantu languages, such impressionistic statements have also been made by Bokamba (1976), Khumalo (1981), Louwrens (1987), Mathibela (1989), Steyn (1991), Khoali (1994) Poulos (1990) and Nkabinde (1999) and some also include claims on the use of differing speech tempo and tonal register (cf. §3.6). Other features differentiating statements from questions, as impressionistically noted by scholars, related to duration of the penultimate vowel (Lanham, 1963; Riordan, 1969, cf. §2.5.1), higher absolute pitch level for questions (Riordan, 1969; Khumalo, 1981) and tonal down-drift for statements as opposed to questions (Louw, 1968; Khumalo, 1981; Laughren, 1984, cf. §2.6.1).

Primarily, the aims of this research of statements and queclaratives was to determine to what extent, duration, pitch and loudness of the syllables and the utterance as a whole facilitated this distinction.

Experimental analysis on the data set of 854 queclarative and statement pairs demonstrated that for Xhosa:

1. Duration of all the syllables and words are significantly longer for statements than queclaratives, with the penultimate syllable being the most significant. This implies that the tempo was higher for queclaratives than for statements. From a linguistic perspective, the research therefore implies that, in Xhosa single-item queclaratives, the rule is that one criterion of question formation is increased tempo, which results in shortened syllables and the utterance as a whole (§2.5.3).
2. In 60% of the cases it was found that the pitch of the whole word and that of every significant vowel was higher for queclaratives than statements, which implies an overall raised register for queclaratives (§2.6.3).
3. It also revealed that down-drift occurred on final syllables in both statements and questions (§2.6.3).

Comparing these results to the impressionistic claims made for Bantu and other languages, it would seem possible to postulate that, for Xhosa, the linguistic components of the formulation of questions resides in the increased tempo of production (which results in shorter duration) and an overall raised register with down-drift on the final syllable.

To date claims relating to prosodic data made by scholars concerning the African languages are not only elusive, but lack authenticity as a result of their impressionistic nature. The way in which scholars have dealt with this subject over many decades, without having questioned the authenticity and correctness of the claims, is surprising, if not disappointing. This becomes even more alarming in the light of this data playing a role in the development of the phonological theory of a language. There are many comprehensive sets of tonal data produced by competent scholars, such as Doke (1926), Doke and Vilikazi (1948), Cope (1956, 1959, 1966), Rycroft and Ngcobo (1979), and Khumalo (1981, 1987) for Zulu and Kropf (1915), Lanham (1958, 1960, 1963), Louw (1968, 1969, 1971, 1995), Davey (1973) and Pahl (1977, 1978, 1989) for Xhosa. However, when comparing this data and its acquisition to other related fields that deal with the scientific aspects of language in terms of neurolinguistics and linguistics (t'Hart et al., 1990; Studdert-Kennedy, 1973) and speech technology (Witten, 1982) it is amazing that no queries have been made about the tonal data in these languages to date. Claughton (1992), for example, used one informant in establishing his set of tonal data and states that

> ...my work, Davey's thesis (1973) and Riordan's work (1969) are all based mainly on the speech of one informant, Stanley Bentele. Fortunately, his tones do not appear to be idiosyncratic in any way. (1992:33)

In view of the use of one informant and the statement that *his tones do not appear to be idiosyncratic* (my emphasis) one therefore has to pose the question, just how authentic can this investigation be into Xhosa data? Claughton (1992:161), in his study on Xhosa tone, also appears to have a problem with the abstract nature of accent as it does not have a clear acoustic correlate. Claughton (1992:161), quoted correctly from Fry (1960), who stated that stress accent is regarded as abstract and "lacks a clear acoustic correlate". However such a statement does not necessarily imply that no such phenomenon exists and therefore this is good enough reason to discard the possibility of its existence at all. According to Roux (1998:41):

> It is unfortunate that phonologists such as Claughton rely on a single reference (cf. Fry, 1960) to discard of the possibility of finding relevant acoustic correlates for a specific phenomenon.

In his second objection, Claughton says that

> There is no evidence that speakers of Xhosa feel that there is accent on the syllable where Goldsmith, Peterson and Drogo postulate an accent. (Claughton, 1992:162)

Bearing this statement in mind, Roux (1998:42) queries who these speakers were that Claughton is referring to and questions their linguistic naiveté and their understanding of the concept of linguistic accent. The question is then posed as to whether this was a general observation or the result of carefully controlled psycholinguistic experimentation, for which the author presented no evidence in the study.

Claughton's data (1992:6), by his own admission, is not only impressionistic, but may also be considered to be the result of a method that has many dangers. It is also lacking in areas of controlled experimentation of data and, therefore, both his analysis and his rejection of a possible alternative description of tonal phenomena (namely the use of accent in Xhosa) surely requires a reappraisal of the approach used in tonal studies specifically within Bantu languages.

Khoali (1993:4), in his thesis on the tonal grammar of Sesotho, is very critical of the tonal descriptions in the Southern African context. However, he himself uses the criterion of consistency in determining reliability of data. He states that existing tonal studies in Sotho show

> ...serious gaps in the presentation of phonetic facts. Some crucial data revealing the subtle phonetic cues in different syntactic contexts are so flimsy as to be regarded as inadmissible as representation sample for making scientific valid generalizations. To make it worse, some of the data is inconsistent with that collected by Charles Kisserberth between 1987 and 1988 as well as our data.

The impressionistic data typically used in phonological analyses by scholars is evidenced by the examples cited above. With regard to the results of this experimentally based study it will now once again be demonstrated how claims made on impressionistic notions cannot be supported, specifically in this instance relating to the differences between questions and statements.

Laughren (1984), in her exposition on tone in Zulu nouns, also discusses the prosodic structure of statements and echo-questions in Zulu. She has also, due to the close relationship between the Nguni languages, often applied the rules of one language to a related language. According to Roux (n.d.), Laughren bases her analysis on the assumption that the phonetic difference between statements and corresponding echo-questions are to be found in, respectively, the presence and absence of down-drift.

According to Laughren (1984:194)

> This descending type of intonation, or down-drift, and final lowering can be schematized as //⌝↘↓// 'declarative'. An intonation contour which eliminates down-stepping and down-drift is used to characterize interrogative and some emotional types of utterances. This can be impressionistically characterized as //—↓// 'interrogative'. The raising of the pitch level is also used as an expressive device.

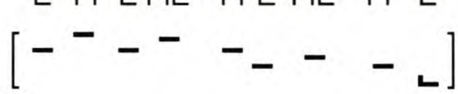
Roux (n.d.) interprets Laughren's statement to mean that she implies, in contrast to the falling curve characteristic of statements, that questions do not show any down-drift as the intonation level stays constant except for the final vowel.

Laughren (1984:224) represents the statement form (sibona inqola encinane / we see the small wagon) as such (i.e. in slow speech):

(4.3) Statement (slow speech)

// sìbónà í!·nqólà é!·ncá:nè //

L H L HL H L HL H L



The transformation from statement (slow speech) to statement (fast speech) resides in the assumption that vowel deletion generally takes place across word boundaries, which has resulted in the down-drift, and is conditioned by word-internal down-step, represented as follows where the final vowel /a/ of /sibona/ is deleted before the initial vowel /i/ of /inqola/ across the word boundary.

(4.4) Statement (normal and fast speech)

// sì + bón # í +!nqól # é + !ncá:nè //

L H~~L~~ HL H~~L~~ HL H L

This lowering of tones seems to be sufficient to account for the classical falling intonation contour for statements. This step in the analysis is totally reliant on the supposition that vowel deletion takes place generally across word boundaries. However, this has been found not to be true by Roux (n.d.), in his investigation on the prosodics of declarative and question sentences in Xhosa. While this may be true at times in quick speech or increased tempo in articulation, a rule based on this impressionistic assumption is not justified (Roux, n.d). Laughren also states that high tones in echo-question sentences are phonetically all equally high (Laughren, 1984:224) This supposition gives rise to the view that falling intonation contours are typical of statements, while question intonation contours typically display no down-drift as the contour stays level except for the final vowel. Although Laughren does not explicitly formulate such a rule, its application would result in the representation of a queclarative as follows (Roux, n.d.):

(4.5) Question

// sì + bón # í +!nqól # é + !ncá:nè //

L H H~~L~~ H H~~L~~ H L

This gives rise to the view that, in contrast to the falling intonation contour of statements, questions typically display no down-drift as the intonation level stays level, except for the final vowel. To achieve this contour it would be necessary to block the internal down-step and would result in the contour as shown in (4.3) above.

According to Roux (n.d.), the analysis presented above of the representation of the question form is based on assumptions concerning the phonetic structure of the language, for which there is very little experimental verification. It is not only based on the impressionistic views of one or two scholars, but also seems to minimize other impressionistic views regarding the differences between statements and questions in the Nguni languages.

Concerning the present corpus of Xhosa data, the possibility of vowel deletion between words (which ultimately is responsible for the down-drift in statements, as described by Laughren) was not tested as the data consisting of single-word queclaratives does not allow for such a test. However, such an impressionistic claim is refuted by Roux (n.d.) in his experimentation, where he found that vowels maintained their identity across word boundaries in both statements and questions and that such a categorical statement on vowel deletion is certainly not justified.

However, in the present corpus, it was also found that both statement and question forms showed down-drift as from the final syllable of the utterance and therefore the proposal that statements are typically characterized by down-drift, whilst in echo-questions there is level intonation except for the last vowel, cannot be supported. Theron (1991), in her experimentally-based investigation of Xhosa declaratives and echo questions, lends support to findings about statements being characterized by tonal down-drift, while identifying a measure of down-drift in echo-questions as well. She also refers to examples where equally sharp decline in the F0 curve may occur in questions as well. Laughren (1984:224) also found high tones in echo questions to be phonetically equally high. This, by implication, means that questions typically display no down-drift, as the intonation level stays level except for the final vowel. If this were the case, it would imply that it would be necessary to block the internal process of word-internal down-step as represented in (4.1) and (4.2) above. This analysis of Laughren's differentiation between statements and echo-questions:

> ...is based on assumptions concerning the phonetical structure of the language for which very little experimental verification may be presented. This analysis is not only based on one or two scholars, but it totally plays down other impressionistic views regarding the differences between statements and questions in the Nguni languages (Roux, n.d.).

Laughren's outputs from the impressionistic contribution on Zulu tonal rules are certainly questionable in the light of experimental claims made by Roux (n.d.), Theron (1991) and from the research results presented here.

Considering that phonological predictions are being made by questioning and subsequently rejecting impressionistic data, replacing the rejected claims with yet 'other' impressionistic data seems both unprofessional and frighteningly unscientific. However, it is this approach, lacking in authenticity and quantitative analysis, which dominates in such phonological outputs. By contrast the type of information gained from an experimental investigation, provides reliable data and therefore has implications for models of intonation involving an auto-segmental phonology approach.

According to House (1990:17)

> A number of recent perception studies involving pitch peak shifts or pitch contour manipulations have been carried out with the objective of testing models and descriptions of intonation (e.g. Bruce, 1977; Gårding and House, 1987; Kohler, 1987; Touati, 1987; Botinis, 1989; Gårding and Eriksson, 1989; Pierrehumbert and Steele, 1989).

Similarly for Xhosa, the same objectives of testing models and description could be applied using the results obtained from research projects which are experimentally based, which is the approach used in this study.

Relying on the perceptual judgements and impressionistic descriptions of persons in tonal descriptions in Xhosa reveals a total lack of credibility with regard to research in this field and leads to many discrepancies, as have been related above. Having scrutinized the contributions on this subject relating to the Southern African Bantu languages, and Xhosa in particular, there is no doubt that, in order to present credible analyses for these languages, research in this field must be based on experimental evidence.

According to Roux (1998:48):

> Developing intricate details of a phonological theory without any sound physical base is just as futile as searching for invariant acoustic and perceptual properties in the speech signals without any reference to linguistic functions.

The outcomes of the experimental approach to the analysis of data results in credible and authentic contributions to the study and development of language from many perspectives, with implications for other related disciplines, as indicated in this chapter.

4.5 Chapter summary

This chapter has established that:

1. the very early recognition of the different forms, notably from the first phoneme, where pitch was the only possible differentiating factor, indicate that pitch could be considered to be perceptually the most significant differentiating factor psycholinguistically.
2. other factors that were perceptually relevant, namely duration and pitch on the penultimate syllable, may be used to reinforce decisions made from the first phoneme of the utterance.
3. there are patterns that reveal themselves from this experimentation, both from a physical and perceptual perspective, which allow one to postulate that specific parameters should be included in the reproduction of different speech forms. This is a vital component in the development of human language technology, which is essential for any country to remain part of the global community and, therefore, research of this nature, which provides necessary, relevant and experimentally proven components of language, makes a vital contribution to speech recognition, an essential constituent of human language technology.
4. the phonological rules and outputs within Bantu languages are predominantly purely impressionistic in nature and are, therefore, unreliable, as seen in the critique of contributions from Laughren (1984) and Claughton (1992). Therefore it is imperative that in future an experimentally-based research methodology be used to produce reliable, valid and authentic results, which supports the argument that quantified data at physical as well as perceptual levels are indispensable for credible descriptions and analyses and, therefore, have definite linguistic implications.

CHAPTER 5

CONCLUSIONS

The thesis of linguistic determinism leads to the realization: language is not simply an instrument we control; it controls us. The same language which adds a dimension of freedom to our existence appears to us at the end of this book as a power in its own right and, indeed as overpowering. To a certain extent language seems to travel past what we proudly feel to be our "real" capacities.

Hörmann, H. (1979:299)

This chapter provides an overview of the aims of this research project and the results achieved which relate to the original objectives. It also puts into context the accomplishments and the contributions that the results may have for other disciplines namely psycholinguistics, linguistics and human language technologies and the value of this approach for future research within the Bantu languages of Southern Africa. Finally a resumé of future research and recent activities in this field will be included showing the state of the art within the South African context.

The **first objective** of this research project was to determine the typical **acoustic differences** which exist between statements and queclaratives in Xhosa. This objective was achieved from the acoustic analysis in that the results revealed that the **duration** of the word and that of every syllable was longer for statements than for queclaratives. Therefore the **tempo** in syllables per second was higher for queclaratives. The duration of the penultimate syllable was the most significant syllable duration feature for 92.3% of the words, although other syllables may play a secondary role in disambiguation 59.0%. The analysis on **pitch** showed that the pitch of the whole word and that of every significant vowel was higher for queclaratives than for statements which implies that the overall **register** for queclaratives is higher than for statements. The pitch on the final vowel was the most significant feature for 66.7% of the words while the pitch on the penultimate vowel was second most significant in 30.8% of cases. **Down-drift** was apparent in both statement (89.7%) and queclarative (66.7%) signals, while throughout a **raised register** remained consistent for queclaratives. In this corpus of data it was found that most queclaratives had a **rising pitch contour** and statements a **falling pitch contour** however, there was also evidence of opposite trends. Regarding the analysis of **loudness** it was evident that queclaratives were louder than statements but that neither the absolute nor the relative loudness of each phoneme within a word was found to be significant in any example. As a result of these inconsistencies within the statistical analysis no further claims could be made regarding this parameter and it was therefore discarded for perceptual testing (cf. Chapter 2).

The **second objective** was to determine which acoustic parameters are relevant for mother-tongue listeners to make perceptual distinctions between statement and queclarative forms.

The second objective was achieved in that the results of the perceptual testing of the acoustically significant parameters indicated that both **duration** and **pitch** on the penultimate syllables showed that people's perception could be changed in both directions, i.e. from queclarative to statement and vice versa. These results were achieved after both durational and pitch manipulations on the penultimate vowels of both queclarative and statement signals were presented as stimuli to listeners in carefully designed perception tests (cf. Chapter 3).

The results of the gating test however also showed that linguistic decisions are made as from the first two phonemes of the utterance. The initial gates retained the pitch, duration, loudness and vowel quality characteristics. As duration and loudness were not found to be statistically significant on the first syllable, while pitch on the first vowel was statistically of secondary importance, one may postulate that **pitch** could be the cue to differentiating between these two forms. Having established this from the gating test, a further perceptual test was included using pitch manipulations on the first vowel.

The perception test of **pitch** on the first vowel showed that people's perception could be changed from the first syllable of the word when differentiating between statements and queclaratives.

One of the most significant results revealed the **early recognition** of the different forms as from the first two phonemes, which allows one to postulate that the most important parameter used in the differentiation process in this set of data was **pitch**.

Duration and pitch on the penultimate syllable, it could be postulated, is therefore only used to reinforce decisions already made at the beginning of the utterance or to make new decisions, if they were not made at the beginning due to confusion arising from some idiosyncrasy of the speaker or from the signal itself.

The **third objective** was to establish whether there were **human universals** present in the disambiguation process which were also applicable to Xhosa in differentiating statements from queclaratives. It has also been established that the use of **duration** and **pitch** and a **raised register** for questions as opposed to statements used as disambiguating factors coincides with findings in several other languages of the world showing some universal tendencies in the strategies used by humans in the disambiguation process. The **early recognition** of the different forms as from the first two phonemes was one of the most significant observations made and has also been referred to in other studies on interrogativity in languages other than the Bantu languages.

This research has, therefore, achieved the objectives initially defined in the identification of the acoustic parameters used in the differentiation between Xhosa copulative queclarative and statement forms and has also identified the parameters used by Xhosa mother-tongue listeners as cues in the disambiguation process through careful perceptual testing of data.

As stated at the outset of this project, experimentally-based research in most languages of the world has been in progress for many decades. However, as this is not generally so for Bantu languages, the results obtained from this project should be seen as an attempt to promote this type of research, as it has demonstrated the interdisciplinary value and the contribution that this type of research can make to the fields of psycholinguistics, linguistics and human language technologies.

This research project, which has included both acoustic and perceptual analyses, will endorse the viability of experimentally-based research and increase the credibility of research output within the Bantu languages. Furthermore, it has also demonstrated the relevance of experimentally-based research in harnessing the power of language as it relates to the field of human language technology. This in turn leads to improvement of communication, the dissemination of information to all and to the development and the upliftment of society in general.

Perhaps one of the most important aspects of this project lies not only in the credibility of research output, but the creation of an awareness among linguistic researchers. It also emphasizes the uses and the urgent necessity for experimentally-based research in the development of language, specifically those spoken in Africa, which has far-reaching implications for the development of the country and it's inhabitants as a whole. As the awareness of the magnitude of the potential of experimental research as applied to development permeates society, it is envisaged that attitudes will be changed. Experimentally-based research is often viewed as élitist through ignorance and a lack of understanding of its value. Its uses are widespread and multifaceted, not only in third world countries but it is also of paramount importance to all developing societies. This attitude towards so-called 'élitist research' which exists stems from naiveté not only on the part of potential researchers, but also of influential persons, who could otherwise sponsor, endorse and encourage research projects of this nature. More importantly, if research in this field is not pursued and research projects are not initiated and sponsored, this society will be impoverished by non-participation in the global community. This surely will have incomprehensible and dire implications.

The future of and the benefits to be gained from human language technologies are extraordinary and can not be overemphasized. Acknowledgement of this fact and the implementation and initiation of projects within this field should be considered to be of highest priority for the future development and participation of all the inhabitants of this country.

Future research and current activities

Within the South African context the need for future research in phonetics and phonology to be experimentally based, allows for credibility of research results, inclusion in the global community and the upliftment and development of society from a multifaceted perspective.

The importance of this approach to research sees an urgent need for a national awareness campaign of experimentally-based research amongst scholars, relevant business communities and government officials for support and endorsement, including financial support for future research projects.

There are also appeals from interested researchers and scholars for consideration of and inclusion of this approach to research within the developmental programmes envisaged by the present government relating to the development of all official languages and, especially, previously disadvantaged languages, namely the Bantu languages of Southern Africa.

Recent developments in this field of research within the African and South African context has seen the following progression in the field to date:

1. During a conference in Eritrea in January 2000, according to the Asmara Declaration on African Languages and Literatures, "the effective and rapid development of science and technology must be used for the development of the African languages".
2. At the Second Language Indaba held in Durban in March 2000, the final draft of the Language Policy was tabled. According to this draft, one of the basic requirements for a language policy for South Africa has to be "Development of an efficient language industry by, among other things, using and developing appropriate technology".
3. ALASA (SIG). A special Interest Group on Language and Speech Technology Development (SIG) of the African Language Association (ALASA) was founded following a workshop on Language and Speech Technology sponsored by PANSALB in 1999. This group has international links of co-operation with the Universities of Helsinki, Stuttgart and Leipzig. In the short term the aim of this

body is capacity building through the organization of training sessions, and in the long term at capacity building through restructuring of curricula at tertiary level.

4. DACST Innovation Fund. The Innovation Fund of the Departement of Arts, Culture, Science and Technology (DACST) is currently sponsoring a major project of the Research Unit for Experimental Phonology (RUEPUS) at Stellenbosch University. This three year project, *Promoting the development of the official languages of South Africa through language and speech technology applications (African Speech Technology)*, commenced in January 2000.
5. Apart from the national government's involvement in projects of this nature there is a newly founded *South African Foundation for Language and Speech Technology Development (SAFLAST[1])* which was founded in 1999 as a non-profit organisation and aimed at facilitating interaction between national and international role players, and organizing workshops and seminars in Human Language Technologies. This organization strives to create an environment conducive to development in this field of research which ultimately focuses on support from the private sector who are possible end-users of the developed systems or products.
6. A module entitled 'Language processing and language technology' has recently been approved at undergraduate level at the University of South Africa for implementation in 2002 with the specific aim of short term capacity building.
7. An undergraduate degree is presently in the developmental stages at the University of Potchefstroom relating to speech technology, artificial intelligence and psycholinguistics. While no date has as yet been set for implementation this qualification also aims at short term capacity building.
8. RUEPUS. This NRF/CSD University of Stellenbosch sponsored research unit is in a process of transformation to become a full fledged text and speech resource centre at the end of its four year cycle in March 2002. This has been approved by its national Advisory Committee provided that the necessary financial support could be obtained.
9. There are at this moment in time a number of national and international companies involved in speech technology applications which have indicated an interest in entering the South African market, while some are already involved. However these companies are focusing on the development of English systems due to the fact that

---

[1] www.saflast.sun.ac.za

appropriate language specific resources in electronic format are not yet available in the indigenous languages.

10. There are a number of academics at the universities of Pretoria, Cape Town, Stellenbosch, the North, Transkei, Natal, Qwa-Qwa, Bloemfontein, Potchefstroom and UNISA with specific interest and in some cases with a research record in this field.

The progression, sensitization and the acknowledgement of the importance of the potential contribution of language processing and speech technology by the national government and business bodes well for South Africa. Future research in this field is particularly important in a country such as South Africa which enjoys a multilingual milieu and has at the same time underdeveloped literacy skills.

This research project which is based on experimental research has shown the importance of this approach to research within languages in general and especially within the Bantu languages of Southern Africa. It is believed that projects of this nature will encourage future researchers within the field of African languages to recognize the significance and the contributions that this approach may make to the development of the previously disadvantaged languages of Southern Africa. It is also believed that this approach creates an awareness of a new and stimulating approach to research within the African languages which in the final analysis will benefit the society as a whole and in so doing also make some contribution to the multifaceted concept of the *African Renaissance.*

BIBLIOGRAPHY

Abrahamson, A.S. 1962. "The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments". Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Allan, K. 1984. "The Component Functions of the High Rise Terminal Contour in Australian Declarative Sentences". *Australian Journal of Linguistics,* 4, 19-32.

Azuma, J. and Tsukuma, Y. 1990. "Prosodic Features Marking the Major Syntactic Boundaries of Japanese." *Proceedings of the 1990 International Conference on Spoken Language Processing,* 453-455.

Berkovits, R. 1981. "Are Spoken Surface Structure Ambiguities Perceptually Unambiguous?" *Journal of Psycholinguistic Research,* 10, 41-56.

Bokamba, Eyamba G. 1976. *Question Formation in Some Bantu Languages.* (Ph.D. thesis). Bloomington, Indiana: Indiana University Press.

Bolinger, Dwight L. 1958. "A Theory of Pitch Accent in English". *Word,* 14, 109-149.

Bolinger, Dwight L. 1964. "Intonation as a Universal". In H. Hunt (ed.), *Proceedings of the Ninth International Congress of Linguistics* (The Hague), 833-844.

Bolinger, Dwight L. 1978. "Intonation Across Languages". In Greenberg et al. (eds), *Universals in Human Language*. Stanford: Stanford University Press.

Borden, Gloria J., and Harris, Katherine S. 1984. *Speech Science Primer Physiology, Acoustics, and Perception of Speech* (Second Edition). U.S.A.: Waverly Press, Inc.

Botinis, A. 1989. *Stress and Prosodic Structure in Greek.* Lund: Lund University Press.

Brown, G., Currie, K.L. and Kenworthy, J. 1980. *Questions of Intonation.* London: Croom Helm.

Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective.* Lund: Gleerups.

Bruce, G. 1985. "Structure and Functions of Prosody". In B. Guerin and R. Carre (eds). *Proceedings of the French Swedish Seminar on Speech.* Grenoble, 549-559.

Chiba, T. 1935. *A Study of Accent, Research into its Nature and Scope in the Light of Experimental Phonetics.* Tokyo: Phonetic Society of Japan.

Chomsky, N. 1972. *Language and Mind* (enlarged edition). New York: Harcourt Brace Jovanovich, Inc.

Clark, John and Yallop, Colin. 1990. *An Introduction to Phonetics and Phonology.* Oxford and Cambridge, U.S.A.: Blackwell Publishers.

Clements, G.N. and Goldsmith J.A. (eds). 1984. *Autosegmental Studies in Bantu Tone* 3. Dordrecht Holland/Annaminson U.S.A.: Foris Publications.

Claughton, John S. 1992. *The Tonology of Xhosa.* (Unpublished Ph.D. dissertation). Rhodes University, Grahamstown.

Cole, D.T. 1955. *An Introduction to Tswana Grammar.* Cape Town: Longman.

Comrie, Bernard. 1984. "Russian". *Typological Studies in Language: Interrogativity* (vol 4). Amsterdam: John Benjamins Publishing Company.

Cooper, W.E. 1976. "Syntactic Control of Timing in Speech Production: A Study on Complement Clauses". *Journal of Phonetics,* 4, 151-171.

Cooper, W.E. and Sorenson, J. 1981. *Fundamental Frequency in Sentence Production.* New York: Springer.

Cope, A.T. 1956. *An Investigation into the Tonal System of Zulu, with Specific Reference to Nouns.* (Unpublished M.A. dissertation). University of Natal, Durban.

Cope, A.T. 1959. "Zulu Tonology". Africa und Übersee, 43, 1-42.

Cope, A. T. 1966. *Zulu Tonology, Tonology and Tonal Grammar*. (Unpublished Ph. D. thesis). University of Natal, Durban.

Cope, A.T. 1970. "Zulu Tonal Morphology". *Journal of African languages,* 9 (3), 111-152.

Cruttenden, A. 1981. "Falls and Rises: Meanings and Universals". *Journal of Linguistics*, 17, 77-91.

Crystal, David. 1985. *A Dictionary of Linguistics and Phonetics* (2nd ed.). Oxford: Basil Blackwell.

Cutler, Anne. 1998. "The Recognition of Spoken Words with Variable Representations". *Proceedings of the ESCA Workshop on Sound Patterns of Spontaneous Speech* (Aix-en-Provence), 83-92.

Cutler, Anne and Otake, Takashi. 1999. "Pitch Accent in Spoken-word Recognition in Japanese". *Journal of the Acoustical Society of America,* 105R, 1877-1888.

Davey, Anthony S. 1973. *The Moods and Tenses of the Verb in Xhosa*. (Unpublished M.A. dissertation). University of South Africa, Pretoria.

De Casper, A.J. and Spence, M.J. 1986. "Prenatal Maternal Speech Influences Newborn's Perception of Speech Sounds". *Infant Behavior and Development,* 1, 36-48.

De Francis, John F. 1963. *Beginning Chinese*. New Haven: Yale University Press.

De Mori, Renato and Probst, David. 1985. "Knowledge-based Computer Recognition of Continuous Speech". In M.R. Schroeder (ed.). *Speech and SpeakerRrecognition* 12 *Bibliotheca Phonetica* (Basel: Karger), 53-79.

Denes, P. B. and Pinson, E.N. 1973. *The Speech Chain.* New York: Doubleday.

Demuth, Katherine. 1995. "Questions, Relatives, and Minimal Projection". *Language Acquisition,* 4 (1&2), 49-71.

Doke, C.M. 1926. "The Phonetics of the Zulu Language". *Bantu Studies,* 2. Johannesburg: Witwatersrand University Press.

Doke, C.M. and Vilikazi, B.W. 1948. *Zulu-English Dictionary.* Johannesburg: Witwatersrand University Press.

Doke, C.M. and Mofokeng, S.M. 1957. *Textbook of Southern Sotho Grammar.* Johannesburg: Longman.

Du Toit, T. 1999. (December 17-last update) "A Short Introduction to Text-to-Speech Synthesis".[MULTITEL-TCTS Lab], [Online] Available: http://tcts.fpms.ac.be/synthesis/introtts.html. (2000).

Eady, S.J. and Cooper, W.E. 1986. "Speech Intonation and Focus Location in Matched Statements and Questions". *Journal of the Acoustical Society of America,* 80, 402-415.

Elman, Jeffrey L. and McClelland, James L. 1985. "An Architecture for Parallel Processing in Speech Recognition: the TRACE Model". In M.R. Schroeder (ed.). *Speech and Speaker Recognition* 12 *Bibliotheca Phonetica* (Basel: Karger), 6-35.

Fonagy, I. 1958. "Electrophysiologische Beiträge zur Akzentfrage". *Phonetica,* 2, 12-58.

Fowler, C.A. 1986. "An Event Approach to the Study of Speech Perception from a Direct-Realist Perspective". *Journal of Phonetics* 14, 3-28.

Fraser, H. 1992. *The Subject of Speech Perception.* Basingstoke, Hampshire and London: Macmillan Press Ltd.

Fries, C. 1964. "On the Intonation of 'Yes-No' Questions in English". In D. Abercrombie, D. Fry, P. Mc Carthy, N. Scotten and J. Trim (eds). *In Honour of Daniel Jones.* London: Longman.

Fromkin, V.A. 1976. "The Interface Between Phonetics and Phonology". *UCLA Working Papers in Phonetics,* 31, 104-107.

Fry, D.B. 1958. "Experiments in the Perception of Stress". *Language and Speech* 1, 126-152.

Fry, D.B. 1960. "Linguistic Theory and Experimental Research". *Transactions of the Philological Society,* 13-39.

Gårding, Eva and Abramson, Arthur S. 1965. "A Study of the Perception of some American English Intonation Contours". *Studia Linguistica,* 19, 61.

Gårding, Eva and House, David. 1987. "Production and Perception of Phrases in some Nordic Dialects". In P. Lilius and M. Saari (eds). *The Nordic Languages and Modern Linguistics,* 6. Helsinki: Helsinki University Press, 163-175.

Gårding, Eva and Eriksson, L. 1989. "Perceptual Cues to some Swedish Prosodic Phrase Patterns - a Peak Shift Experiment". *Speech Transmission Laboratory Quarterly Progress and Status Report,* 1, Stockholm: Royal Institute of Technology, 13-16.

Garman, M. 1990. *Psycholinguistics.* Cambridge: Cambridge University Press.

Garnes, S. 1973. "Phonetic Evidence Supporting a Phonological Analysis". *Journal of Phonetics,* 1, 273-283.

Garnes, S. 1974. "Suprasegmental Aspects of Icelandic Vowel Quality". *Working Papers in Linguistics,* 17, 144-159.

Garret, M.F. 1988. "Processes in Language Production". In F.J. Newmeyer (ed.). *Linguistics: the Cambridge Survey: Language: Psychological and Biological Aspects* (Vol III). Cambridge: Cambridge University Press, 69-96.

Garro, L. and Parker, F. 1982. "Some Suprasegmental Characteristics of Relative Clauses in English". *Journal of Phonetics,* 10, 149-161.

Garro, L. and Parker, F. 1983. "Relative Clauses in Spanish: Some Suprasegmental Characteristics". *Journal of Phonetics,* 11, 85-89.

Geluykens, Ronald. 1987. "Intonation and Speech Act Type: An Experimental Approach to Rising Intonation in Queclaratives". *Journal of Pragmatics,* 11 (4), 483-494.

Geluykens, Ronald. 1988. "On the Myth of Rising Intonation in Polar Questions". *Journal of Pragmatics,* 12, 467-485.

Geluykens, Ronald. 1989. "R(a)ising Questions: Question Intonation Revisited". *Journal of Pragmatics,* 13, 567-575.

Goldsmith, J.A., Peterson, K. and Drogo, J. 1989. "Tone and Accent in the Xhosa Verbal System". In P. Newman and R. Botne (eds). *Current Approaches to African Linguistics,* 5, Dordrecht: Foris Publications, 157-178.

Grice, M.L. 1995. *The Intonation of Interrogation in Palermo Italian.* Tubingen: Niemeeyer.

Grosjean, F. G. 1996. "Gating". *Language and Cognitive Processes,* 11 (6), 597-604.

Guy, G.R. and Vonwiller, J. 1984. "The Meaning of Intonation in Australian English". *Australian Journal of Linguistics,* 4, 1-17.

Hadding-Koch, Kerstin. 1961. *Acoustico-Phonetic Studies in the Intonation of Southern Swedish.* Lund, Sweden: C.W.K. Gleerup.

Hadding-Koch, Kerstin. 1965. "On the Physiological Background of Intonation". *Studia Linguistica,* 19, 55.

Hadding-Koch, K. and Studdert-Kennedy, M. 1963. "A Study of Semantic and Psychophysical Test Responses to Controlled Variations in Fundamental Frequency". *Studia Linguistica,* 17, 65.

Hadding-Koch, Kerstin and Studdert-Kennedy, Michael. 1964. "An Experimental Study of some Intonation Contours". *Phonetica*, 11, 175.

Hadding-Koch, Kerstin and Studdert-Kennedy, Michael. 1965. "Intonation Contours Evaluated by American and Swedish Test Subjects". *Proceedings of the Vth International Congress of Phonetic Science, Munster, 1964* (Basel), 326.

Halle, M. 1954. "Why and How do We Study the Sounds of Speech". In Meuller, H.J. (ed.) *Report of the Fifth Annual Round Table Meeting on Linguistics Language Teaching.* Washington DC: Georgetown University Press.

Halle, M. and Stevens, K.N. 1964. "Speech Recognition: A Model and a Program for Research". In Fodor, J.A. and Katz, J.J. (eds) *The Structure of Language: Readings in the Philosophy of Language.* Engelwood Cliffs: Prentice-Hall, 604-612.

Harris, Alice C. 1984. "Georgian". In W.S. Chisholm Jr. (ed.). *Typological Studies in Language: Interrogativity,* (Vol. 4). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Herman, Rebecca. 1996. "Final Lowering in Kipare". *Phonology,* 13 (2), 171-196.

Hermann, Eduard. 1942. "Problem der Frage", *Nachrichten von der Akademie der Wissenschaften in Göttingen, Philologische Historische Klasse.* Göttingen: Vandenhoek & Ruprecht, 121-408.

Hertz, Susan. 1997. (January 12-last update). "The Technology of Text-to-Speech". [Homepage of Eloquent Technologies, Inc.], [Online]. Available: www.eloq.com/SuePap.htm. (2000)

Hinds, John. 1984. "Japanese". In W.S. Chisholm Jr. (ed.). *Typological Studies in Language: Interrogativity* (Vol. 4). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hirose, K. 1997. "Disambiguating Recognition Results by Prosodic Features". In Y. Sagisaka, N. Campbell and N. Higuchi (eds). *Computing Prosody.* Heidelberg: Springer, 327-342.

Hirose, K. and Iwano, K. 1997. "A Method of Representing Fundamental Frequency Contours of Japanese using Statistical Models of Moraic Transition". In *Proceedings of EUROSPEECH 97.* Rhodes, 311-314.

Hirschberg, Julia and Ward, Gregory. 1992. "The Influence of Pitch Range, Duration, Amplitude and Spectral Features on the Interpretation of the Rise-Fall-Rise Intonation Contour in English". *Journal of Phonetics,* 20, 241-251.

Ho, Aichen T. 1977. "Intonation Variations in a Mandarin Sentence for Three Expressions: Interrogative, Exclamatory, and Declarative". *Phonetica,* 34, 446-456.

Hörmann, H. 1979. *Psycholinguistics An Introduction to Research and Theory* (Second Edition, Revised). Springer-Verlag, Inc: New York.

House, David. 1990. *Tonal Perception in Speech.* Sweden: Lund University Press.

Jones, D. 1909. *Intonation Curves.* Leipzig and Berlin: B.G. Teubner.

Jones, C.J.J.; Louw, J.A.N. and Roux, J.C. 1998. "Queclaratives in Xhosa". *South African Journal of Linguistics*, 36, 3-18.

Jones, C.J.J.; Louw, J.A.N. and Roux, J.C. 1998. "Perceptual Experiments on Queclaratives in Xhosa". *South African Journal of Linguistics*, 36, 19-32.

Jun, Sun-Ah and Oh, Mira. 1994. "A Prosodic Analysis of Three Sentence Types with 'WH' Words in Korean". *Proceedings of the 1994 International Conference on Spoken Language Processing.* Tokyo: The Acoustical Society of Japan, 323-326.

Jun, Sun-Ah and Oh, Mira. 1996. "A Prosodic Analysis of Three Types of Wh-Phrases in Korean". *Language and Speech,* 39 (1), 37-61.

Jusczyk, P. W.; Friederici, A.D.; Wessels, J.; Svenkerud, V.Y. and Jusczyk, A.M. 1993. "Infants' Sensitivity to the Sound Patterns of Native Language Words". *Journal of Memory and Language,* 32, 402-420.

Jusczyk, P. W., Houston, D. and Goodman, M. 1998. "Speech Perception During the First Year". In Alan Slater (ed.). *Perceptual Development Visual, Auditory, and Speech Perception in Infancy.* Psychology Press Ltd.: East Essex, UK.

Keating, P.A. 1988. "The Phonology-Phonetics Interface". In Newmeyer, F.J. (ed.), *Linguistics: The Cambridge Survey,* Vol. 1. *Linguistic Theory: Foundations.* Cambridge: Cambridge University Press, 281-302.

Khoali, Benjamin T. 1993. *A Sesotho Tonal Grammar.* Ann Arbor: UMI Press.

Khoali, Maphutshe H.E. 1994. *Interrogative Structures in Setswana - a Functional Approach.* (Unpublished M.A. dissertation). Potchefstroom University for Christian Higher Education, Potchefstroom.

Khumalo, James S.M. 1981. *Zulu Tonology* (Unpublished M.A. dissertation). University of the Witwatersrand, Johannesburg.

Khumalo, James S.M. 1987. *An Autosegmental Account of Zulu Phonology*. (Unpublished Ph.D. thesis). University of the Witwatersrand, Johannesburg.

Klatt, D.H. 1976. "Linguistic Uses of Segmental Duration in English: Acoustic Perceptual Evidence". *Journal of the Acoustical Society of America,* 59, 1208-1221.

Kohler, K.J. 1987. "Catergorical Pitch Perception". In U. Viks (ed.). *Proceedings of the Eleventh International Congress of Phonetic Sciences,* 5. Tallinn: Academy of Sciences of the Estonian S.S.R., 331-333.

Kreyszig, E. 1988. *Advanced Engineering Mathematics,* (Second Edition). New York: John Wiley & Sons.

Kropf, A. 1915. *A Kafir-English Dictionary.* Lovedale, South Africa: Lovedale Mission Press.

Kunene, D.P. 1961. *The Sound System of Southern Sotho.* (Unpublished Ph. D thesis). University of Cape Town, Cape Town.

Ladefoged, P. 1988. *The Many Interfaces between Phonetics and Phonology.* Papers presented at the Sixth International Phonology Conference, Krems, Austria.

"Language Engineering; Harnessing the power of language". 2000. (last update unavailable) [Online]. Available: www.linglink.lu/le/en/broch/harness.html. (2000)

Lanham, L.W. 1958. "The Tonemes of Xhosa". *African Studies,* 17, 65-81.

Lanham, L.W. 1960. "The Comparative Phonology of Nguni". (Unpublished Ph. D. thesis). University of the Witwatersrand, Johannesburg.

Lanham, L.W. 1963. "The Tonemes of Xhosa: a Restatement". *Studies in Linguistics*, 17, 35-58.

Laughren, Mary. 1984. "Tone in Zulu Nouns". In G.N. Clements and J. Goldsmith (eds). *Autosegmental Studies in Bantu Tone.* Dordrecht: Foris Publications.

Laver, John. 1989. "Cognitive Science and Speech: a Framework for Research". In H. Schnelle and N-O. Bernsen (eds.). *Logic and Linguistics.* Hillsdale NJ, Hove and London: Lawrence Erlbaum Associates, 37-69. (Reprinted in Laver (1991) *The Gift of Speech: Papers in the Analysis of Speech and Voice.* Edinburgh: Edinburgh University Press.

Laver, John.1994. *Principles of Phonetics.* Cambridge: Cambridge University Press.

Lecanuet, Jean-Pierre. 1998. "Foetal Responses to Auditory and Speech Stimuli". In Alan Slater (ed.) *Perceptual Development Visual, Auditory, and Speech Perception in Infancy.* East Essex, UK: Psychology Press Ltd., 313-355.

Lehiste, I. 1970. *Suprasegmentals.* Cambridge, MA.: MIT Press.

Lehiste, I., Olive, J.P. and Streeter, L.Y. 1976. "Role of Duration in Syntactic Ambiguous Sentences". *Journal of the Acoustical Society of America,* 60, 1199-1202.

Lehiste, I. 1980. "Phonetic Manifestation of Syntactic Structure in English". *Annual Bulletin*, RILP, 14. University of Tokyo, 1-27.

Lewis, M. 1936. *Infant Speech: A Study of the Beginnings of Language*. New York: Harcourt Brace.

Li, Xai and Crane, Nancy B.1993. *Electronic Style: A Guide to Citing Electronic Information.* United States of America: Meckler Publishing.

Liberman, A.M., Cooper, F.S., Shankweiler,D.S. and Studdert-Kennedy, M. 1967. "Perception of the Speech Code". *Psychological Review,* 74, 431-461.

Liberman, A.M. and Studdert-Kennedy, M. 1978. "Phonetic Perception". In R. Held, H.W. Leibowitz and H-L. Teuber (eds). *Handbook of Sensory Physiology,* Volume VIII. Berlin: Springer, 143-178.

Liberman, A.M. and Mattingly, I.G. 1985. "The Motor Theory of Speech Perception Revised". *Cognition,* 21, 1-36.

Lieberman, P. 1967. *Intonation, Perception and Language, Research Monograph,* 38. Cambridge, MA: M.I.T. Press.

Lombard, D.P. 1980. "Duur en Lengte in Noord-Sotho". *Studies in Bantoetale. Jaargang* 7.1, Universiteit van Pretoria.

Louw, J.A. 1968. *The Intonation of the Sentence and its Constituent Parts in Xhosa and Tsonga.* Pretoria: University of South Africa.

Louw, J.A. 1969. "Some Remarks on Zulu tone". *Limi,*7, 45-49.

Louw, J.A. 1971. "The Tonal Paradigm of the Verb in Xhosa". In V. Six, N. Cyffer, E. Wolff, L. Gerhardt and H. Meyer-Bahlburg (eds). The Festschrift in honour of Prof J. Lukas. *Afrikanische Sprachen und Kulturen-Ein Querschnitt.* Deutsches Institut für Afrika-Forschung, Hamburg, 102-113.

Louw, J.A. 1995. "Xhosa Tone". In A. Traill, R. Vossen and M. Biesele (eds). *The Complete Linguist* (Papers in memory of Patrick J. Dickens). Köln: Rüdiger Köppe, 237-270.

Louwrens, L.J. 1987. "Some Pragmatic Functions of Interrogative Particles in Northern Sotho". *South African Journal of African Languages*, 7 (4), 121-124.

Louwrens, L.J. 1991. *Aspects of Northern Sotho Grammar.* Pretoria: Via Afrika Ltd.

Massaro, Dominic W. 1975. *Experimental Psychology and Information Processing.* Chicago, USA: Rand McNally College Publishing Company.

Massaro, D.W. and Cohen, M.M. 1975. "Preperceptual Auditory Storage in Speech Perception". In A. Cohen and S. G. Nooteboom (eds). *Structure and Process in Speech Perception.* Berlin, Heidelberg and New York: Springer Verlag.

Mathibela, Mathlodi. A. 1989. *Question Formation in Tswana.* (Unpublished Honours article). University of South Africa, Pretoria.

Mehler, Jacques; Bertoncini, Josiane; Dupoux, Emmanuel and Pallier, Christophe. 1996. "The Role of Suprasegmentals in Speech Perception and Acquisition". In Takashi Otake and Anne Cutler (eds). *Phonological Structure and Language Processing: Cross Linguistic Studies, Speech Research,* 12. Berlin, New York: Mouton de Gruyter, 145-169.

Mehler, J., Lamberz, G., Jusczyk, P. and Amiel-Tison, C. 1986. "Discrimination de la langue maternelle par le nouveau-nϑ". *Compte rendus de l' Acad mie des Sciences.* Paris, 303, 637-640.

Mehler, J.; Jusczyk, P.; Lamberz, G.; Halsted, N.; Bertoncini, J. and Amiel-Tison, C. 1988. "A Precursor of Language Aquisition in Young Infants". *Cognition,* 29, 143-178.

Mehta, G. and Cutler, A. 1988. "Detection of Target Phonemes in Spontaneous and Read Speech". *Language and Speech,* 31, 135-156.

Miura, Ichiro and Hara, Noriyo. 1995. "Production and Perception of Rhetorical Questions in Osaka Japanese". *Journal of Phonetics,* 23, 291-303.

Moon, C., Panneton-Cooper, R.P. and Fifer, W.P. 1993. "Two-day-olds Prefer their Native Language". *Infant Behavior and Development,* 16, 495-500.

Moulines, E. and Charpentier, F. 1990. "Pitch-synchronous Waveform Processing Techniques for Text-to-Speech Synthesis using Diphones". *Speech Communication,* 9, 453-467.

Myers, S. 1999. "Downdrift and Pitch Range in Chichewa Intonation". *Proceedings of the 14th International Congress of Phonetic Sciences.* San Francisco, USA: ICPHS.

Neisser, U. 1976. *Cognition and Reality.* San Francisco: W.H. Freeman and Company.

Nkabinde, A.C. 1999. "Some Features of Zulu Nouns". In R. Finlayson (ed.). *African Mosaic*, Pretoria: Unisa Press, 377-386.

Nooteboom, Sibout G., Brokx, Johannes P.L. and De Rooij, Jacobus J. 1978. "Contributions of Prosody to Speech Perception". In W.J.M. Levelt and Flores d'Arcais (eds). *Studies in the Perception of Speech.* Bath: Pitman Press Ltd., 75-107.

Ohala, J.J. 1978. "Production of Tone". In V. Fromkin (ed.) *Tone: a Linguistic Survey.* New York: Academic, 5-39.

Ohala, J.J. 1983. "Cross-language use of Pitch: An Ethological View". *Phonetica* 40, 1-18.

Ohala, J.J. 1990. "There is No Interface between Phonology and Phonetics: A Personal View". *Journal of Phonetics,* Vol. 18, no.2, 153-171.

Ọmọruyi, Thomas, O. 1988. "On the Formation of Questions in Ẹdo". *Journal of African Linguistics,* 10 (1), 19-31.

O' Shaughnessy, D. 1979. "Linguistic Features in Fundamental Frequency Patterns". *Journal of Phonetics,* 7, 119-145.

Otake, Takashi and Cutler, Anne. 1999. "Perception of Suprasegmental Structure in a Nonnative Dialect". *Journal of Phonetics,* 27, 229-253.

Pahl, H.W. 1977. "The Distribution and Functional Roles of Certain Significant Tones and Tonal Sequences in Xhosa 1. The Falling Tone". *Limi,* 5, 1-18.

Pahl, H.W. 1978. "The Distribution and Roles of Certain Significant Tones and Tonal Sequences in Xhosa. II. The Terminal HH Tone Sequence and the Reversal of Tones". *Limi* 6, 53-65.

Pahl, H.W. (ed.). 1989. *The Greater Dictionary of Xhosa*. Alice: University of Fort Hare.

Picket, M. 1980. *The Sounds of Speech Communication.* Baltimore: University Park Press.

Pierrehumbert, J. 1979. "The Perception of Fundamental Frequency Declination". *Journal of the Acoustical Society of America*, 66 (2), 363-369.

Pierrehumbert, J.B. and Steele, S.A. 1989. *The Phonology and Phonetics of English Intonation.* (Unpublished Ph.D. thesis). Cambridge, MA: MIT.

Pike, K.L. 1945. *The Intonation of American English.* Ann Arbor: University of Michigan Press.

Pisoni, D. B. and Sawusch, J.R. 1975. "Some Stages of Processing in Speech Perception". In A. Cohen and S.G. Nooteboom (eds). *Structure and Process in Speech Perception.* Berlin, Heidelberg and New York: Springer Verlag.

Ponelis, Fritz. 1974. "On the Dynamics of Velarization and Labialization: Some Bantu Evidence". *Studies in African Linguistics,* Vol. 5, no. 1, 27-58.

Poulos, George. 1990. *A Linguistic Analysis of Venda.* Pretoria: Via Afrika Ltd.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S. and Fong, C. 1991. "The Use of Prosody in Syntactic Disambiguation". *Journal of the Acoustical Society of America*, 90, 2956-2970.

Repp, B. 1988. "Integration and Segregation in Speech Perception". *Language Speech,* 31, 239-271.

Restan, Per. 1972. *Sintaksis voprositel'nogo predloženija: obŝĉij vopros (glavnym obrazom na materiale sovremennogo russkogo jazyka).* [Syntax of the Interrogative Sentence: the General Question (Mainly on the Basis of Modern Russian Material)]. Oslo-Bergen-Troms : Universitetsforlaget.

Riordan, J. 1969. *Lumko Xhosa Self instruction Course.* Lady Frere, South Africa: Lumko Institute.

Roux, Justus C. "On the Prosodics of Declarative and Question Sentences in Xhosa". (unpublished article).

Roux, Justus C. 1989. "Grapheme-to-Phoneme Conversions in Xhosa". *South African Journal of African Languages,* 9 (2), 74-78.

Roux, Justus C. 1978. "Phonetic Data and Phonological Analysis". *SPIL: Stellenbosch Papers in Linguistics* 1, 105-133.

Roux, Justus C. 1991. "On the Integration of Phonetics and Phonology". *South African Journal of Linguistics,* 11, 34-52.

Roux, Justus C. 1998. "Xhosa: A Tone or Pitch-accent Language". *South African Journal of Linguistics,* 36, 33-50.

Roux, Justus, C., Botha, E.C. and Du Preez, J.A. 2000. "Developing a Multilingual Telephone Based Information System in African Languages". *Proceedings of the Second International Conference on Language Resources and Evaluation, Athens, Greece.* ELRA, 975-980.

Rycroft, D.K. and Ngcobo, A.B. 1979. *Say it in Zulu.* London: School of Oriental and African Studies.

Sachs, L. 1984. *Applied Statistics.* New York: Springer Verlag.

Saha, P.K. 1984. "Bengali". *Typological Studies in Language: Interrogativity* (Vol.4). Amsterdam: John Benjamins Publishing Company.

Scott, D.R. 1982. "Duration as a Cue to the Perception of a Phrase Boundary". *Journal of the Acoustical Society of America,* 66, 363-369.

Sekiguchi, T. and Nakajima, Y. 1999. "The Use of Lexical Prosody for Lexical Access of the Japanese Language". *Journal of Psycholinguistic Research,* 28, 439-454.

Shen, Xiao-nan S. 1990. "The Prosody of Mandarin Chinese". Berkeley: University of California Press.

Shen, Jong. 1985. "Bějīng huà shēngdiào de yīnyù hé yǔdiào". In Tao Lin and Lijia Wang (eds). *Běijīng yǔyīn shíyàn.* Beijing: Beijing University Press, 73-130.

Snyman, J.W., Le Roux J.C. and Le Roux, M. 1991. *Tswana for Beginners.* Pretoria: University of South Africa.

Solomon, R.C. 1996. *Statistics.* London: John Murray Publisher.

Stevens, K.N. and Blumenstein, S.E. 1981. "The Search for Invariant Acoustic Correlates of Phonetic Features". In P. Eimas and J. Miller (eds). *Perspectives on the Study of Speech.* Hillsdale: Erlbaum, 1-38.

Stevens, K.N. and Halle, M. 1967. "Remarks on Analysis by Synthesis and Distinctive Features". In W. Wathen-Dunn (ed.). *Models for the Perception of Speech and Visual Form.* Cambridge, MA: MIT Press, 88-102.

Steyn, Elna. 1991. *Vraagwoorde en Vraagpartikels in Noord-Sotho*. (Unpublished M.A. dissertation). University of Pretoria, Pretoria.

Strange, Winifred. 1995. "Cross-language Studies of Speech Perception: A Historical Review". *Speech Perception and Linguistic Experience.* Maryland, USA: York Press, Inc.

Streeter, L.A. 1978. "Acoustic Determinants of Phrase Boundary Perception". *Journal of the Acoustical Society of America,* 64, 1582-1592.

Studdert-Kennedy, Michael and Hadding, Kerstin. 1973. "Auditory and Linguistic Processes in the Perception of Intonation Contours". *Language and Speech*, 16, 293-313.

Studdert-Kennedy, Michael. 1998. "Introduction: the Emergence of Phonology". In James, R. Hurford, Michael Studdert-Kennedy and Chris Knight (eds). *Approaches to the Evolution of Language*. Cambridge: Cambridge University Press, 169-176.

Svensson, S.G. 1974. *Prosody and Grammar in Speech Perception.* Monographs from the Institute of Linguistics 2. Stockholm: University of Stockholm.

T'Hart, J., Collier, R. and Cohen, R. 1990. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody.* Cambridge: Cambridge University Press.

Theron, S.C. 1991. *Fonetiese en Fonologiese Aspekte van Stel- en Vraagsinne in Xhosa.* (Unpublished M.A. dissertation). University of Stellenbosch, Stellenbosch.

Thomas, Elaine. 1974. "Engenni". In J. Bendor-Samuel (ed.). *Ten Nigerian Tone Systems (Studies in Nigerian Languages,* 4*).* Jos: Institute of Linguistics, 13-26.

Touati, P. 1987. *Structure prosodiques du su9dois et du francais.* Lund: Lund University Press.

Thorsen, Nina G. 1980. "A Study of the Perception of Sentence Intonation. Evidence from Danish". *Journal of the Acoustical Society of America,* 67, 1014-1030.

Tucker, A.N. 1929. *The Comparative Phonetics of the Suto-Chuana Group of Bantu Languages.* Longmans, Green and Company.

Ultan, R. 1978. "Some General Characteristics of Interrogative Systems". In J.P. Greenberg, C.A. Ferguson and E.A. Moravscik (eds). *Universals of Human Language: Syntax* (Vol.4). Stanford: Stanford University Press, 211-248.

Ultan, R. 1969. "Some General Characteristics of Interrogative Systems". *Working Papers on Language Universals* I, 41-63a.

Uyeno, T., Hayashibe, H., Imai, K., Imagawa H. and Kiritani, S. 1980. "Comprehension of Relative Clause Construction and Pitch Contours in Japanese". *Annual Bulletin,* RILP, 14. University of Tokyo, 225-236.

Waibel, Alex and Lee, Kai-Fu (eds), 1990. *Readings in Speech Recognition.* United States: Morgan Kaufmann Publishers, Inc.

Wales, R. and Taylor, S. 1987 "Intonation Cues to Questions and Statements: How are they Perceived?" *Language and Speech,* 30 (3), 199-211.

Ward, G. and Hirschberg, J. 1988. "Intonation and Propositional Attitude: The Pragmatics of L* + HLH %". *Proceedings of theFifth Meeting of the Eastern States Conference on Linguistics (ESCOL)* Vol. 5, Philadelphia, 512-522.

Witten, I. H. 1982. *Principles of Computer Speech.* New York: Academic Press.

Woodward, S.C. 1992. *The Transmission of Music into the Human Uterus and the Response to Music of the Human Fetus and Neonate.* (Unpublished Ph.D. thesis). University of Cape Town., Cape Town.

Ziervogel, D. 1976. *Handboek van Noord-Sotho.* Van Schaik: Pretoria.

APPENDIX A

ACOUSTIC ANALYSIS DATA

A.1 Corpus

Table A.1 Data corpus with transcription and tags.

Word	Transcription	Tags	Identifier
ngumntu	ŋgumnt'u	n1 u1 m n2 u2	A01
ngumfazi	ŋgumfazi	n u m f a z i	A02
ngumfana	ŋgumfana	n1 u m f a1 n2 a2	A03
ngabantu	ŋgaɓant'u	n1 a1 b a2 n2 u	B01
ngabafazi	ŋgaɓafazi	n a1 b a2 f a3 z i	B02
ngabafana	ŋgaɓafana	n1 a1 b a2 f a3 n2 a4	B03
ngumthi	ŋgumtʰi	n u m t i	C01
ngumlilo	ŋgumlilɔ	n u m l1 i l2 o	C02
ngumfula	ŋgumfula	n u1 m f u2 l a	C03
yimithi	jimitʰi	y i1 m i2 t i3	D01
yimililo	jimililɔ	y i1 m i2 l1 i3 l2 o	D02
yimifula	jimifula	y i1 m i2 f u l a	D03
lilitye	lilic'ɛ	l1 i1 l2 i2 t e	E01
liliso	lilisɔ	l1 i1 l2 i2 s o	E02
lilifu	lilifu	l1 i1 l2 i2 f u	E03
ngamatye	ŋgamac'ɛ	n a1 m a2 t e	F01
ngamehlo	ŋgamɛɬɔ	n a m e h o	F02
ngamafu	ŋgamafu	n a1 m a2 f u	F03
sisitya	sisic'a	s1 i1 s2 i2 t a	G01
sisifo	sisifɔ	s1 i1 s2 i2 f o	G02
sisilo	sisilɔ	s1 i1 s2 i2 l o	G03
zizitya	zizic'a	z1 i1 z2 i2 t a	H01
zizifo	zizifɔ	z1 i1 z2 i2 f o	H02
zizilo	zizilɔ	z1 i1 z2 i2 l o	H03
yinja	jiɲdʒa	y i n a	I01
yintloko	jintɬ'ɔk'ɔ	y i n o1 k o2	I02
yinkomo	jiŋk'ɔmɔ	y i n o1 m o2	I03
zizinja	ziziɲdʒa	z1 i1 z2 i2 n a	J01
ziintloko	ziːntɬ'ɔk'ɔ	z i n o1 k o2	J02
ziinkomo	ziːŋk'ɔmɔ	z i n o1 m o2	J03
luluthi	lulutʰi	l1 u1 l2 u2 t i	K01
lulwimi	lulwimi	l1 u l2 i1 m i2	K02
lulwandle	lulwandɮɛ	l1 u l2 a n e	K03
bubuso	ɓuɓusɔ	b1 u1 b2 u2 s o	L01
bubuhlanti	ɓuɓuɬant'i	b1 u1 b2 u2 h a n i	L02
bubusika	ɓuɓusik'a	b1 u1 b2 u2 s i k a	L03
kukutya	k'uk'uc'a	k1 u1 k2 u2 t a	M01
kukwindla	k'uk'windɮa	k1 u k2 i n a	M02
kukunene	k'uk'unɛnɛ	k1 u1 k2 u2 n1 e1 n2 e2	M03

A.2 Duration data

Table A.2 Mean phoneme and word duration [in ms].

ngumntu (A01)									
	d_{n1}	d_{u1}	d_m	d_{n2}	d_{u2}				$d_{ngumntu}$
Queclarative	39.9	108.4	135.3	118.1	80.9				482.6
Statement	97.5	133.4	235.1	144	74.5				684.5
% change	144.4%	23.1%	73.8%	21.9%	-7.9%				41.8%
ngumfazi (A02)									
	d_n	d_u	d_m	d_f	d_a	d_z	d_i		$d_{ngumfazi}$
Queclarative	44	49.3	103.5	118.4	191.2	137.5	40.8		684.7
Statement	39.8	77.7	121	127.1	323.5	142.7	34.6		866.5
% change	-9.5%	57.6%	16.9%	7.3%	69.2%	3.8%	-15.2%		26.6%
ngumfana (A03)									
	d_{n1}	d_u	d_m	d_f	d_{a1}	d_{n2}	d_{a2}		$d_{ngumfana}$
Queclarative	44.6	57.1	107.2	105.6	182.7	79.2	66.9		643.5
Statement	61.2	61.7	120.6	131.1	333.4	86.7	66.9		861.7
% change	37.2%	8.1%	12.5%	24.1%	82.5%	9.5%	0%		33.9%
ngabantu (B01)									
	d_{n1}	d_{a1}	d_b	d_{a2}	d_{n2}	d_u			$d_{ngabantu}$
Queclarative	69.6	106.5	63.7	187	126.9	56.7			610.4
Statement	79.2	106.4	76.9	320.1	157.4	48.9			788.9
% change	13.8%	-0.1%	20.7%	71.2%	24%	-13.8%			29.2%
ngabafazi (B02)									
	d_n	d_{a1}	d_b	d_{a2}	d_f	d_{a3}	d_z	d_i	$d_{ngabafazi}$
Queclarative	57.7	81.6	56	101	114	198.9	112.8	57.6	779.6
Statement	70.7	89.6	60.5	104.6	134.9	329.5	125.2	39.8	954.8
% change	22.5%	9.8%	8%	3.6%	18.3%	65.7%	11%	-30.9%	22.5%
ngabafana (B03)									
	d_{n1}	d_{a1}	d_b	d_{a2}	d_f	d_{a3}	d_{n2}	d_{a4}	$d_{ngabafana}$
Queclarative	67.8	88.2	56.5	103.6	118	168.1	80.1	57.8	740.1
Statement	80.9	85.4	62.3	103.6	145.8	328.7	94	53.9	954.7
% change	19.3%	-3.2%	10.3%	0%	23.6%	95.5%	17.4%	-6.7%	29%
ngumthi (C01)									
	d_n	d_u	d_m	d_t	d_i				$d_{ngumthi}$
Queclarative	74.5	109.6	182.6	106.3	129.1				602.1
Statement	74.5	144.1	272.5	184.4	122.6				798.1
% change	0%	31.5%	49.2%	73.5%	-5%				32.6%
ngumlilo (C02)									
	d_n	d_u	d_m	d_{l1}	d_i	d_{l2}	d_o		$d_{ngumlilo}$
Queclarative	50.5	78.9	88.4	67.2	169	65.5	81.7		601.2
Statement	59.3	71	118.1	103.4	281.5	109.1	77.4		819.7
% change	17.4%	-10%	33.6%	53.9%	66.6%	66.6%	-5.3%		36.3%
ngumfula (C03)									
	d_n	d_{u1}	d_m	d_f	d_{u2}	d_l	d_a		$d_{ngumfula}$
Queclarative	56.2	76.6	106.5	129.2	153.7	85.2	74.6		682
Statement	68.9	61.7	143.4	163.3	291.4	89.5	77		895.2
% change	22.6%	-19.5%	34.6%	26.4%	89.6%	5%	3.2%		31.3%

Table A.2 Mean phoneme and word duration [in ms] (continued).

yimithi (D01)									
	d_y	d_{i1}	d_m	d_{i2}	d_t	d_{i3}			$d_{yimithi}$
Queclarative	37.1	83.2	88.1	167.9	129.2	108.7			614.2
Statement	48.3	86.1	113.6	293.2	178.8	118.1			838.2
% change	30.2%	3.5%	28.9%	74.6%	38.4%	8.6%			36.5%
yimililo (D02)									
	d_y	d_{i1}	d_m	d_{i2}	d_{l1}	d_{i3}	d_{l2}	d_o	$d_{yimililo}$
Queclarative	28.1	72	84.7	89.4	81.3	143.5	77	77.1	653.1
Statement	24.1	83.1	92.9	102.5	103.3	277.3	96	72.6	851.8
% change	-14.2%	15.4%	9.7%	14.7%	27.1%	93.2%	24.7%	-5.8%	30.4%
yimifula (D03)									
	d_y	d_{i1}	d_m	d_{i2}	d_f	d_u	d_l	d_a	$d_{yimifula}$
Queclarative	15.8	79.8	80	85.1	153.7	146.2	75.4	73	709
Statement	26.8	85.8	88	107.5	172.8	277.6	90.5	68.1	917
% change	69.6%	7.5%	10%	26.3%	12.4%	89.9%	20%	-6.7%	29.3%
lilitye (E01)									
	d_{l1}	d_{i1}	d_{l2}	d_{i2}	d_t	d_e			$d_{lilitye}$
Queclarative	53	86	86.9	162.2	108	93.5			589.6
Statement	61.4	110.8	110.6	320.6	124.3	100.5			828.1
% change	15.8%	28.8%	27.3%	97.7%	15.1%	7.5%			40.5%
liliso (E02)									
	d_{l1}	d_{i1}	d_{l2}	d_{i2}	d_s	d_o			d_{liliso}
Queclarative	51.4	93.5	74.1	146	145.6	109.3			620.1
Statement	57.2	104.5	105.1	257.7	191.1	102.3			817.8
% change	11.3%	11.8%	41.8%	76.5%	31.3%	-6.4%			31.9%
lilifu (E03)									
	d_{l1}	d_{i1}	d_{l2}	d_{i2}	d_f	d_u			d_{lilifu}
Queclarative	49.1	102	83	156	151.2	86.1			627.3
Statement	62.1	104.3	116.5	278.8	186.1	84.4			832.1
% change	26.5%	2.3%	40.4%	78.7%	23.1%	-2%			32.6%
ngamatye (F01)									
	d_n	d_{a1}	d_m	d_{a2}	d_t	d_e			$d_{ngamatye}$
Queclarative	73.9	99.1	93.2	180.7	115.4	86.9			649.2
Statement	80.7	116	116.5	344.9	137.9	93.5			889.4
% change	9.2%	17.1%	25%	90.9%	19.5%	7.6%			37%
ngamehlo (F02)									
	d_n	d_a	d_m	d_e	d_h	d_o			$d_{ngamehlo}$
Queclarative	64.3	110.1	90.5	183.8	141.8	78.9			669.3
Statement	76.5	118.5	113.8	334.6	178.8	76.1			898.4
% change	19%	7.6%	25.7%	82%	26.1%	-3.5%			34.2%
ngamafu (F03)									
	d_n	d_{a1}	d_m	d_{a2}	d_f	d_u			$d_{ngamafu}$
Queclarative	61.8	113.2	87.9	183.6	139.9	78.5			664.8
Statement	70.4	118.4	101.4	319.3	168.2	69.1			846.7
% change	13.9%	4.6%	15.4%	73.9%	20.2%	-12%			27.4%
sisitya (G01)									
	d_{s1}	d_{i1}	d_{s2}	d_{i2}	d_t	d_a			$d_{sisitya}$
Queclarative	88.7	92.6	133.6	164	100.2	96.5			675.6
Statement	130.7	113.2	172.7	293	120.4	96.1			926.1
% change	47.4%	22.2%	29.3%	78.7%	20.2%	-0.4%			37.1%

Table A.2 Mean phoneme and word duration [in ms] (continued).

sisifo (G02)							
	d_{s1}	d_{i1}	d_{s2}	d_{i2}	d_f	d_o	d_{sisifo}
Queclarative	99.3	97.6	120.2	138.2	128.5	124.5	708.3
Statement	124.6	106	166.5	260.4	173.9	129.7	961.2
% change	25.5%	8.6%	38.5%	88.4%	35.3%	4.2%	35.7%
sisilo (G03)							
	d_{s1}	d_{i1}	d_{s2}	d_{i2}	d_l	d_o	d_{sisilo}
Queclarative	106.8	93.3	147.9	148.8	92	87.6	676.3
Statement	117.1	108.6	184.9	295.8	101.3	104.6	912.4
% change	9.6%	16.4%	25%	98.8%	10.1%	19.4%	34.9%
zizitya (H01)							
	d_{z1}	d_{i1}	d_{z2}	d_{i2}	d_t	d_a	$d_{zizitya}$
Queclarative	85.9	126.3	105.1	186.4	98.8	101.5	703.9
Statement	109.8	131.3	159.2	312.5	120.6	101.5	935
% change	27.8%	4%	51.5%	67.7%	22.1%	0%	32.8%
zizifo (H02)							
	d_{z1}	d_{i1}	d_{z2}	d_{i2}	d_f	d_o	d_{zizifo}
Queclarative	81.8	138.8	88.9	126.1	125.5	114.3	675.5
Statement	109.8	161.7	123.7	232.9	178.4	133.6	940
% change	34.2%	16.5%	39.1%	84.7%	42.2%	16.9%	39.2%
zizilo (H03)							
	d_{z1}	d_{i1}	d_{z2}	d_{i2}	d_l	d_o	d_{zizilo}
Queclarative	83.1	120	115.3	156.4	86.7	101.9	663.3
Statement	106.2	136.5	155.8	286.3	103.6	104.2	892.4
% change	27.8%	13.8%	35.1%	83.1%	19.5%	2.3%	34.5%
yinja (I01)							
	d_y	d_i	d_n	d_a			d_{yinja}
Queclarative	49.1	186.6	143.8	139.5			519
Statement	78.2	274.6	196.1	145.6			694.5
% change	59.3%	47.2%	36.4%	4.4%			33.8%
yintloko (I02)							
	d_y	d_i	d_n	d_{o1}	d_k	d_{o2}	$d_{yintloko}$
Queclarative	39.3	95	145.8	195.3	90	78.1	643.5
Statement	17.4	116.3	189.4	336.5	99.1	74.5	833.2
% change	-55.7%	22.4%	29.9%	72.3%	10.1%	-4.6%	29.5%
yinkomo (I03)							
	d_y	d_i	d_n	d_{o1}	d_m	d_{o2}	$d_{yinkomo}$
Queclarative	33.9	86.7	133.7	183.1	87.9	99.2	624.6
Statement	31.5	115	164.7	331.5	109.2	114.8	866.7
% change	-7.1%	32.6%	23.2%	81%	24.2%	15.7%	38.8%
zizinja (J01)							
	d_{z1}	d_{i1}	d_{z2}	d_{i2}	d_n	d_a	$d_{zizinja}$
Queclarative	90.6	122.1	90.9	142.4	117.8	131.9	695.6
Statement	123.3	156	142.2	231.8	172.8	132.4	958.4
% change	36.1%	27.8%	56.4%	62.8%	46.7%	0.4%	37.8%
ziintloko (J02)							
	d_z	d_i	d_n	d_{o1}	d_k	d_{o2}	$d_{ziintloko}$
Queclarative	100.9	171	150.8	196.3	81	76	776
Statement	132.2	217.8	217.2	335.1	102.7	78.9	1083.8
% change	31%	27.4%	44%	70.7%	26.8%	3.8%	39.7%

Table A.2 Mean phoneme and word duration [in ms] (continued).

ziinkomo (J03)								
	d_z	d_i	d_n	d_{o1}	d_m	d_{o2}		$d_{ziinkomo}$
Queclarative	106.4	165.5	134.7	181.6	82.2	104.4		774.8
Statement	135.5	194.8	180.1	322	119.5	108.5		1060.4
% change	27.3%	17.7%	33.7%	77.3%	45.4%	3.9%		36.9%
luluthi (K01)								
	d_{l1}	d_{u1}	d_{l2}	d_{u2}	d_t	d_i		$d_{luluthi}$
Queclarative	47.9	108	85.9	174	133.8	112.7		662.4
Statement	77.5	119.7	120.5	313.8	194.2	144		969.8
% change	61.8%	10.8%	40.3%	80.3%	45.1%	27.8%		46.4%
lulwimi (K02)								
	d_{l1}	d_u	d_{l2}	d_{i1}	d_m	d_{i2}		$d_{lulwimi}$
Queclarative	55.4	114.5	149.8	152.5	94.2	83.8		650.2
Statement	80.5	142.2	215.5	278	124.4	83.3		924
% change	45.3%	24.2%	43.9%	82.3%	32.1%	-0.6%		42.1%
lulwandle (K03)								
	d_{l1}	d_u	d_{l2}	d_a	d_n	d_e		$d_{lulwandle}$
Queclarative	58.6	108.9	141.2	177	168.5	79.3		733.5
Statement	57.3	137.6	181.3	309.3	199.9	73.8		959.2
% change	-2.2%	26.4%	28.4%	74.7%	18.6%	-6.9%		30.8%
bubuso (L01)								
	d_{b1}	d_{u1}	d_{b2}	d_{u2}	d_s	d_o		d_{bubuso}
Queclarative	34.2	107.6	60.4	174.4	129.4	119.9		625.9
Statement	46	123.4	82.9	309.2	185.1	115		861.6
% change	34.5%	14.7%	37.3%	77.3%	43%	-4.1%		37.7%
bubuhlanti (L02)								
	d_{b1}	d_{u1}	d_{b2}	d_{u2}	d_h	d_a	d_n	d_i	$d_{bubuhlanti}$
Queclarative	27.7	83.8	58.3	85.7	129.7	178.9	135.9	72.9	772.9
Statement	30.7	99.3	67.4	106.7	160.1	342.2	149.9	69.1	1025.5
% change	10.8%	18.5%	15.6%	24.5%	23.4%	91.3%	10.3%	-5.2%	32.7%
bubusika (L03)								
	d_{b1}	d_{u1}	d_{b2}	d_{u2}	d_s	d_i	d_k	d_a	$d_{bubusika}$
Queclarative	25.7	83.9	43	115.5	166.3	178.2	77	91	780.5
Statement	28.7	102.2	60.5	110.7	207.8	302.1	131.7	76.1	1019.7
% change	11.7%	21.8%	40.7%	-4.2%	25%	69.5%	71%	-16.4%	30.6%
kukutya (M01)								
	d_{k1}	d_{u1}	d_{k2}	d_{u2}	d_t	d_a			$d_{kukutya}$
Queclarative	29.2	117.2	78.6	170	91.8	127			613.9
Statement	35.8	134.2	118.4	298.7	146.9	136.7			870.7
% change	22.6%	14.5%	50.6%	75.7%	60%	7.6%			41.8%
kukwindla (M02)								
	d_{k1}	d_u	d_{k2}	d_i	d_n	d_a			$d_{kukwindla}$
Queclarative	25.3	112.7	129.3	136.6	147.7	136.2			687.8
Statement	30.5	152	160.4	235.5	198.9	142.2			919.4
% change	20.6%	34.9%	24.1%	72.4%	34.7%	4.4%			33.7%
kukunene (M03)								
	d_{k1}	d_{u1}	d_{k2}	d_{u2}	d_{n1}	d_{e1}	d_{n2}	d_{e2}	$d_{kukunene}$
Queclarative	26.3	97.2	70.9	75.7	102.5	192.9	97	87.5	750
Statement	28.6	109.5	99	76.1	143.7	322.9	121.7	75.4	976.9
% change	8.7%	12.7%	39.6%	0.5%	40.2%	67.4%	25.5%	-13.8%	30.3%

Table A.3 Mean word tempo [in syllables/s].

Word	Queclarative	Statement	Percentage change
ngumntu (A01)	6.22	4.38	-29.6 %
ngumfazi (A02)	5.84	4.62	-20.9 %
ngumfana (A03)	6.22	4.64	-25.4 %
ngabantu (B01)	4.92	3.8	-22.8 %
ngabafazi (B02)	5.13	4.19	-18.3 %
ngabafana (B03)	5.4	4.19	-22.4 %
ngumthi (C01)	4.98	3.76	-24.5 %
ngumlilo (C02)	6.65	4.88	-26.6 %
ngumfula (C03)	5.86	4.47	-23.7 %
yimithi (D01)	4.88	3.58	-26.6 %
yimililo (D02)	6.12	4.7	-23.2 %
yimifula (D03)	5.64	4.36	-22.7 %
lilitye (E01)	5.09	3.62	-28.9 %
liliso (E02)	4.84	3.67	-24.2 %
lilifu (E03)	4.78	3.61	-24.5 %
ngamatye (F01)	4.62	3.37	-27.1 %
ngamehlo (F02)	4.48	3.34	-25.4 %
ngamafu (F03)	4.51	3.54	-21.5 %
sisitya (G01)	4.44	3.24	-27.0 %
sisifo (G02)	4.24	3.12	-26.4 %
sisilo (G03)	4.44	3.29	-25.9 %
zizitya (H01)	4.26	3.21	-24.6 %
zizifo (H02)	4.44	3.19	-28.2 %
zizilo (H03)	4.52	3.36	-25.7 %
yinja (I01)	3.85	2.88	-25.2 %
yintloko (I02)	4.66	3.6	-22.7 %
yinkomo (I03)	4.8	3.46	-27.9 %
zizinja (J01)	4.31	3.13	-27.4 %
ziintloko (J02)	3.87	2.77	-28.4 %
ziinkomo (J03)	3.87	2.83	-26.9 %
luluthi (K01)	4.53	3.09	-31.8 %
lulwimi (K02)	4.61	3.25	-29.5 %
lulwandle (K03)	4.09	3.13	-23.5 %
bubuso (L01)	4.79	3.48	-27.3 %
bubuhlanti (L02)	5.18	3.9	-24.7 %
bubusika (L03)	5.12	3.92	-23.4 %
kukutya (M01)	4.89	3.45	-29.4 %
kukwindla (M02)	4.36	3.26	-25.2 %
kukunene (M03)	5.33	4.09	-23.3 %

Table A.4 Significance of syllables based on duration analyses.

Word	Sign test	Wilcoxon test	$\overline{\mu} / \sigma$
ngumntu	**ngu**$_1$ m ntu	**ngu**$_1$ **m**$_2$ ntu	**ngu**$_{1-}$ **m**$_{2-}$ ntu
ngumfazi	ngu m **fa**$_1$ zi	ngu **m**$_2$ **fa**$_1$ zi	ngu **m**$_{2-}$ **fa**$_{1-}$ zi
ngumfana	ngu m **fa**$_1$ na	ngu **m**$_2$ **fa**$_1$ na	ngu **m**$_{2-}$ **fa**$_{1-}$ na
ngabantu	nga **ba**$_1$ ntu	nga **ba**$_1$ ntu	nga **ba**$_{1-}$ ntu
ngabafazi	nga ba **fa**$_1$ zi	nga ba **fa**$_1$ zi	nga ba **fa**$_{1-}$ zi
ngabafana	nga ba **fa**$_1$ na	nga ba **fa**$_1$ na	nga ba **fa**$_{1-}$ na
ngumthi	ngu m thi	**ngu**$_3$ **m**$_2$ **thi**$_1$	**ngu**$_{3-}$ **m**$_{2-}$ **thi**$_{1-}$
ngumlilo	ngu **m**$_1$ **li**$_1$ lo	ngu **m**$_1$ **li**$_1$ **lo**$_2$	ngu **m**$_{3-}$ **li**$_{1-}$ **lo**$_{2-}$
ngumfula	ngu m **fu**$_1$ la	ngu **m**$_2$ **fu**$_1$ la	ngu **m**$_{2-}$ **fu**$_{1-}$ la
yimithi	yi **mi**$_1$ thi	yi **mi**$_1$ **thi**$_2$	yi **mi**$_{1-}$ **thi**$_{2-}$
yimililo	yi mi **li**$_1$ lo	yi **mi**$_2$ **li**$_1$ lo	yi **mi**$_{2-}$ **li**$_{1-}$ lo
yimifula	yi **mi**$_1$ **fu**$_1$ la	yi **mi**$_2$ **fu**$_1$ la	yi **mi**$_{2-}$ **fu**$_{1-}$ la
lilitye	li **li**$_1$ tye	li **li**$_1$ tye	li **li**$_{1-}$ tye
liliso	li **li**$_1$ so	li **li**$_1$ **so**$_2$	li **li**$_{1-}$ **so**$_{2-}$
lilifu	li **li**$_1$ fu	li **li**$_1$ fu	li **li**$_{1-}$ fu
ngamatye	nga **ma**$_1$ tye	nga **ma**$_1$ tye	nga **ma**$_{1-}$ tye
ngamehlo	nga **me**$_1$ hlo	nga **me**$_1$ hlo	nga **me**$_{1-}$ hlo
ngamafu	nga **ma**$_1$ fu	nga **ma**$_1$ fu	nga **ma**$_{1-}$ fu
sisitya	**si**$_1$ **si**$_1$ tya	**si**$_1$ **si**$_1$ **tya**$_2$	**si**$_{2-}$ **si**$_{1-}$ **tya**$_{3-}$
sisifo	si **si**$_1$ **fo**$_1$	si **si**$_1$ **fo**$_2$	si **si**$_{1-}$ **fo**$_{2-}$
sisilo	si **si**$_1$ lo	si **si**$_1$ lo	si **si**$_{1-}$ lo
zizitya	zi **zi**$_1$ tya	zi **zi**$_1$ **tya**$_2$	zi **zi**$_{1-}$ **tya**$_{2-}$
zizifo	zi **zi**$_1$ fo	**zi**$_2$ **zi**$_1$ **fo**$_3$	**zi**$_{3-}$ **zi**$_{1-}$ **fo**$_{2-}$
zizilo	zi **zi**$_1$ lo	zi **zi**$_1$ lo	zi **zi**$_{1-}$ lo
yinja	**yi**$_1$ nja	**yi**$_1$ **nja**$_2$	**yi**$_{1-}$ **nja**$_{2-}$
yintloko	yi **ntlo**$_1$ ko	yi **ntlo**$_1$ ko	yi **ntlo**$_{1-}$ ko
yinkomo	yi **nko**$_1$ mo	yi **nko**$_1$ **mo**$_2$	yi **nko**$_{1-}$ **mo**$_{2-}$
zizinja	zi zi **nja**$_1$	**zi**$_2$ **zi**$_1$ **nja**$_2$	**zi**$_{3-}$ **zi**$_{1-}$ **nja**$_{2-}$
ziintloko	zii **ntlo**$_1$ ko	**zii**$_2$ **ntlo**$_1$ ko	**zii**$_{2-}$ **ntlo**$_{1-}$ ko
ziinkomo	zii **nko**$_1$ mo	**zii**$_3$ **nko**$_1$ **mo**$_2$	**zii**$_{3-}$ **nko**$_{1-}$ **mo**$_{2-}$
luluthi	lu lu thi	lu **lu**$_2$ **thi**$_1$	lu **lu**$_{2-}$ **thi**$_{1-}$
lulwimi	lu **lwi**$_1$ mi	lu **lwi**$_1$ mi	lu **lwi**$_{1-}$ mi
lulwandle	lu **lwa**$_1$ ndle	lu **lwa**$_1$ ndle	lu **lwa**$_{1-}$ ndle
bubuso	bu **bu**$_1$ so	**bu**$_3$ **bu**$_1$ **so**$_2$	**bu**$_{3-}$ **bu**$_{1-}$ **so**$_{2-}$
bubuhlanti	bu **bu**$_1$ **hla**$_1$ nti	bu **bu**$_1$ **hla**$_1$ nti	bu **bu**$_{2-}$ **hla**$_{1-}$ nti
bubusika	bu bu **si**$_1$ **ka**$_1$	bu bu **si**$_1$ **ka**$_2$	bu bu **si**$_{1-}$ **ka**$_{2-}$
kukutya	ku **ku**$_1$ **tya**$_1$	ku **ku**$_1$ **tya**$_1$	ku **ku**$_{1-}$ **tya**$_{2-}$
kukwindla	**ku**$_1$ **kwi**$_1$ **ndla**$_1$	**ku**$_1$ **kwi**$_2$ **ndla**$_3$	**ku**$_{3-}$ **kwi**$_{1-}$ **ndla**$_{2-}$
kukunene	ku ku **ne**$_1$ ne	ku **ku**$_2$ **ne**$_1$ ne	ku **ku**$_{2-}$ **ne**$_{1-}$ ne

A.3 Pitch data

Table A.5 Median vowel and word pitch (includes the syllabic /m/ where appropriate).

ngumntu (A01)					
	p_{u1}	p_m	p_{u2}		$p_{ngumntu}$
Queclarative	155.9 Hz	234.2 Hz	204.2 Hz		208.1 Hz
Statement	133.9 Hz	169.7 Hz	137.7 Hz		159.5 Hz
% change	-14.1%	-27.5%	-32.6%		-23.4%
ngumfazi (A02)					
	p_u	p_m	p_a	p_i	$p_{ngumfazi}$
Queclarative	147.4 Hz	180.8 Hz	250 Hz	0	203.3 Hz
Statement	128 Hz	141.4 Hz	158.1 Hz	0	143.9 Hz
% change	-13.2%	-21.8%	-36.8%	0%	-29.2%
ngumfana (A03)					
	p_u	p_m	p_{a1}	p_{a2}	$p_{ngumfana}$
Queclarative	140.9 Hz	170.5 Hz	257.6 Hz	198.3 Hz	229.7 Hz
Statement	116.5 Hz	135 Hz	158.8 Hz	140.1 Hz	148.6 Hz
% change	-17.3%	-20.8%	-38.4%	-29.3%	-35.3%
ngabantu (B01)					
	p_{a1}	p_{a2}	p_u		$p_{ngabantu}$
Queclarative	140.8 Hz	240.3 Hz	187.8 Hz		208.5 Hz
Statement	116.2 Hz	153.5 Hz	125 Hz		146.2 Hz
% change	-17.5%	-36.1%	-33.4%		-29.9%
ngabafazi (B02)					
	p_{a1}	p_{a2}	p_{a3}	p_i	$p_{ngabafazi}$
Queclarative	151.7 Hz	228.8 Hz	238.3 Hz	85.7 Hz	199.9 Hz
Statement	131.8 Hz	163.6 Hz	164 Hz	0 Hz	151.4 Hz
% change	-13.1%	-28.5%	-31.2%	-100%	-24.3%
ngabafana (B03)					
	p_{a1}	p_{a2}	p_{a3}	p_{a4}	$p_{ngabafana}$
Queclarative	142.9 Hz	203.8 Hz	272.4 Hz	201.3 Hz	238.4 Hz
Statement	129.4 Hz	159.6 Hz	163.4 Hz	122.3 Hz	153.2 Hz
% change	-9.4%	-21.7%	-40%	-39.2%	-35.7%
ngumthi (C01)					
	p_u	p_m	p_i		$p_{ngumthi}$
Queclarative	232.6 Hz	230.9 Hz	273.3 Hz		242.6 Hz
Statement	167.9 Hz	172 Hz	146.7 Hz		171.7 Hz
% change	-27.8%	-25.5%	-46.3%		-29.2%
ngumlilo (C02)					
	p_u	p_m	p_i	p_o	$p_{ngumlilo}$
Queclarative	166.1 Hz	204.1 Hz	278.2 Hz	132.8 Hz	232.7 Hz
Statement	137.1 Hz	165.5 Hz	177 Hz	105.9 Hz	163.5 Hz
% change	-17.5%	-18.9%	-36.4%	-20.3%	-29.7%
ngumfula (C03)					
	p_{u1}	p_m	p_{u2}	p_a	$p_{ngumfula}$
Queclarative	157.3 Hz	204.6 Hz	275.6 Hz	125.1 Hz	222.9 Hz
Statement	133.9 Hz	161.7 Hz	176 Hz	102 Hz	162.8 Hz
% change	-14.9%	-21%	-36.1%	-18.5%	-27%

Table A.5 Median vowel and word pitch (continued).

yimithi (D01)						
	p_{i1}	p_m	p_{i2}	p_{i3}		$p_{yimithi}$
Queclarative	169.4 Hz	234.6 Hz	233.9 Hz	259.9 Hz		239.8 Hz
Statement	142.9 Hz	172.1 Hz	171.3 Hz	139.3 Hz		161.7 Hz
% change	-15.6%	-26.6%	-26.8%	-46.4%		-32.6%
yimililo (D02)						
	p_{i1}	p_m	p_{i2}	p_{i3}	p_o	$p_{yimililo}$
Queclarative	157.8 Hz	202.6 Hz	250.8 Hz	264.9 Hz	128 Hz	232.2 Hz
Statement	141.5 Hz	161.9 Hz	187.8 Hz	163.5 Hz	111.5 Hz	160.3 Hz
% change	-10.3%	-20.1%	-25.1%	-38.3%	-12.9%	-31%
yimifula (D03)						
	p_{i1}	p_m	p_{i2}	p_u	p_a	$p_{yimifula}$
Queclarative	155.2 Hz	195 Hz	236.8 Hz	256.9 Hz	125 Hz	216.5 Hz
Statement	136.1 Hz	156.5 Hz	192.5 Hz	165.4 Hz	108.2 Hz	166.6 Hz
% change	-12.3%	-19.7%	-18.7%	-35.6%	-13.4%	-23%
lilitye (E01)						
	p_{i1}	p_{i2}	p_e			$p_{lilitye}$
Queclarative	169.1 Hz	265.2 Hz	197.8 Hz			234.5 Hz
Statement	152.8 Hz	186 Hz	136.8 Hz			170.8 Hz
% change	-9.6%	-29.9%	-30.8%			-27.2%
liliso (E02)						
	p_{i1}	p_{i2}	p_o			p_{liliso}
Queclarative	222.9 Hz	219.1 Hz	199.5 Hz			241.5 Hz
Statement	181.8 Hz	161.6 Hz	132.1 Hz			165.7 Hz
% change	-18.4%	-26.2%	-33.8%			-31.4%
lilifu (E03)						
	p_{i1}	p_{i2}	p_u			p_{lilifu}
Queclarative	168.2 Hz	267.5 Hz	201.9 Hz			218.2 Hz
Statement	161.3 Hz	188.8 Hz	133.2 Hz			178.8 Hz
% change	-4.1%	-29.4%	-34%			-18.1%
ngamatye (F01)						
	p_{a1}	p_m	p_{a2}	p_e		$p_{ngamatye}$
Queclarative	143.5 Hz	163.6 Hz	235.5 Hz	187.8 Hz		208.1 Hz
Statement	134.1 Hz	142.2 Hz	176.2 Hz	143.4 Hz		153.4 Hz
% change	-6.6%	-13.1%	-25.2%	-23.6%		-26.3%
ngamehlo (F02)						
	p_a	p_m	p_e	p_o		$p_{ngamehlo}$
Queclarative	163.1 Hz	219.9 Hz	238.5 Hz	137.6 Hz		200.1 Hz
Statement	138.6 Hz	164.9 Hz	176.3 Hz	115.5 Hz		164.2 Hz
% change	-15%	-25%	-26.1%	-16.1%		-17.9%
ngamafu (F03)						
	p_{a1}	p_m	p_{a2}	p_u		$p_{ngamafu}$
Queclarative	144.1 Hz	176.2 Hz	265.9 Hz	201.8 Hz		209.5 Hz
Statement	125.8 Hz	147.8 Hz	186 Hz	140.5 Hz		162.1 Hz
% change	-12.7%	-16.1%	-30%	-30.4%		-22.6%
sisitya (G01)						
	p_{i1}	p_{i2}	p_a			$p_{sisitya}$
Queclarative	188.6 Hz	273 Hz	189.8 Hz			256.5 Hz
Statement	161.2 Hz	197.6 Hz	148.1 Hz			180.4 Hz
% change	-14.5%	-27.6%	-22%			-29.7%

Table A.5 Median vowel and word pitch (continued).

sisifo (G02)					
	p_{i1}	p_{i2}	p_{o}		p_{sisifo}
Queclarative	268.3 Hz	217.7 Hz	240.4 Hz		243.2 Hz
Statement	218.4 Hz	178.2 Hz	141 Hz		184.6 Hz
% change	-18.6%	-18.1%	-41.3%		-24.1%
sisilo (G03)					
	p_{i1}	p_{i2}	p_{o}		p_{sisilo}
Queclarative	196.8 Hz	271.6 Hz	196.1 Hz		267.9 Hz
Statement	162.7 Hz	201.1 Hz	148.1 Hz		183.7 Hz
% change	-17.3%	-26%	-24.5%		-31.4%
zizitya (H01)					
	p_{i1}	p_{i2}	p_{a}		$p_{zizitya}$
Queclarative	164.6 Hz	226.8 Hz	236.3 Hz		197.4 Hz
Statement	134.3 Hz	173.5 Hz	159.9 Hz		153 Hz
% change	-18.4%	-23.5%	-32.3%		-22.5%
zizifo (H02)					
	p_{i1}	p_{i2}	p_{o}		p_{zizifo}
Queclarative	249.6 Hz	178.6 Hz	282.1 Hz		239.8 Hz
Statement	208.2 Hz	132.8 Hz	140.6 Hz		160.6 Hz
% change	-16.6%	-25.6%	-50.2%		-33%
zizilo (H03)					
	p_{i1}	p_{i2}	p_{o}		p_{zizilo}
Queclarative	162.1 Hz	195.3 Hz	245.5 Hz		205.7 Hz
Statement	140.5 Hz	169.5 Hz	153.3 Hz		161.1 Hz
% change	-13.3%	-13.2%	-37.6%		-21.7%
yinja (I01)					
	p_{i}	p_{a}			p_{yinja}
Queclarative	254.6 Hz	203.8 Hz			222.1 Hz
Statement	183.8 Hz	121.1 Hz			160.2 Hz
% change	-27.8%	-40.6%			-27.9%
yintloko (I02)					
	p_{i}	p_{o1}	p_{o2}		$p_{yintloko}$
Queclarative	157.7 Hz	261.7 Hz	183.1 Hz		213 Hz
Statement	125.9 Hz	171.7 Hz	131.4 Hz		160.6 Hz
% change	-20.2%	-34.4%	-28.2%		-24.6%
yinkomo (I03)					
	p_{i}	p_{o1}	p_{m}	p_{o2}	$p_{yinkomo}$
Queclarative	157.7 Hz	225.1 Hz	195.4 Hz	254.6 Hz	223 Hz
Statement	145.9 Hz	172 Hz	119.5 Hz	132.8 Hz	151 Hz
% change	-7.5%	-23.6%	-38.8%	-47.8%	-32.3%
zizinja (J01)					
	p_{i1}	p_{i2}	p_{a}		$p_{zizinja}$
Queclarative	264.7 Hz	185.8 Hz	204.4 Hz		217.3 Hz
Statement	192.1 Hz	116.5 Hz	117.1 Hz		135.7 Hz
% change	-27.4%	-37.3%	-42.7%		-37.6%
ziintloko (J02)					
	p_{i}	p_{o1}	p_{o2}		$p_{ziintloko}$
Queclarative	243.3 Hz	266.9 Hz	200.9 Hz		236.6 Hz
Statement	216.5 Hz	149.7 Hz	129.7 Hz		178.1 Hz
% change	-11%	-43.9%	-35.4%		-24.7%

Table A.5 Median vowel and word pitch (continued).

ziinkomo (J03)					
	p_i	p_{o1}	p_m	p_{o2}	$p_{ziinkomo}$
Queclarative	254.6 Hz	179.9 Hz	189.5 Hz	269.6 Hz	223.7 Hz
Statement	211.7 Hz	129.9 Hz	119.3 Hz	134.1 Hz	167.9 Hz
% change	-16.8%	-27.8%	-37%	-50.3%	-24.9%
luluthi (K01)					
	p_{u1}	p_{u2}	p_i		$p_{luluthi}$
Queclarative	239	207.6	275		231.7
Statement	201.1	176.7	140.2		188.4
% change	-15.9%	-14.9%	-49%		-18.7%
lulwimi (K02)					
	p_u	p_{i1}	p_m	p_{i2}	$p_{lulwimi}$
Queclarative	273.2	276.2	176.8	141.8	237.2
Statement	237.9	161.6	113.1	111.1	182
% change	-12.9%	-41.5%	-36%	-21.7%	-23.3%
lulwandle (K03)					
	p_u	p_a	p_e		$p_{lulwandle}$
Queclarative	277	275.4	132		237.9
Statement	207.3	162.4	94.6		168.5
% change	-25.2%	-41%	-28.3%		-29.2%
bubuso (L01)					
	p_{u1}	p_{u2}	p_o		p_{bubuso}
Queclarative	233	224.5	245.5		237.4
Statement	196.1	182	135.8		177.8
% change	-15.8%	-18.9%	-44.7%		-25.1%
bubuhlanti (L02)					
	p_{u1}	p_{u2}	p_a	p_i	$p_{bubuhlanti}$
Queclarative	201.9	247.2	250.6	183.4	238.7
Statement	161.9	186.4	177.6	139.3	174.9
% change	-19.8%	-24.6%	-29.1%	-24%	-26.7%
bubusika (L03)					
	p_{u1}	p_{u2}	p_i	p_a	$p_{bubusika}$
Queclarative	237.8	231.5	229.5	137.7	232.7
Statement	209.7	185.7	165.1	111.7	178.4
% change	-11.8%	-19.8%	-28.1%	-18.9%	-23.3%
kukutya (M01)					
	p_{u1}	p_{u2}	p_a		$p_{kukutya}$
Queclarative	260.9	215	260		252.9
Statement	201	161.3	139.6		174.4
% change	-23%	-25%	-46.3%		-31%
kukwindla (M02)					
	p_u	p_i	p_a		$p_{kukwindla}$
Queclarative	260.8	211.9	222		242
Statement	206	158.1	121.2		177.8
% change	-21%	-25.4%	-45.4%		-26.5%
kukunene (M03)					
	p_{u1}	p_{u2}	p_{e1}	p_{e2}	$p_{kukunene}$
Queclarative	262.4	262.9	258.2	130.7	245.6
Statement	215	204.9	153.6	99	175
% change	-18.1%	-22.1%	-40.5%	-24.3%	-28.7%

Table A.6 Vowel pitch slope patterns (includes the syllabic /m/ where appropriate).

Word	Queclarative	Statement
ngumntu (A01)	+−	+−
ngumfazi (A02)	++−	++−
ngumfana (A03)	++−	++−
ngabantu (B01)	+−	+−
ngabafazi (B02)	++−	++−
ngabafana (B03)	++−	++−
ngumthi (C01)	−+	+−
ngumlilo (C02)	++−	++−
ngumfula (C03)	++−	++−
yimithi (D01)	+−+	+−−
yimililo (D02)	+++−	++−−
yimifula (D03)	+++−	++−−
lilitye (E01)	+−	+−
liliso (E02)	−−	−−
lilifu (E03)	+−	+−
ngamatye (F01)	++−	++−
ngamehlo (F02)	++−	++−
ngamafu (F03)	++−	++−
sisitya (G01)	+−	+−
sisifo (G02)	−+	−−
sisilo (G03)	+−	+−
zizitya (H01)	++	+−
zizifo (H02)	−+	−+
zizilo (H03)	++	+−
yinja (I01)	−	−
yintloko (I02)	+−	+−
yinkomo (I03)	+−+	+−+
zizinja (J01)	−+	−+
ziintloko (J02)	+−	−−
ziinkomo (J03)	−++	−−+
luluthi (K01)	−+	−−
lulwimi (K02)	+−−	−−−
lulwandle (K03)	−−	−−
bubuso (L01)	−+	−−
bubuhlanti (L02)	++−	+−−
bubusika (L03)	−−−	−−−
kukutya (M01)	−+	−−
kukwindla (M02)	−+	−−
kukunene (M03)	+−−	−−−

Table A.7 Significance of vowels based on pitch analyses (includes the syllabic /m/ where appropriate).

Word	Sign test	Wilcoxon test	$\bar{\mu} / \sigma$
ngumntu	ng**u**$_1$ **m**$_1$ nt**u**$_1$	ng**u**$_1$ **m**$_1$ nt**u**$_2$	ng**u**$_{3+}$ **m**$_{2+}$ nt**u**$_{1+}$
ngumfazi	ngu **m**$_1$ f**a**$_1$ zi	ngu **m**$_1$ f**a**$_1$ zi	ngu **m**$_{2+}$ f**a**$_{1+}$ zi
ngumfana	ng**u**$_1$ **m**$_1$ f**a**$_1$ na	ng**u**$_1$ **m**$_1$ f**a**$_1$ n**a**$_2$	ng**u**$_{3+}$ **m**$_{2+}$ f**a**$_{1+}$ n**a**$_{4+}$
ngabantu	ng**a**$_1$ b**a**$_1$ ntu	ng**a**$_1$ b**a**$_1$ ntu	ng**a**$_{2+}$ b**a**$_{1+}$ ntu
ngabafazi	ng**a**$_1$ b**a**$_1$ f**a**$_1$ zi	ng**a**$_2$ b**a**$_1$ f**a**$_1$ z**i**$_3$	ng**a**$_{3+}$ b**a**$_{2+}$ f**a**$_{1+}$ z**i**$_{4+}$
ngabafana	nga b**a**$_1$ f**a**$_1$ n**a**$_1$	ng**a**$_2$ b**a**$_1$ f**a**$_1$ n**a**$_3$	ng**a**$_{3+}$ b**a**$_{2+}$ f**a**$_{1+}$ n**a**$_{4+}$
ngumthi	ng**u**$_1$ **m**$_1$ th**i**$_1$	ng**u**$_1$ **m**$_1$ th**i**$_1$	ng**u**$_{3+}$ **m**$_{2+}$ th**i**$_{1+}$
ngumlilo	ng**u**$_1$ **m**$_1$ l**i**$_1$ l**o**$_1$	ng**u**$_1$ **m**$_1$ l**i**$_1$ l**o**$_1$	ng**u**$_{3+}$ **m**$_{2+}$ l**i**$_{1+}$ l**o**$_{4+}$
ngumfula	ng**u**$_1$ **m**$_1$ f**u**$_1$ l**a**$_1$	ng**u**$_1$ **m**$_2$ f**u**$_1$ l**a**$_3$	ng**u**$_{3+}$ **m**$_{2+}$ f**u**$_{1+}$ l**a**$_{4+}$
yimithi	y**i**$_1$ m**i**$_1$ th**i**$_1$	y**i**$_2$ m**i**$_1$ th**i**$_1$	y**i**$_{3+}$ m**i**$_{2+}$ th**i**$_{1+}$
yimililo	yi m**i**$_1$ l**i**$_1$ l**o**$_1$	y**i**$_2$ m**i**$_1$ l**i**$_1$ l**o**$_1$	y**i**$_{4+}$ m**i**$_{2+}$ l**i**$_{1+}$ l**o**$_{3+}$
yimifula	yi m**i**$_1$ f**u**$_1$ l**a**$_1$	y**i**$_3$ m**i**$_1$ f**u**$_1$ l**a**$_2$	y**i**$_{4+}$ m**i**$_{2+}$ f**u**$_{1+}$ l**a**$_{3+}$
lilitye	l**i**$_1$ l**i**$_1$ ty**e**$_1$	l**i**$_1$ l**i**$_1$ ty**e**$_1$	l**i**$_{3+}$ l**i**$_{2+}$ ty**e**$_{1+}$
liliso	l**i**$_1$ l**i**$_1$ s**o**$_1$	l**i**$_1$ l**i**$_1$ s**o**$_2$	l**i**$_{2+}$ l**i**$_{1+}$ s**o**$_{3+}$
lilifu	l**i**$_1$ l**i**$_1$ fu	l**i**$_1$ l**i**$_1$ f**u**$_2$	l**i**$_{2+}$ l**i**$_{1+}$ f**u**$_{3+}$
ngamatye	ng**a**$_1$ m**a**$_1$ ty**e**$_1$	ng**a**$_1$ m**a**$_1$ ty**e**$_1$	ng**a**$_{3+}$ m**a**$_{1+}$ ty**e**$_{2+}$
ngamehlo	ng**a**$_1$ m**e**$_1$ hl**o**$_1$	ng**a**$_2$ m**e**$_1$ hl**o**$_3$	ng**a**$_{2+}$ m**e**$_{1+}$ hl**o**$_{3+}$
ngamafu	ng**a**$_1$ m**a**$_1$ fu	ng**a**$_2$ m**a**$_1$ f**u**$_3$	ng**a**$_{2+}$ m**a**$_{1+}$ f**u**$_{3+}$
sisitya	s**i**$_1$ s**i**$_1$ ty**a**$_1$	s**i**$_2$ s**i**$_1$ ty**a**$_2$	s**i**$_{3+}$ s**i**$_{1+}$ ty**a**$_{2+}$
sisifo	si s**i**$_1$ f**o**$_1$	s**i**$_3$ s**i**$_1$ f**o**$_2$	s**i**$_{3+}$ s**i**$_{1+}$ f**o**$_{2+}$
sisilo	s**i**$_1$ s**i**$_1$ l**o**$_1$	s**i**$_2$ s**i**$_1$ l**o**$_1$	s**i**$_{3+}$ s**i**$_{1+}$ l**o**$_{2+}$
zizitya	z**i**$_1$ z**i**$_1$ ty**a**$_1$	z**i**$_2$ z**i**$_1$ ty**a**$_1$	z**i**$_{3+}$ z**i**$_{1+}$ ty**a**$_{2+}$
zizifo	z**i**$_1$ z**i**$_1$ f**o**$_1$	z**i**$_2$ z**i**$_1$ f**o**$_1$	z**i**$_{3+}$ z**i**$_{2+}$ f**o**$_{1+}$
zizilo	z**i**$_1$ z**i**$_1$ l**o**$_1$	z**i**$_2$ z**i**$_1$ l**o**$_1$	z**i**$_{3+}$ z**i**$_{2+}$ l**o**$_{1+}$
yinja	y**i**$_1$ nj**a**$_1$	y**i**$_1$ nj**a**$_1$	y**i**$_{2+}$ nj**a**$_{1+}$
yintloko	y**i**$_1$ ntl**o**$_1$ k**o**$_1$	y**i**$_2$ ntl**o**$_1$ k**o**$_3$	y**i**$_{2+}$ ntl**o**$_{1+}$ k**o**$_{3+}$
yinkomo	yi nk**o**$_1$ m**o**$_1$	y**i**$_2$ nk**o**$_1$ m**o**$_1$	y**i**$_{3+}$ nk**o**$_{2+}$ m**o**$_{1+}$
zizinja	z**i**$_1$ z**i**$_1$ nj**a**$_1$	z**i**$_2$ z**i**$_1$ nj**a**$_1$	z**i**$_{3+}$ z**i**$_{2+}$ nj**a**$_{1+}$
ziintloko	z**ii**$_1$ ntl**o**$_1$ k**o**$_1$	z**ii**$_1$ ntl**o**$_1$ k**o**$_2$	z**ii**$_{2+}$ ntl**o**$_{1+}$ k**o**$_{3+}$
ziinkomo	z**ii**$_1$ nk**o**$_1$ m**o**$_1$	z**ii**$_1$ nk**o**$_2$ m**o**$_1$	z**ii**$_{2+}$ nk**o**$_{3+}$ m**o**$_{1+}$
luluthi	l**u**$_1$ l**u**$_1$ th**i**$_1$	l**u**$_2$ l**u**$_1$ th**i**$_1$	l**u**$_{3+}$ l**u**$_{2+}$ th**i**$_{1+}$
lulwimi	l**u**$_1$ lw**i**$_1$ m**i**$_1$	l**u**$_1$ lw**i**$_1$ m**i**$_1$	l**u**$_{2+}$ lw**i**$_{1+}$ m**i**$_{3+}$
lulwandle	l**u**$_1$ lw**a**$_1$ ndl**e**$_1$	l**u**$_1$ lw**a**$_1$ ndl**e**$_2$	l**u**$_{2+}$ lw**a**$_{1+}$ ndl**e**$_{3+}$
bubuso	b**u**$_1$ b**u**$_1$ s**o**$_1$	b**u**$_1$ b**u**$_1$ s**o**$_1$	b**u**$_{3+}$ b**u**$_{2+}$ s**o**$_{1+}$
bubuhlanti	b**u**$_1$ b**u**$_1$ hl**a**$_1$ nti	b**u**$_1$ b**u**$_1$ hl**a**$_1$ nti	b**u**$_{3+}$ b**u**$_{2+}$ hl**a**$_{1+}$ nti
bubusika	bu b**u**$_1$ s**i**$_1$ k**a**$_1$	b**u**$_2$ b**u**$_1$ s**i**$_1$ k**a**$_1$	b**u**$_{4+}$ b**u**$_{3+}$ s**i**$_{1+}$ k**a**$_{2+}$
kukutya	k**u**$_1$ k**u**$_1$ ty**a**$_1$	k**u**$_1$ k**u**$_1$ ty**a**$_1$	k**u**$_{3+}$ k**u**$_{2+}$ ty**a**$_{1+}$
kukwindla	k**u**$_1$ kw**i**$_1$ ndl**a**$_1$	k**u**$_1$ kw**i**$_1$ ndl**a**$_2$	k**u**$_{3+}$ kw**i**$_{1+}$ ndl**a**$_{2+}$
kukunene	k**u**$_1$ k**u**$_1$ n**e**$_1$ n**e**$_1$	k**u**$_1$ k**u**$_1$ n**e**$_1$ n**e**$_1$	k**u**$_{3+}$ k**u**$_{2+}$ n**e**$_{1+}$ n**e**$_{4+}$

A.4 Loudness data

Table A.8 Significance of phonemes based on relative loudness analyses.

Word	Sign test	Wilcoxon test	$\bar{\mu} / \sigma$
ngumntu	**ng**$_1$u m ntu	**ng**$_1$u m ntu	**ngu**$_{1-}$ m ntu
ngumfazi	ng**u**$_1$ m fa zi	ng**u**$_1$ **m**$_2$ fa **z**$_3$**i**$_4$	ng**u**$_{4+}$ **m**$_{1+}$ fa **z**$_{2+}$**i**$_{3+}$
ngumfana	ngu m fa na	**ng**$_2$u m **f**$_1$a na	**ng**$_{2+}$u m **f**$_{1-}$a na
ngabantu	nga ba ntu	nga ba nt**u**$_1$	nga ba nt**u**$_{1+}$
ngabafazi	nga ba fa zi	nga ba **f**$_1$a **z**$_2$**i**$_3$	nga ba **f**$_{1+}$a **z**$_{3+}$**i**$_{2+}$
ngabafana	nga ba fa na	**ng**$_1$**a**$_2$ ba fa na	**ng**$_{2-}$**a**$_{1-}$ ba fa na
ngumthi	ngu m th**i**$_1$	ngu m th**i**$_1$	ngu m th**i**$_{1+}$
ngumlilo	ngu m li **l**$_1$o	ngu m li **l**$_1$**o**$_2$	ngu m li **l**$_{1+}$**o**$_{2+}$
ngumfula	ngu m fu **l**$_1$**a**$_1$	ngu m fu **l**$_2$**a**$_1$	ngu m fu **l**$_{1+}$**a**$_{2+}$
yimithi	yi mi **th**$_1$i	yi mi **th**$_2$**i**$_1$	yi mi **th**$_{2+}$**i**$_{1+}$
yimililo	yi mi li lo	yi mi **li**$_2$ **l**$_1$**o**$_3$	yi mi **li**$_{2+}$ **l**$_{1+}$**o**$_{3+}$
yimifula	yi mi fu **l**$_1$a	yi mi fu **l**$_1$**a**$_2$	yi mi fu **l**$_{1+}$**a**$_{2+}$
lilitye	li li tye	li li tye	li li tye
liliso	li li so	li **l**$_3$i **s**$_1$**o**$_2$	li **l**$_{2+}$i **s**$_{3+}$**o**$_{1+}$
lilifu	li li fu	li li fu	li li fu
ngamatye	nga ma tye	nga **m**$_1$a tye	nga **m**$_{1+}$a tye
ngamehlo	nga me hlo	nga me hlo	nga me hlo
ngamafu	nga **m**$_1$a fu	ng**a**$_2$ **m**$_1$a f**u**$_2$	ng**a**$_{3+}$ **m**$_{1+}$a f**u**$_{2+}$
sisitya	si si tya	si si tya	si si tya
sisifo	si si fo	si si fo	si si fo
sisilo	si si lo	s**i**$_1$ si lo	s**i**$_{1-}$ si lo
zizitya	zi zi tya	**z**$_1$i **z**$_2$i tya	**z**$_{2+}$i **z**$_{1+}$i tya
zizifo	zi z**i**$_1$ fo	zi z**i**$_1$ fo	zi z**i**$_{1+}$ fo
zizilo	zi zi lo	zi zi lo	zi zi lo
yinja	yi nja	yi nj**a**$_1$	yi nj**a**$_{1+}$
yintloko	yi ntlo ko	yi ntlo ko	yi ntlo ko
yinkomo	yi nko **m**$_1$o	yi nko **m**$_1$o	yi nko **m**$_{1+}$o
zizinja	zi zi **nj**$_1$**a**$_1$	**z**$_4$i z**i**$_3$ **nj**$_1$**a**$_2$	**z**$_{4+}$i z**i**$_{3+}$ **nj**$_{1+}$**a**$_{2+}$
ziintloko	zii ntlo ko	zii ntlo ko	zii ntlo ko
ziinkomo	zii nko **m**$_1$**o**$_1$	zii nk**o**$_3$ **m**$_1$**o**$_2$	zii nk**o**$_{1+}$ **m**$_{2+}$**o**$_{3+}$
luluthi	lu lu thi	**l**$_3$u l**u**$_2$ th**i**$_1$	**l**$_{3-}$u l**u**$_{2+}$ th**i**$_{1+}$
lulwimi	lu lwi **m**$_1$**i**$_1$	l**u**$_3$ lw**i**$_3$ **m**$_1$**i**$_2$	l**u**$_{2-}$ lw**i**$_{4+}$ **m**$_{1+}$**i**$_{3+}$
lulwandle	lu lwa ndle	lu lwa **ndl**$_1$**e**$_2$	lu lwa **ndl**$_{1+}$**e**$_{2+}$
bubuso	b**u**$_1$ bu so	b**u**$_1$ bu so	b**u**$_{1-}$ bu so
bubuhlanti	bu bu hla nti	bu bu hla nti	bu bu hla nti
bubusika	bu bu si **k**$_1$**a**$_1$	**b**$_5$u **b**$_4$u s**i**$_3$ **k**$_2$**a**$_1$	**b**$_{4-}$u **b**$_{5-}$u s**i**$_{3+}$ **k**$_{2+}$**a**$_{1+}$
kukutya	ku ku **ty**$_1$**a**$_1$	ku ku **ty**$_1$**a**$_2$	ku ku **ty**$_{2+}$**a**$_{1+}$
kukwindla	ku kwi **ndl**$_1$**a**$_1$	ku kw**i**$_3$ **ndl**$_1$**a**$_2$	ku kw**i**$_{3+}$ **ndl**$_{2+}$**a**$_{1+}$
kukunene	ku ku ne **n**$_1$**e**$_1$	ku ku ne **n**$_1$**e**$_2$	ku ku ne **n**$_{1+}$**e**$_{2+}$

Table A.9 Mean absolute loudness of words.

Word	Queclarative	Statement	Percentage change
ngumntu (A01)	0.031233	0.020553	-34.2 %
ngumfazi (A02)	0.014601	0.008445	-42.2 %
ngumfana (A03)	0.024512	0.011362	-53.6 %
ngabantu (B01)	0.018788	0.012163	-35.3 %
ngabafazi (B02)	0.015531	0.009627	-38.0 %
ngabafana (B03)	0.020591	0.009522	-53.8 %
ngumthi (C01)	0.033395	0.024689	-26.1 %
ngumlilo (C02)	0.035449	0.016579	-53.2 %
ngumfula (C03)	0.024789	0.010798	-56.4 %
yimithi (D01)	0.042171	0.019826	-53.0 %
yimililo (D02)	0.035132	0.021606	-38.5 %
yimifula (D03)	0.028882	0.015302	-47.0 %
lilitye (E01)	0.028486	0.015323	-46.2 %
liliso (E02)	0.026216	0.012743	-51.4 %
lilifu (E03)	0.031289	0.014808	-52.7 %
ngamatye (F01)	0.017498	0.01228	-29.8 %
ngamehlo (F02)	0.018722	0.01295	-30.8 %
ngamafu (F03)	0.019769	0.011981	-39.4 %
sisitya (G01)	0.017383	0.0101	-41.9 %
sisifo (G02)	0.022016	0.01307	-40.6 %
sisilo (G03)	0.022539	0.020414	-9.43 %
zizitya (H01)	0.017083	0.010426	-39.0 %
zizifo (H02)	0.023401	0.010351	-55.8 %
zizilo (H03)	0.019013	0.014631	-23.0 %
yinja (I01)	0.034006	0.014797	-56.5 %
yintloko (I02)	0.01816	0.009571	-47.3 %
yinkomo (I03)	0.018872	0.013708	-27.4 %
zizinja (J01)	0.017869	0.012332	-31.0 %
ziintloko (J02)	0.019488	0.012068	-38.1 %
ziinkomo (J03)	0.028998	0.016363	-43.6 %
luluthi (K01)	0.043112	0.028896	-33.0 %
lulwimi (K02)	0.036264	0.019873	-45.2 %
lulwandle (K03)	0.023532	0.014883	-36.8 %
bubuso (L01)	0.022053	0.01325	-39.9 %
bubuhlanti (L02)	0.02028	0.013925	-31.3 %
bubusika (L03)	0.013811	0.010293	-25.5 %
kukutya (M01)	0.018709	0.011063	-40.9 %
kukwindla (M02)	0.021743	0.011652	-46.4 %
kukunene (M03)	0.023339	0.013319	-42.9 %

Table A.10 Significance of phonemes based on absolute loudness analyses.

Word	Sign test	Wilcoxon test	$\bar{\mu} / \sigma$
ngumntu	ngu m ntu	ngu $\mathbf{m}_{2}$ $\mathbf{nt}_{1}$u	ngu $\mathbf{m}_{1+}$ $\mathbf{nt}_{2+}$u
ngumfazi	ngu $\mathbf{m}_{1}$ fa zi	ng$\mathbf{u}_{2}$ $\mathbf{m}_{1}$ fa $\mathbf{z}_{2}\mathbf{i}_{3}$	ng$\mathbf{u}_{2+}$ $\mathbf{m}_{1+}$ fa $\mathbf{z}_{4+}\mathbf{i}_{3+}$
ngumfana	ng$\mathbf{u}_{1}$ $\mathbf{m}_{1}$ f$\mathbf{a}_{1}$ $\mathbf{n}_{1}$a	ng$\mathbf{u}_{1}$ $\mathbf{m}_{2}$ f$\mathbf{a}_{3}$ $\mathbf{n}_{3}\mathbf{a}_{4}$	ng$\mathbf{u}_{2+}$ $\mathbf{m}_{1+}$ f$\mathbf{a}_{4+}$ $\mathbf{n}_{3+}\mathbf{a}_{5+}$
ngabantu	ng$\mathbf{a}_{1}$ ba ntu	ng$\mathbf{a}_{2}$ ba nt$\mathbf{u}_{1}$	ng$\mathbf{a}_{2+}$ ba nt$\mathbf{u}_{1+}$
ngabafazi	nga ba $\mathbf{f}_{1}$a zi	nga $\mathbf{b}_{3}$a $\mathbf{f}_{1}\mathbf{a}_{2}$ $\mathbf{z}_{4}\mathbf{i}_{5}$	nga $\mathbf{b}_{3+}$a $\mathbf{f}_{1+}\mathbf{a}_{2+}$ $\mathbf{z}_{5+}\mathbf{i}_{4+}$
ngabafana	nga ba f$\mathbf{a}_{1}$ na	nga $\mathbf{b}_{3}$a $\mathbf{f}_{2}\mathbf{a}_{1}$ $\mathbf{n}_{2}$a	nga $\mathbf{b}_{4+}$a $\mathbf{f}_{3+}\mathbf{a}_{1+}$ $\mathbf{n}_{2+}$a
ngumthi	ngu m th$\mathbf{i}_{1}$	ngu m th$\mathbf{i}_{1}$	ngu m th$\mathbf{i}_{1+}$
ngumlilo	ng$\mathbf{u}_{1}$ $\mathbf{m}_{1}$ l$\mathbf{i}_{1}$ $\mathbf{l}_{1}\mathbf{o}_{1}$	ng$\mathbf{u}_{1}$ $\mathbf{m}_{2}$ $\mathbf{l}_{4}\mathbf{i}_{3}$ $\mathbf{l}_{1}\mathbf{o}_{1}$	ng$\mathbf{u}_{1+}$ $\mathbf{m}_{2+}$ $\mathbf{l}_{6+}\mathbf{i}_{4+}$ $\mathbf{l}_{5+}\mathbf{o}_{3+}$
ngumfula	ngu $\mathbf{m}_{1}$ fu $\mathbf{l}_{1}\mathbf{a}_{1}$	ng$\mathbf{u}_{2}$ $\mathbf{m}_{3}$ f$\mathbf{u}_{4}$ $\mathbf{l}_{1}\mathbf{a}_{1}$	ng$\mathbf{u}_{1+}$ $\mathbf{m}_{4+}$ f$\mathbf{u}_{5+}$ $\mathbf{l}_{3+}\mathbf{a}_{2+}$
yimithi	yi mi $\mathbf{th}_{1}\mathbf{i}_{1}$	y$\mathbf{i}_{3}$ $\mathbf{m}_{3}\mathbf{i}_{3}$ $\mathbf{th}_{1}\mathbf{i}_{2}$	y$\mathbf{i}_{3+}$ $\mathbf{m}_{4+}\mathbf{i}_{5+}$ $\mathbf{th}_{2+}\mathbf{i}_{1+}$
yimililo	yi m$\mathbf{i}_{1}$ li lo	yi m$\mathbf{i}_{1}$ l$\mathbf{i}_{2}$ $\mathbf{l}_{3}\mathbf{o}_{2}$	yi m$\mathbf{i}_{2+}$ l$\mathbf{i}_{3+}$ $\mathbf{l}_{4+}\mathbf{o}_{1+}$
yimifula	yi mi f$\mathbf{u}_{1}$ $\mathbf{l}_{1}\mathbf{a}_{1}$	y$\mathbf{i}_{5}$ $\mathbf{m}_{2}\mathbf{i}_{4}$ $\mathbf{f}_{3}\mathbf{u}_{1}$ $\mathbf{l}_{1}\mathbf{a}_{2}$	y$\mathbf{i}_{5+}$ $\mathbf{m}_{4+}\mathbf{i}_{6+}$ $\mathbf{f}_{7+}\mathbf{u}_{2+}$ $\mathbf{l}_{1+}\mathbf{a}_{3+}$
lilitye	li $\mathbf{l}_{1}\mathbf{i}_{1}$ tye	li $\mathbf{l}_{2}\mathbf{i}_{1}$ $\mathbf{ty}_{3}\mathbf{e}_{4}$	li $\mathbf{l}_{4+}\mathbf{i}_{1+}$ $\mathbf{ty}_{2+}\mathbf{e}_{3+}$
liliso	li $\mathbf{l}_{1}\mathbf{i}_{1}$ $\mathbf{s}_{1}\mathbf{o}_{1}$	li $\mathbf{l}_{1}\mathbf{i}_{1}$ $\mathbf{s}_{2}\mathbf{o}_{3}$	li $\mathbf{l}_{3+}\mathbf{i}_{1+}$ $\mathbf{s}_{4+}\mathbf{o}_{2+}$
lilifu	li $\mathbf{l}_{1}$i fu	l$\mathbf{i}_{4}$ $\mathbf{l}_{1}\mathbf{i}_{2}$ f$\mathbf{u}_{3}$	l$\mathbf{i}_{3+}$ $\mathbf{l}_{4+}\mathbf{i}_{2+}$ f$\mathbf{u}_{1+}$
ngamatye	nga ma tye	nga ma tye	nga ma tye
ngamehlo	nga me hlo	nga me hl$\mathbf{o}_{1}$	nga me hl$\mathbf{o}_{1+}$
ngamafu	ng$\mathbf{a}_{1}$ ma fu	$\mathbf{ng}_{4}\mathbf{a}_{2}$ $\mathbf{m}_{3}$a f$\mathbf{u}_{1}$	$\mathbf{ng}_{4+}\mathbf{a}_{2+}$ $\mathbf{m}_{3+}$a f$\mathbf{u}_{1+}$
sisitya	si si tya	$\mathbf{s}_{1}$i si tya	$\mathbf{s}_{1+}$i si tya
sisifo	si si f$\mathbf{o}_{1}$	s$\mathbf{i}_{2}$ s$\mathbf{i}_{3}$ f$\mathbf{o}_{1}$	s$\mathbf{i}_{2+}$ s$\mathbf{i}_{3+}$ f$\mathbf{o}_{1+}$
sisilo	si si lo	$\mathbf{s}_{1}$i si lo	$\mathbf{s}_{1+}$i si lo
zizitya	$\mathbf{z}_{1}$i zi tya	$\mathbf{z}_{1}\mathbf{i}_{3}$ $\mathbf{z}_{2}$i $\mathbf{ty}_{4}$a	$\mathbf{z}_{1+}\mathbf{i}_{2+}$ $\mathbf{z}_{4+}$i $\mathbf{ty}_{3+}$a
zizifo	zi $\mathbf{z}_{1}\mathbf{i}_{1}$ fo	$\mathbf{z}_{6}\mathbf{i}_{5}$ $\mathbf{z}_{2}\mathbf{i}_{1}$ $\mathbf{f}_{4}\mathbf{o}_{3}$	$\mathbf{z}_{6+}\mathbf{i}_{5+}$ $\mathbf{z}_{4+}\mathbf{i}_{1+}$ $\mathbf{f}_{3+}\mathbf{o}_{2+}$
zizilo	zi $\mathbf{z}_{1}$i lo	zi $\mathbf{z}_{1}$i lo	zi $\mathbf{z}_{1+}$i lo
yinja	y$\mathbf{i}_{1}$ $\mathbf{nj}_{1}\mathbf{a}_{1}$	y$\mathbf{i}_{2}$ $\mathbf{nj}_{1}\mathbf{a}_{1}$	y$\mathbf{i}_{3+}$ $\mathbf{nj}_{2+}\mathbf{a}_{1+}$
yintloko	yi ntlo ko	$\mathbf{y}_{3}\mathbf{i}_{1}$ $\mathbf{ntl}_{4}$o $\mathbf{k}_{2}$o	$\mathbf{y}_{2+}\mathbf{i}_{1+}$ $\mathbf{ntl}_{4+}$o $\mathbf{k}_{3+}$o
yinkomo	yi nko $\mathbf{m}_{1}$o	yi nko $\mathbf{m}_{1}\mathbf{o}_{2}$	yi nko $\mathbf{m}_{1+}\mathbf{o}_{2+}$
zizinja	zi zi nja	$\mathbf{z}_{2}$i z$\mathbf{i}_{3}$ $\mathbf{nj}_{1}\mathbf{a}_{2}$	$\mathbf{z}_{3+}$i z$\mathbf{i}_{4+}$ $\mathbf{nj}_{2+}\mathbf{a}_{1+}$
ziintloko	zii ntlo ko	zi$\mathbf{i}_{2}$ $\mathbf{ntl}_{3}\mathbf{o}_{1}$ k$\mathbf{o}_{4}$	zi$\mathbf{i}_{1+}$ $\mathbf{ntl}_{2+}\mathbf{o}_{3+}$ k$\mathbf{o}_{4+}$
ziinkomo	zii nk$\mathbf{o}_{1}$ $\mathbf{m}_{1}\mathbf{o}_{1}$	zii $\mathbf{nk}_{3}\mathbf{o}_{2}$ $\mathbf{m}_{1}\mathbf{o}_{1}$	zii $\mathbf{nk}_{4+}\mathbf{o}_{2+}$ $\mathbf{m}_{1+}\mathbf{o}_{3+}$
luluthi	lu $\mathbf{l}_{1}$u thi	lu $\mathbf{l}_{3}\mathbf{u}_{4}$ $\mathbf{th}_{2}\mathbf{i}_{1}$	lu $\mathbf{l}_{3+}\mathbf{u}_{2+}$ $\mathbf{th}_{4+}\mathbf{i}_{1+}$
lulwimi	lu $\mathbf{lw}_{1}\mathbf{i}_{1}$ $\mathbf{m}_{1}$i	lu $\mathbf{lw}_{4}\mathbf{i}_{2}$ $\mathbf{m}_{1}\mathbf{i}_{3}$	lu $\mathbf{lw}_{3+}\mathbf{i}_{2+}$ $\mathbf{m}_{1+}\mathbf{i}_{4+}$
lulwandle	lu lwa $\mathbf{ndl}_{1}\mathbf{e}_{1}$	lu $\mathbf{lw}_{3}$a $\mathbf{ndl}_{1}\mathbf{e}_{2}$	lu $\mathbf{lw}_{3+}$a $\mathbf{ndl}_{1+}\mathbf{e}_{2+}$
bubuso	bu b$\mathbf{u}_{1}$ $\mathbf{s}_{1}$o	bu $\mathbf{b}_{4}\mathbf{u}_{1}$ $\mathbf{s}_{3}\mathbf{o}_{2}$	bu $\mathbf{b}_{3+}\mathbf{u}_{1+}$ $\mathbf{s}_{4+}\mathbf{o}_{2+}$
bubuhlanti	bu bu hla nti	bu bu hl$\mathbf{a}_{1}$ nt$\mathbf{i}_{2}$	bu bu hl$\mathbf{a}_{1+}$ nt$\mathbf{i}_{2+}$
bubusika	bu bu $\mathbf{s}_{1}\mathbf{i}_{1}$ $\mathbf{k}_{1}\mathbf{a}_{1}$	bu b$\mathbf{u}_{3}$ $\mathbf{s}_{1}\mathbf{i}_{1}$ $\mathbf{k}_{2}\mathbf{a}_{2}$	bu b$\mathbf{u}_{3+}$ $\mathbf{s}_{5+}\mathbf{i}_{4+}$ $\mathbf{k}_{2+}\mathbf{a}_{1+}$
kukutya	ku ku $\mathbf{ty}_{1}$a	ku $\mathbf{k}_{3}$u $\mathbf{ty}_{1}\mathbf{a}_{2}$	ku $\mathbf{k}_{3+}$u $\mathbf{ty}_{2+}\mathbf{a}_{1+}$
kukwindla	ku kw$\mathbf{i}_{1}$ $\mathbf{ndl}_{1}\mathbf{a}_{1}$	ku kw$\mathbf{i}_{1}$ $\mathbf{ndl}_{2}\mathbf{a}_{1}$	ku kw$\mathbf{i}_{3+}$ $\mathbf{ndl}_{2+}\mathbf{a}_{1+}$
kukunene	ku k$\mathbf{u}_{1}$ $\mathbf{n}_{1}$e $\mathbf{n}_{1}$e	ku k$\mathbf{u}_{1}$ $\mathbf{n}_{3}\mathbf{e}_{4}$ $\mathbf{n}_{2}\mathbf{e}_{4}$	ku k$\mathbf{u}_{2+}$ $\mathbf{n}_{5+}\mathbf{e}_{3+}$ $\mathbf{n}_{1+}\mathbf{e}_{4+}$

APPENDIX B

PERCEPTION EXPERIMENT DATA

B.1 Stimuli duration

Table B.1 Perception Tests A1 and B1: Duration of the penultimate vowels of stimuli.

	Original	Stimulus A	Stimulus B	Stimulus C	Stimulus D	Stimulus E
ngumfana (A03)						
Queclarative	177.7 ms	240.7 ms	303.7 ms	370.2 ms	433.2 ms	496.2 ms
Statement	370.2 ms	310.1 ms	243.6 ms	176.7 ms	116 ms	
ngabafana (B03)						
Queclarative	127.7 ms	190.4 ms	253.1 ms	319.1 ms	378.5 ms	441.2 ms
Statement	317.6 ms	255.2 ms	193 ms	125 ms	69.8 ms	
ngumthi (C01)						
Queclarative	119.5 ms	176.5 ms	233.5 ms	294.3 ms	351.3 ms	412.1 ms
Statement	295.5 ms	240.1 ms	175.2 ms	118.6 ms	79.2 ms	
yimililo (D02)						
Queclarative	112.8 ms	160.8 ms	208.8 ms	256.8 ms	304.8 ms	352.8 ms
Statement	256.8 ms	208.7 ms	169.7 ms	112.6 ms	77.5 ms	
lilitye (E01)						
Queclarative	158.9 ms	221.6 ms	290 ms	358.4 ms	421.1 ms	489.5 ms
Statement	358.5 ms	289.9 ms	226.1 ms	151.2 ms	86.7 ms	
ngamafu (F03)						
Queclarative	176.9 ms	247.7 ms	318.5 ms	389.3 ms	460.1 ms	530.9 ms
Statement	390.7 ms	316.3 ms	247 ms	172.4 ms	93.9 ms	
sisilo (G03)						
Queclarative	136.3 ms	196.3 ms	250.3 ms	304.3 ms	362 ms	423 ms
Statement	308.5 ms	248.7 ms	194.1 ms	139.1 ms	89.5 ms	
zizifo (H02)						
Queclarative	108.7 ms	137.4 ms	170.2 ms	198.9 ms	227.6 ms	256.3 ms
Statement	198.8 ms	165.7 ms	140.2 ms	107 ms	82.6 ms	
yintloko (I02)						
Queclarative	205.9 ms	267.9 ms	329.9 ms	391.9 ms	453.9 ms	515.9 ms
Statement	393.5 ms	329.8 ms	265.7 ms	207.4 ms	148.9 ms	
ziinkomo (J03)						
Queclarative	201.8 ms	243.2 ms	289.2 ms	326 ms	372 ms	413.4 ms
Statement	329.5 ms	290 ms	242.5 ms	203.3 ms	169.9 ms	
lulwimi (K02)						
Queclarative	153.5 ms	189.5 ms	225.5 ms	255.5 ms	293.8 ms	324.3 ms
Statement	258.7 ms	220.8 ms	188.3 ms	149.2 ms	115.1 ms	
bubuhlanti (L02)						
Queclarative	178.4 ms	226.5 ms	270.9 ms	319 ms	363.4 ms	407.8 ms
Statement	318.3 ms	271.9 ms	225.3 ms	178.7 ms	145.4 ms	
kukutya (M01)						
Queclarative	123.9 ms	167.9 ms	211.9 ms	251.5 ms	295.5 ms	339.5 ms
Statement	253.9 ms	209 ms	163.4 ms	117 ms	86.5 ms	

B.2 Stimuli pitch

Table B.2 Perception Test A2: Pitch on the penultimate vowels of stimuli.

	Stimulus A	Stimulus B_1	Stimulus B_2	Stimulus C
ngumfana (A03)				
Queclarative	256 Hz	232 Hz	232 Hz	212 Hz
Statement	182 Hz	196 Hz		208 Hz
lilitye (E01)				
Queclarative	333 Hz	294 Hz	294 Hz	263 Hz
Statement	212 Hz	239 Hz		
yintloko (I02)				
Queclarative	297 Hz	272 Hz	272 Hz	236 Hz
Statement	181 Hz	192 Hz		204 Hz
lulwimi (K02)				
Queclarative	287 Hz	270 Hz	270 Hz	254 Hz
Statement	200 Hz	217 Hz		
bubuhlanti (L02)				
Queclarative	244 Hz	223 Hz	223 Hz	205 Hz
Statement	167 Hz	189 Hz		

Table B.3 Perception Test A2: Pitch on the first vowel of stimulus I02.

	Stimulus D	Stimulus E_1	Stimulus E_2	Stimulus F
yintloko (I02)				
Queclarative	193 Hz	171 Hz	171 Hz	161 Hz
Statement	149 Hz	163 Hz		178 Hz

Table B.4 Perception Test A2: Pitch on the final vowel of stimulus I02.

	Stimulus G	Stimulus H_1	Stimulus H_2	Stimulus I
yintloko (I02)				
Queclarative	274 Hz	250 Hz	250 Hz	233 Hz
Statement	178 Hz	191 Hz		203 Hz

Table B.5 Perception Test B2: Pitch on the penultimate vowels of stimuli.

	Original	Stimulus D	Stimulus E	Stimulus F	Stimulus G
ngumfana (A03)					
Queclarative	297 Hz	252 Hz	219 Hz	193 Hz	168 Hz
Statement	168 Hz	198 Hz	223 Hz	251 Hz	273 Hz
lilitye (E01)					
Queclarative	347 Hz	304 Hz	275 Hz	250 Hz	228 Hz
Statement	187 Hz	218 Hz	243 Hz	270 Hz	298 Hz
ngamafu (F03)					
Queclarative	339 Hz	309 Hz	281 Hz	257 Hz	244 Hz
Statement	175 Hz	207 Hz	229 Hz	268 Hz	309 Hz
yintloko (I02)					
Queclarative	320 Hz	279 Hz	250 Hz	228 Hz	200 Hz
Statement	172 Hz	198 Hz	228 Hz	261 Hz	292 Hz
lulwimi (K02)					
Queclarative	347 Hz	325 Hz	302 Hz	280 Hz	250 Hz
Statement	194 Hz	213 Hz	239 Hz	277 Hz	294 Hz
bubuhlanti (L02)					
Queclarative	285 Hz	270 Hz	241 Hz	219 Hz	193 Hz
Statement	151 Hz	182 Hz	208 Hz	233 Hz	255 Hz

Table B.6 Perception Test B3: Pitch on the first vowels of stimuli.

	Original	Stimulus A	Stimulus B	Stimulus C
ngumfana (A03)				
Queclarative	160 Hz	148 Hz	139 Hz	130 Hz
Statement	122 Hz	132 Hz	140 Hz	150 Hz
ngabafana (B03)				
Queclarative	156 Hz	148 Hz	141 Hz	135 Hz
Statement	129 Hz	135 Hz	142 Hz	149 Hz
ngumthi (C01)				
Queclarative	304 Hz	253 Hz	223 Hz	205 Hz
Statement	205 Hz	232 Hz	261 Hz	291 Hz
yimililo (D02)				
Queclarative	172 Hz	163 Hz	156 Hz	149 Hz
Statement	141 Hz	149 Hz	156 Hz	164 Hz
lilitye (E01)				
Queclarative	222 Hz	207 Hz	195 Hz	187 Hz
Statement	149 Hz	161 Hz	173 Hz	184 Hz
ngamafu (F03)				
Queclarative	189 Hz	172 Hz	164 Hz	154 Hz
Statement	131 Hz	147 Hz	159 Hz	173 Hz
sisilo (G03)				
Queclarative	213 Hz	203 Hz	195 Hz	192 Hz
Statement	161 Hz	169 Hz	178 Hz	185 Hz
zizifo (H02)				
Queclarative	344 Hz	297 Hz	262 Hz	241 Hz
Statement	209 Hz	246 Hz	268 Hz	292 Hz
yintloko (I02)				
Queclarative	204 Hz	196 Hz	185 Hz	174 Hz
Statement	130 Hz	140 Hz	151 Hz	162 Hz
ziinkomo (J03)				
Queclarative	346 Hz	322 Hz	294 Hz	278 Hz
Statement	313 Hz	322 Hz	334 Hz	358 Hz
lulwimi (K02)				
Queclarative	361 Hz	322 Hz	303 Hz	277 Hz
Statement	262 Hz	270 Hz	277 Hz	286 Hz
bubuhlanti (L02)				
Queclarative	286 Hz	263 Hz	250 Hz	238 Hz
Statement	249 Hz	269 Hz	285 Hz	303 Hz
kukutya (M01)				
Queclarative	303 Hz	293 Hz	277 Hz	262 Hz
Statement	200 Hz	212 Hz	222 Hz	238 Hz

B.3 Duration of gates

Table B.7 Perception Tests A3 and B4: Duration of gates [in ms].

ngumfana (A03)								
	d_{ng}	d_{ngu}	d_{ngum}	d_{ngumf}	d_{ngumfa}	$d_{ngumfan}$	$d_{ngumfana}$	
Queclarative	78.2	96.7	152.1	242.5	381.1	503.4	545.5	
Statement	137.4	162.4	238.8	355.9	599.4	827.2	891.2	
ngabafana (B03)								
	d_{ng}	d_{nga}	d_{ngab}	d_{ngaba}	d_{ngabaf}	$d_{ngabafa}$	$d_{ngabafan}$	$d_{ngabafana}$
Queclarative	97.2	145.2	206.9	272.1	365.7	480	580.6	634.3
Statement	121.7	164.9	224.1	305.7	417.7	633.7	840.1	907.3

Table B.7 Perception Tests A3 and B4: Duration of gates (continued).

ngumthi (C01)									
	d_{ng}	d_{ngu}	d_{ngum}	d_{ngumth}	$d_{ngumthi}$				
Queclarative	50.7	129.3	265.4	368.4	463.6				
Statement	106.4	178.2	389.5	622.9	755.5				
yimililo (D02)									
	d_{y}	d_{yi}	d_{yim}	d_{yimi}	d_{yimil}	d_{yimili}	$d_{yimilil}$	$d_{yimililo}$	
Queclarative	38.9	49.2	71.5	130.8	275.9	373	487	574.1	
Statement	47.1	61.2	119.9	206.1	292.3	491.6	690.9	750.1	
lilitye (E01)									
	d_{l}	d_{li}	d_{lil}	d_{lili}	d_{lility}	$d_{lilitye}$			
Queclarative	52.8	93.4	173.9	299	428.1	511.5			
Statement	64.9	111.1	200.5	429.8	670.5	764.1			
ngamafu (F03)									
	d_{ng}	d_{nga}	d_{ngam}	d_{ngama}	d_{ngamaf}	$d_{ngamafu}$			
Queclarative	46.8	95.5	174.4	299.1	452.9	533.8			
Statement	63.7	110.3	196.6	426.9	701.1	805.7			
sisilo (G03)									
	d_{s}	d_{si}	d_{sis}	d_{sisi}	d_{sisil}	d_{sisilo}			
Queclarative	30.1	53.3	152	277	386.4	466.3			
Statement	49.2	72.7	184.9	401.6	605.6	691			
zizifo (H02)									
	d_{z}	d_{zi}	d_{ziz}	d_{zizi}	d_{zizif}	d_{zizifo}			
Queclarative	60.7	130.7	221.2	301.3	410.3	528.6			
Statement	112.8	174.2	289.7	452.2	633.6	781.3			
yintloko (I02)									
	d_{y}	d_{yi}	d_{yin}	d_{yintl}	d_{yintlo}	$d_{yintlok}$	$d_{yintloko}$		
Queclarative	56.8	98.2	150.2	210.9	350.4	486.9	540.8		
Statement	32.4	69	130.8	230.6	473.6	711.2	767.4		
ziinkomo (J03)									
	d_{z}	d_{zii}	d_{ziin}	d_{ziink}	d_{ziinko}	$d_{ziinkom}$	$d_{ziinkomo}$		
Queclarative	39.4	111.9	204.2	254.9	374.9	505.1	595.1		
Statement	119.7	229	371.2	444	636.5	844.5	938.2		
lulwimi (K02)									
	d_{l}	d_{lu}	d_{lul}	d_{lulw}	d_{lulwi}	d_{lulwim}	$d_{lulwimi}$		
Queclarative	49.6	96.9	182.5	248.5	352.5	472	548.3		
Statement	58.4	109.1	205.5	279	440.1	602.4	669.6		
bubuhlanti (L02)									
	d_{b}	d_{bu}	d_{bub}	d_{bubu}	d_{bubuhl}	$d_{bubuhla}$	$d_{bubuhlan}$	$d_{bubuhlant}$	$d_{bubuhlanti}$
Queclarative	39.1	67.4	125.3	198.5	296.5	434.6	543.3	596.4	660.2
Statement	38.3	66.2	129.3	213.9	333.8	559.8	756.7	829	882.8
kukutya (M01)									
	d_{k}	d_{ku}	d_{kuk}	d_{kuku}	d_{kukuty}	$d_{kukutya}$			
Queclarative	41.1	68.6	157.9	259.4	374	486.7			
Statement	30.7	80.6	174.5	327.3	519.7	640			

B.4 Perception test evaluation criteria

Table B.8 Criteria used in interpreting the perception test results.

Abbreviation	Description
sm$X$	significant majority X responses (χ^2)
mX	majority X responses
sc	significant change (χ^2)
scp	significant change in perception
cp	change in perception
ncp	no change in perception
mi	misidentification
mrt+	mean response time increased
mrt−	mean response time decreased
Q	number of Queclarative responses
S	number of Statement responses
coi	cross over point improbable
co at x	cross over point at stimulus x
pco $< x$	possible cross over point before stimulus x
pco $> x$	possible cross over point after stimulus x

B.5 Results of Experiment A

Table B.9 Results of Perception Test A1.

	Original	Stimulus A	Stimulus B_1	Stimulus B_2	Stimulus C	Comments
ngumfana (A03)						
Queclarative	smQ	smQ	mS	mQ	smS	co at B, scp
Statement	smS	smS	smS	smS	smS	coi, cp
ngabafana (B03)						
Queclarative	smQ	smQ	smQ	smQ	smS	co at C, scp
Statement	smS	smS	smS	smS	smS	coi, cp
ngumthi (C01)						
Queclarative	smQ	smQ	smQ	smQ	smQ	coi, ncp
Statement	smS	smS	smS	smS	smS	coi, ncp
yimililo (D02)						
Queclarative	smQ	smQ	smQ	smQ	smQ	pco > C, cp
Statement	smS	smS	smS	smS	smS	coi, ncp
lilitye (E01)						
Queclarative	smQ	smQ	smQ	mQ	smQ	pco > C, cp
Statement	smS	smS	smS	smS	smS	coi, cp
ngamafu (F03)						
Queclarative	smQ	smQ	mS	mQ	mQ	co at B, scp
Statement	smS	smS	smS	smS	smS	coi, cp
sisilo (G03)						
Queclarative	smQ	smQ	smQ	smQ	smQ	coi, ncp
Statement	smS	smS	smS	smS	smS	coi, ncp
zizifo (H02)						
Queclarative	smQ	smQ	smQ	smQ	smQ	coi, ncp
Statement	smS	smS	smS	smS	smS	coi, ncp

Table B.9 Results of Perception Test A1 (continued).

	Original	Stimulus A	Stimulus B_1	Stimulus B_2	Stimulus C	Comments
yintloko (I02)						
Queclarative	smQ	smQ	smQ	mQ	mQ	pco > C, cp
Statement	smS	smS	smS	smS	smS	coi, ncp
ziinkomo (J03)						
Queclarative	smQ	smQ	smQ	smQ	smQ	coi, ncp
Statement	smS	smS	smS	smS	smS	coi, ncp
lulwimi (K02)						
Queclarative	smQ	smQ	smQ	smQ	smQ	coi, ncp
Statement	smS	smS	smS	smS	smS	coi, ncp
bubuhlanti (L02)						
Queclarative	smQ	smQ	smQ	smQ	smQ	coi, cp
Statement	smS	smS	smS	smS	smS	coi, cp
kukutya (M01)						
Queclarative	smQ	smQ	smQ	smQ	smQ	coi, ncp
Statement	smS	smS	smS	smS	smS	coi, ncp

Table B.10 Results of Perception Test A2.

	Stimulus A	Stimulus B_1	Stimulus B_2	Stimulus C
ngumfana (A03)				
Queclarative	90%	92%	85%	73%
Statement	90%	93%		88%
lilitye (E01)				
Queclarative	88%	93%	95%	70%
Statement	92%	88%		
yintloko (I02)				
Queclarative	88%	92%	90%	70%
Statement	95%	95%		85%
lulwimi (K02)				
Queclarative	93%	85%	88%	93%
Statement	90%	87%		
bubuhlanti (L02)				
Queclarative	83%	65%	82%	57%
Statement	83%	85%		

Table B.11 Results of Perception Test A2: pitch manipulation on the first vowel.

	Stimulus D	Stimulus E_1	Stimulus E_2	Stimulus F
yintloko (I02)				
Queclarative	88%	92%	90%	92%
Statement	90%	92%		90%

Table B.12 Results of Perception Test A2: pitch manipulation on the final vowel.

	Stimulus G	Stimulus H_1	Stimulus H_2	Stimulus I
yintloko (I02)				
Queclarative	95%	92%	93%	90%
Statement	93%	90%		92%

Table B.13 Results of Perception Test A3.

	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5	Gate 6	Gate 7	Gate 8	Gate 9	Gate 10
ngumfana (A03)										
	ngum	$ngumf_1$	$ngumf_2$	$ngumfa_1$	$ngumfa_2$	ngumfan	ngumfana	ngumfana		
% Queclarative	90%	83.3%	75%	93.3%	90%	85%	75%	88.3%		
% Statement	80%	91.7%	90%	96.7%	86.7%	86.7%	88.3%	91.7%		

Table B.13 Results of Perception Test A3 (continued).

	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5	Gate 6	Gate 7	Gate 8	Gate 9	Gate 10
ngabafana (B03)										
	ngab	ngaba$_1$	ngaba$_2$	ngabaf	ngabafa	ngabafan$_1$	ngabafan$_2$	ngabafana	ngabafana	
% Queclarative	66.7%	88.3%	88.3%	91.7%	95%	85%	96.7%	88.3%	95%	
% Statement	90%	83.3%	86.7%	86.7%	85%	85%	93.3%	88.3%	90%	
ngumthi (C01)										
	ngum	ngumth$_1$	ngumth$_2$	ngumthi$_1$	ngumthi$_2$	ngumthi				
% Queclarative	96.7%	80%	91.7%	88.3%	96.7%	80%				
% Statement	86.7%	83.3%	80%	86.7%	96.7%	90%				
yimililo (D02)										
	yim	yimi$_1$	yimi$_2$	yimil	yimili	yimilil$_1$	yimilil$_2$	yimililo	yimililo	
% Queclarative	81.7%	71.7%	75%	91.7%	93.3%	81.7%	95%	83.3%	91.7%	
% Statement	75%	76.7%	85%	88.3%	91.7%	88.3%	96.7%	91.7%	90%	
lilitye (E01)										
	lil	lili$_1$	lili$_2$	lility$_1$	lility$_2$	lilitye	lilitye			
% Queclarative	90%	93.3%	91.7%	95%	90%	90%	83.3%			
% Statement	86.7%	90%	93.3%	93.3%	93.3%	91.7%	91.7%			
ngamafu (F03)										
	ngam	ngama$_1$	ngama$_2$	ngamaf$_1$	ngamaf$_2$	ngamafu	ngamafu			
% Queclarative	80%	88.3%	93.3%	91.7%	96.7%	95%	88.3%			
% Statement	88.3%	90%	93.3%	91.7%	91.7%	96.7%	90%			
sisilo (G03)										
	sis	sisi$_1$	sisi$_2$	sisil$_1$	sisil$_2$	sisilo	sisilo			
% Queclarative	71.7%	95%	96.7%	95%	78.3%	93.3%	71.7%			
% Statement	88.3%	95%	90%	86.7%	90%	91.7%	83.3%			
zizifo (H02)										
	ziz	zizi$_1$	zizi$_2$	zizif$_1$	zizif$_2$	zizifo	zizifo			
% Queclarative	93.3%	86.7%	93.3%	96.7%	98.3%	90%	90%			
% Statement	88.3%	93.3%	98.3%	96.7%	91.7%	91.7%	91.7%			
yintloko (I02)										
	yin	yintl$_1$	yintl$_2$	yintlo	yintlok$_1$	yintlok$_2$	yintloko	yintloko		
% Queclarative	83.3%	91.7%	90%	91.7%	91.7%	93.3%	91.7%	95%		
% Statement	83.3%	88.3%	91.7%	86.7%	90%	91.7%	91.7%	88.3%		
ziinkomo (J03)										
	ziin	ziink$_1$	ziink$_2$	ziinko	ziinkom$_1$	ziinkom$_2$	ziinkomo	ziinkomo		
% Queclarative	95%	93.3%	95%	98.3%	96.7%	90%	91.7%	93.3%		
% Statement	91.7%	88.3%	81.7%	83.3%	90%	93.3%	91.7%	90%		
lulwimi (K02)										
	lul	lulw$_1$	lulw$_2$	lulwi	lulwim$_1$	lulwim$_2$	lulwimi	lulwimi		
% Queclarative	88.3%	93.3%	95%	95%	91.7%	91.7%	80%	96.7%		
% Statement	91.7%	88.3%	88.3%	93.3%	85%	95%	83.3%	91.7%		
bubuhlanti (L02)										
	bub	bubu$_1$	bubu$_2$	bubuhl	bubuhla	bubuhlan$_1$	bubuhlan$_2$	bubuhlant	bubuhlanti	bubuhlanti
% Queclarative	88.3%	88.3%	88.3%	83.3%	88.3%	93.3%	86.7%	81.7%	95%	93.3%
% Statement	48.3%	53.3%	81.7%	85%	93.3%	93.3%	85%	91.7%	91.7%	91.7%
kukutya (M01)										
	kuk	kuku$_1$	kuku$_2$	kukuty$_1$	kukuty$_2$	kukutya	kukutya			
% Queclarative	80%	91.7%	86.7%	88.3%	83.3%	86.7%	88.3%			
% Statement	88.3%	93.3%	86.7%	91.7%	93.3%	90%	91.7%			

B.6 Results of Experiment B

Table B.14 Results of Perception Test B1.

	Original	A	B	C	D	E	Comments
ngumfana (A03)							
Queclarative	smQ	smQ	mQ, sc	smS, sc	smS	smS	co at C, scp
Statement	smS	smS	smS	smS, sc	smS		pco > D, cp, mrt+
ngabafana (B03)							
Queclarative	smQ	smQ	mQ, sc	mQ	mQ	mS, sc	co at E, scp
Statement	smS	smS	smS	mS, sc	mS		pco > D, cp, mrt+
ngumthi (C01)							
Queclarative	smQ	smQ	smQ	smQ	smQ	smQ	pco > E, cp, mrt+
Statement	smS	smS	smS	smS	smS		coi, ncp, mrt+
yimililo (D02)							
Queclarative	smQ	smQ	smQ, sc	mQ, sc	mQ	mQ	pco > E, cp
Statement	smS	smS	smS	smS	smS		coi, cp, mrt+
lilitye (E01)							
Queclarative	smQ	smQ	smQ, sc	Q=S	mS	mS	co at C, scp, mrt-
Statement	smS	smS	smS	smS	smS		coi, cp, mrt+
ngamafu (F03)							
Queclarative	smQ	smQ	Q=S, sc	mQ	smS, sc	mS	co at D, scp, mrt-
Statement	smS	smS	smS	mS, sc	smS		pco > D, cp, mrt+
sisilo (G03)							
Queclarative	smQ	smQ	mS, sc	smQ	mS, sc	mQ	co at D, scp, mrt+
Statement	smS	smS	smS, sc	smS	smS		pco > D, cp, mrt+
zizifo (H02)							
Queclarative	smQ	smQ	smQ	smQ	smQ	smQ	coi,ncp, mrt+
Statement	smS	smS	smS	smS	smS		coi, ncp
yintloko (I02)							
Queclarative	smQ	smQ	mS, sc	mS	mQ	smS, sc	co at B, scp, mrt-
Statement	smS	smS	smS	smS	mS, sc		pco > D, cp, mrt+
ziinkomo (J03)							
Queclarative	smQ	smQ	smQ	smQ	smQ, sc	smQ, sc	coi, cp, mrt-
Statement	smS	smS	smS	smS	smS		coi, ncp
lulwimi (K02)							
Queclarative	smQ	smQ	smQ, sc	smQ	smQ	smQ, sc	coi, cp, mrt-
Statement	smS	smS	smS	smS	smS		coi, ncp
bubuhlanti (L02)							
Queclarative	smQ	smQ	smQ	mQ, sc	mQ	mS	co at E, scp
Statement	smS	smS	smS	smS	smS		coi, ncp, mrt+
kukutya (M01)							
Queclarative	smQ	smQ	smQ	smQ	smQ	smQ	coi, cp, mrt-
Statement	smS	smS	smS	smS	smS		coi, cp, mrt-

Table B.15 Results of Perception Test B2.

	Stimulus D	Stimulus E	Stimulus F	Stimulus G	Comments
ngumfana (A03)					
Queclarative	smQ	smQ	mS, sc	smS	co at F, scp, mrt-
% Queclarative	77.8%	65.1%	38.1%	30.2%	
Statement	smS	smS	mQ, sc	smQ, sc	co at F, scp, mrt+
% Statement	88.9%	76.2%	46%	22.2%	
lilitye (E01)					
Queclarative	smQ	smQ	smQ	mQ	pco > G, cp, mrt+
% Queclarative	90.5%	79.4%	69.8%	60.3%	
Statement	smS	smS	smS, sc	smS	cp, mrt+
% Statement	84.1%	88.9%	74.6%	82.5%	

Table B.15 Results of Perception Test B2 (continued).

	Stimulus D	Stimulus E	Stimulus F	Stimulus G	Comments
ngamafu (F03)					
Queclarative	smQ	smQ	smQ	smQ	cp, mrt+
% Queclarative	85.7%	79.4%	73%	71.4%	
Statement	smS	smS	smS	smS	cp, mrt
% Statement	90.5%	88.9%	88.9%	77.8%	
yintloko (I02)					
Queclarative	smQ	mQ	smQ, sc	mQ, sc	cp, mrt+
% Queclarative	71.4%	55.6%	81%	55.6%	
Statement	smS	smS	smS	smS, sc	pco > G, cp, mrt+
% Statement	87.3%	85.7%	82.5%	63.5%	
lulwimi (K02)					
Queclarative	smQ	smQ	smQ	mQ	pco > G, cp, mrt+
% Queclarative	87.3%	81%	74.6%	60.3%	
Statement	smS	smS	smS	mS	pco > G, cp, mrt+
% Statement	90.5%	84.1%	65.1%	50.8%	
bubuhlanti (L02)					
Queclarative	smQ	smQ	mQ, sc	mS, sc	co at G, scp, mrt+
% Queclarative	85.7%	93.7%	58.7%	41.3%	
Statement	smS	smS	Q=S, sc	mQ	co F, scp, mrt+
% Statement	84.1%	73%	49.2%	41.3%	

Table B.16 Results of Perception Test B3.

	Stimulus A	Stimulus B	Stimulus C	Comments
ngumfana (A03)				
Queclarative	smQ	mS, sc	smS	co at B, scp, mrt+
% Queclarative	55.6%	38.1%	23.8%	
Statement	smS	smS	mS	pco > C, cp, mrt-
% Statement	60.3%	57.1%	46%	
ngabafana (B03)				
Queclarative	smQ	smQ	mQ	pco > C, cp, mrt+
% Queclarative	61.9%	60.3%	41.3%	
Statement	smS	smS	mS	pco > C, cp, mrt+
% Statement	73%	61.9%	52.4%	
ngumthi (C01)				
Queclarative	smQ	mQ, sc	smQ, sc	pco > C, cp, mrt+
% Queclarative	81%	52.4%	69.8%	
Statement	smQ	smQ	smQ	mi
% Statement	30.2%	19%	19%	
yimililo (D02)				
Queclarative	smS	mS	mS	mi, mrt+
% Queclarative	30.2%	39.7%	34.9%	
Statement	smS	mS	mS	pco > C, cp, mrt-
% Statement	61.9%	52.4%	55.6%	
lilitye (E01)				
Queclarative	mQ	smQ	mQ	ncp, mrt+
% Queclarative	52.4%	61.9%	50.8%	
Statement	smS	smS	smS	cp, mrt+
% Statement	76.2%	73%	66.7%	
ngamafu (F03)				
Queclarative	mQ	smQ	mS, sc	co at C, scp, mrt+
% Queclarative	58.7%	65.1%	42.9%	
Statement	mS	mS	Q=S	pco > C, cp, mrt-
% Statement	55.6%	55.6%	47.6%	

Table B.16 Results of Perception Test B3 (continued).

	Stimulus A	Stimulus B	Stimulus C	Comments
sisilo (G03)				
Queclarative	mS	mS	mQ	mi, mrt+
% Queclarative	44.4%	41.3%	52.4%	
Statement	mS	mQ	mS	ncp, mrt-
% Statement	55.6%	44.4%	54%	
zizifo (H02)				
Queclarative	smQ	smQ	smQ	cp, mrt+
% Queclarative	82.5%	79.4%	68.3%	
Statement	mQ	smQ, sc	smQ	mi
% Statement	34.9%	14.3%	22.2%	
yintloko (I02)				
Queclarative	mQ	mQ	mS	co at C, scp, mrt+
% Queclarative	58.7%	58.7%	42.9%	
Statement	smS	smS	smS	cp, mrt+
% Statement	76.2%	66.7%	68.3%	
ziinkomo (J03)				
Queclarative	smQ	smQ	smQ	cp, mrt+
% Queclarative	82.5%	76.2%	69.8%	
Statement	smQ	smQ, sc	smQ	mi, mrt+
% Statement	36.5%	17.5%	7.9%	
lulwimi (K02)				
Queclarative	smQ	smQ	mQ	pco > C, cp, mrt+
% Queclarative	76.2%	73%	60.3%	
Statement	mS	mS	mQ	co at C, scp, mrt-
% Statement	50.8%	50.8%	39.7%	
bubuhlanti (L02)				
Queclarative	mQ	smQ	mQ	ncp, mrt+
% Queclarative	50.8%	65.1%	52.4%	
Statement	mQ	mQ	smQ, sc	mi, mrt-
% Statement	41.3%	38.1%	17.5%	
kukutya (M01)				
Queclarative	smQ	smQ	mQ	pco > C, cp, mrt+
% Queclarative	68.3%	65.1%	54%	
Statement	smS	smS	mS, sc	pco > C, cp, mrt+
% Statement	69.8%	71.4%	52.4%	

Table B.17 Results of Perception Test B4[1].

	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5	Gate 6	Gate 7	Gate 8	Gate 9	Comment
ngumfana (A03)										
	ng	ngu	ngum	ngumf	ngumfa	ngumfan	ngumfana			
Queclarative	smS	**mQ**	smQ	smQ	smQ	smQ	smQ			mrt-
% Queclarative	30.2%	**44.4%**	61.9%	74.6%	87.3%	81%	88.9%			
Statement	**smS**	smS	mS	smS	smS	smS	smS			mrt-
% Statement	**69.8%**	74.6%	50.8%	69.8%	77.8%	76.2%	81%			
ngabafana (B03)										
	ng	nga	ngab	ngaba	ngabaf	ngabafa	ngabafan	ngabafana		
Queclarative	mS	**mQ**	mQ	smQ, sc	smQ	smQ	smQ	smQ		mrt-
% Queclarative	44.4%	**54%**	55.6%	77.8%	84.1%	84.1%	88.9%	84.1%		
Statement	**smS**	smS	smS	smS, sc	smS	smS	smS	smS		mrt-
% Statement	**65.1%**	77.8%	68.3%	87.3%	79.4%	85.7%	90.5%	93.7%		

1 The bold typeface indicates the first gate for which a majority correct responses have been obtained.

Table B.17 Results of Perception Test B4[2] (continued).

	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5	Gate 6	Gate 7	Gate 8	Gate 9	Comment
ngumthi (C01)										
	ng	ngu	ngum	ngumth	ngumthi					
Queclarative	mS	**smQ,sc**	smQ	smQ	smQ					mrt-
% Queclarative	41.3%	**76.2%**	87.3%	85.7%	84.1%					
Statement	**smS**	smS	smS	smS, sc	smS					mrt-
% Statement	**69.8%**	65.1%	60.3%	82.5%	87.3%					
yimililo (D02)										
	y	yi	yim	yimi	yimil	yimili	yimilil	yimililo		
Queclarative	smS	mS	**mQ**	smQ	smQ, sc	smQ	smQ	smQ		mrt-
% Queclarative	28.6%	41.3%	**52.4%**	68.3%	82.5%	87.3%	84.1%	84.1%		
Statement	**smS**	smS	smS	smS	smS	smS	smS	smS		mrt-
% Statement	**65.1%**	77.8%	71.4%	74.6%	82.5%	85.7%	87.3%	87.3%		
lilitye (E01)										
	l	li	lil	lili	lility	lilitye				
Queclarative	smS	**mQ, sc**	smQ, sc	smQ	smQ	smQ				mrt-
% Queclarative	31.7%	**55.6%**	85.7%	90.5%	87.3%	92.1%				
Statement	**mS**	smS, sc	smS	smS, sc	smS	smS				mrt-
% Statement	**50.8%**	74.6%	71.4%	87.3%	90.5%	93.7%				
ngamafu (F03)										
	ng	nga	ngam	ngama	ngamaf	ngamafu				
Queclarative	smS	**smQ,sc**	smQ	smQ	smQ	smQ				mrt-
% Queclarative	22.2%	**65.1%**	79.4%	88.9%	90.5%	93.7%				
Statement	**smS**	smS	smS	smS	smS	smS				mrt-
% Statement	**76.2%**	79.4%	74.6%	81%	90.5%	88.9%				
sisilo (G03)										
	s	si	sis	sisi	sisil	sisilo				
Queclarative	mS	**mQ, sc**	mQ	smQ	smQ, sc	smQ				mrt-
% Queclarative	36.5%	**55.6%**	57.1%	73%	88.9%	79.4%				
Statement	**smS**	smS	mS	smS, sc	smS	smS				mrt-
% Statement	**71.4%**	65.1%	60.3%	85.7%	81%	87.3%				
zizifo (H02)										
	z	zi	ziz	zizi	zizif	zizifo				
Queclarative	smS	**smQ,sc**	smQ	smQ	smQ	smQ				mrt-
% Queclarative	27%	**79.4%**	81%	90.5%	79.4%	87.3%				
Statement	**smS**	smS	smS, sc	smS, sc	smS	smS				mrt-
% Statement	**81%**	82.5%	66.7%	85.7%	87.3%	90.5%				
yintloko (I02)										
	y	yi	yin	yintl	yintlo	yintlok	yintloko			
Queclarative	mS	**smQ,sc**	smQ	smQ	smQ	smQ	smQ			mrt-
% Queclarative	46%	**68.3%**	74.6%	82.5%	90.5%	93.7%	88.9%			
Statement	**smS**	smS	smS, sc	smS, sc	smS	smS	smS			mrt-
% Statement	**68.3%**	82.5%	66.7%	79.4%	81%	82.5%	88.9%			

2 The bold typeface indicates the first gate for which a majority correct responses have been obtained.

Table B.17 Results of Perception Test B4[3] (continued).

ziinkomo (J03)										
	z	zii	ziin	ziink	ziinko	ziinkom	ziinkomo			
Queclarative	**mQ**	smQ, sc	smQ	smQ	smQ	smQ	smQ			mrt-
% Queclarative	**57.1%**	74.6%	82.5%	84.1%	82.5%	81%	84.1%			
Statement	smS	mQ, sc	mS	mQ	**smS, sc**	smS	smS			mrt-
% Statement	60.3%	46%	52.4%	44.4%	**85.7%**	92.1%	95.2%			
lulwimi (K02)										
	l	lu	lul	lulw	lulwi	lulwim	lulwimi			
Queclarative	**smQ**	smQ	smQ	smQ	smQ	smQ	smQ			mrt-
% Queclarative	**65.1%**	81%	77.8%	85.7%	87.3%	84.1%	81%			
Statement	**mS**	smS, sc	smS	smS, sc	smS	smS	smS			mrt-
% Statement	**55.6%**	82.5%	73%	88.9%	87.3%	87.3%	84.1%			
bubuhlanti (L02)										
	b	bu	bub	bubu	bubuhl	bubuhla	bubuhlan	bubuhlant	bubuhlanti	
Queclarative	**mQ**	mQ	smQ	smQ	smQ	smQ	smQ	smQ	smQ	mrt-
% Queclarative	**54%**	58.7%	63.5%	76.2%	79.4%	82.5%	90.5%	85.7%	82.5%	
Statement	mQ	mQ	mQ	**smS, sc**	smS, sc	smS, sc	smS, sc	smS, sc	smS	mrt-
% Statement	42.9%	36.5%	44.4%	**81%**	65.1%	93.7%	79.4%	96.8%	90.5%	
kukutya (M01)										
	k	ku	kuk	kuku	kukuty	kukutya				
Queclarative	**smQ**	mQ, sc	smQ	smQ	smQ	smQ				mrt-
% Queclarative	**74.6%**	55.6%	63.5%	71.4%	77.8%	82.5%				
Statement	**mS**	smS, sc	mQ, sc	smS, sc	smS	smS				mrt-
% Statement	**54%**	76.2%	46%	79.4%	84.1%	84.1%				

3 The bold typeface indicates the first gate for which a majority correct responses have been obtained.