THE VOICE QUALITY DISTINCTION IN DINKA SONGS

Exam number:

6320247

A dissertation submitted for the fulfilment of the degree of

Master of Science in Speech and Language Processing

The University of Edinburgh

Academic year 2009/2010

August 2010

Abstract

The purpose of this work is to study the distinction of the voice quality in Dinka songs, in particular to determine whether and how the phonemic distinction between modal and breathy vowel is conveyed in singing.

The methodology adopted consists in the application of two acoustic measurements that have already yielded interesting results on Dinka speech data, namely formant tracking and spectral regression. The outcome of these measurements clearly points towards a neutralisation of the distinction of voice quality in songs. The results of a perception experiment performed on human listeners show more evidence to support the conclusion that no difference in voice quality is conveyed in songs.

In the absence of any acoustic cue that could be used by the listeners to detect voice quality, further research is suggested. In particular, the presentation of the same perception experiment to Dinka native speakers could give crucial evidence on the nature of the acoustic cues that may or may not still be present in songs.

Declaration

I have read and understood The University of Edinburgh guidelines on Plagiarism and declare that this written dissertation is all my own work except when I indicate otherwise by proper use of quotes and references.

Acknowledgements

I wish to express my utmost gratitude to:

My supervisors, Bob Ladd and Bert Remijsen, for the support, interest and inspiration they have always given to my work.

The University of Edinburgh, that generously granted the means to pay the subjects of my perception experiment for their time.

Chris and Charitini, brothers in arms throughout this rewarding challenge, who have never been parsimonious in their friendship and moral support.

Sa, Michel, Mr X, Kat and Fed, who finally managed to perpetrate the deed of making me join them to find a new way in the UK.

Diletta, Franco and Nonna Ermanna, for being the best Famiglia that one person can possibly have.

Carla, who has made every single day I have spent in Edinburgh sunny and warm.

A tutti voi, grazie.

Table of Contents

Abstract	i
Declaration	ii
Acknowledgements	iii
Chapter 1. Motivations for this study	1
1.1 Research questions	1
1.2 The reasons for working on songs	2
1.3 Hypotheses	3
1.4 Structure of this work	4
Chapter 2. Background	5
2.1 Voice quality	5
2.1.1 Measuring voice quality: state of the art	8
2.2 Dinka language: typological and geographical information	12
2.2.1 The sound system of Dinka	14
2.2.2 Dinka songs	17
2.3 Preliminary analysis of spectrograms	18
Chapter 3. Methodology	22
3.1 Acoustic analysis	22
3.1.2 Data set	22
3.1.3 Selecting criteria	23
3.1.4 Code scheming	24
3.1.5 Praat and Praat scripts	24
3.1.6 Analysis of the results	25
3.2 Experiment design	26
3.2.1 Aims	26
3.2.2 Stimuli	26
3.2.3 Listeners and procedure	26
3.2.4 Structure	27
Chapter 4. Analysis of the results	30
4.1 Results of the acoustic measurements	30
4.1.1 Formant tracking	30
4.1.2 Spectral regression	33
4.1.3 Discussion of the results of the acoustic analysis	34

4.2 Results of the experiment 34

Chapter 5. Conclusion 37

References 41

Chapter 1.
Motivations for this study

Dinka has a rich suprasegmental system which includes phonological contrasts involving tone, vowel length and voice quality. While tone and vowel quality can be effectively measured considering their acoustic correlates, respectively fundamental frequency and formant values, this is not possible for voice quality. A variety of methods to measure it have been proposed in the literature and these measurements are all connected with the analysis of spectral tilt, that is to say, with aspects connected to the slope of the spectrum of a vowel. However, so far none of them has proved entirely satisfactory.

The main aim of this study is to determine whether and how the voice quality distinction is conveyed in Dinka songs. The wider context for this work is the research unit on Nilotic Prosody hosted by the Department of Linguistics and English Language of the University of Edinburgh, in particular the "Metre and Melody in Dinka Speech and Songs" project, which was recipient of a Large Grant as part of the “Beyond the Text” strategic programme coordinated by the Arts and Humanities Research Council (AHRC). One of the goals of the project is to take some steps towards an understanding of how the non-textual suprasegmental properties of Dinka language are conveyed in different communication formats (speech, song, writing), and how this mediation is influenced. In this context, the study of Dinka songs is considered particularly interesting as a part of the overall exploration of the three suprasegmental dimensions of the language.

Research on voice quality in Dinka speech data is being carried out by Timothy Mills and Bert Remijsen (unpublished data), using a series of measurements taken from the literature and implemented in Praat scripts as their main tools. These scripts and the results obtained so far will be an important reference for this study, which is also aimed at comparing how voice quality is conveyed in songs in respect to speech.

The data used in this study will be drawn from the extensive database of the "Metre and Melody in Dinka Speech and Song” project, which includes several hours of annotated recordings from a variety of Dinka native speakers and singers.

1.1 Research questions

Is the phonemic distinction between modal and breathy vowel conveyed in Dinka song? This is the main research question at the basis of this study. If the answer to this question is positive, a variety of other research questions can be raised: is the difference in voice quality conveyed in the same way it is in speech? Are the measurements used to determine voice quality as effective for the song data as for the speech data?

A crucial point in this dissertation is therefore the acoustic measurement of the difference in voice quality in Dinka songs in order to determine whether phonemic distinctions are also realised in singing and, eventually, how.

This project provides also an interesting chance to compare the validity of the different proposed measures of voice quality, testing them in a fresh perspective. The interpretation of the data collected involves a comparison between the outcomes of the similar studies applied to speech data, and this will probably shed more light on the effectiveness of some of the mostly often used spectral measurements.

A problem deeply intertwined with the search for a valid measurement is the definition of which acoustic correlates characterise and distinguish voice quality. This dissertation could contribute to this discussion as well, trying to asses directly the problem of voice quality in a phonetic perspective. This research would also consider the contributions of studies related to the variety of disciplines involved in the study of voice quality, from biomedicine to musicology.

1.2. The reasons for working on songs

Working with songs raises a number of peculiar side questions that must be taken into consideration in this study. The main issue is the dualism between melodic and phonological constraints, which is a typical phenomenon in songs, especially in the cases of languages with a rich suprasegmental system (Wong & Diehl, 2002). In these contexts, a reduction of the importance of the phonological constraints has been often observed in favour of compromising solutions which could work effectively under the rules of music as well. This is particularly interesting for tone languages, where the tone dimension expected in free speech inevitably clashes with the melodic structure of songs. It is therefore worthwhile to check whether other suprasegmental features such as vowel length and voice quality are conveyed at all; if they are, then it will be useful to investigate how these features are mediated when they are subordinated to the rules of music.

To sum up, a comparative approach studying prosody over speech and songs gives an innovative perspective on the phonetic study of suprasegmental systems: so far songs have been rarely studied from the point of view of acoustic phonetics, for the main focus of the studies involving the acoustic analysis of songs has been usually limited to the study of music (especially in the subfield of opera singing) or to the study of pathological speech (Titze 2008). In this sense the work of the "Metre and Melody in Dinka Speech and Songs" project is ground breaking.

1.3 Hypotheses

The principal hypothesis is that the difference in voice quality, which is phonemically distinctive in Dinka, should be also realised in songs. If the main hypothesis is confirmed, then further analysis will be carried out to determine whether the difference in voice quality is realised in the same way or in a somewhat different way.

In contrast, the null hypothesis is that the difference in voice quality is neutralised in songs. In this case, it will be worthwhile to determine whether the production favours one type of phonation, allegedly the neutral modal quality, rather than the other or if it simply yields a casual distribution among the two modes of voice quality.

1.4 Structure of this work

This dissertation will be organized in five main chapters, including the present introductory chapter, which is devoted to the presentation of the problem together with the research questions and the motivations that have driven this study.

The second chapter will deal with the background at the basis of this work. A first section will present the concept of voice quality and the problem of measuring voice quality, considering the contributions from a variety of perspectives, including the aerodynamic and physiological dimensions of the phenomenon. Afterwards, the controversial state of the art of the acoustic measurements of phonation types proposed in the literature will be presented, along with a review and critical comments on the most important contributions in defining effective methods for measuring voice quality. The second section of Chapter 2 will be then dedicated to a brief description of the sound system of Luanyjang Dinka, the local variety of Dinka mainly studied in this project, followed by a brief socio-cultural note about the nature of songs and the role that they play in Dinka culture. In the third part of the chapter, modal and breathy vowels will be informally compared analysing spectrograms generated from the data used in this study, both in speech and songs. This first analysis will be a preliminary step to have an impression on the acoustic reality of the voice quality distinction in Dinka, before the acoustic measurements and the perceptual experiment at the centre of this work are described.

The methodology adopted in this work will be described in the third chapter, which will present the automatic acoustic measurements used to determine voice quality differences together with the description of the data set, the respective selecting criteria and the software tools used. This section will also report the setup and the presentation of a perception experiment based on the evaluation of the voice quality distinction by human listeners, carried out in order to flank the data generated by the acoustic measurements with results obtained from a perceptual perspective.

The fourth chapter will present the results of the acoustic analysis, which will be plotted in graphs and discussed, using the available data on Dinka speech as reference materials to check for the validity of the measurements and their results. Similarly, the second section of Chapter 4 will present the outcome of the perception experiment, which will be also statistically analysed and discussed.

The last chapter will be devoted to the conclusions to be drawn from the results of this study. It will be also the starting point for a series of reflections on possible improvements or new directions for future research which could be based on the outcomes of this work.

Chapter 2.
Background

2.1 Voice quality

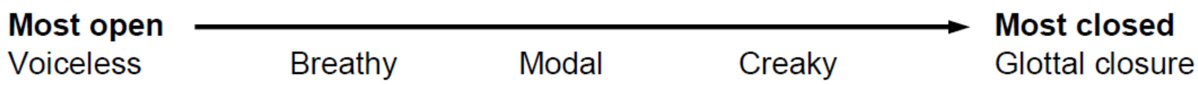
The concept of voice quality is closely related to that of phonation. In a classic study, Catford (1977) identified the two basic levels of any produced speech sound as initiation and articulation, but at the same time he acknowledged the existence of a third component, to be known as phonation. This was defined as “any laryngeal activity of speech that has neither initiatory nor articulatory function”, in the sense that the function of phonation is not involved in generating the air pressure that will be the initiation of a speech sound, and at the same time it cannot be considered as an articulator shaping the airflow into the final speech sound (Catford 1977).

The most typical phonetic phenomenon connected with phonation is voice or voicing, that is, the production of speech sounds involving the vibration of the vocal folds. Nevertheless, there is a variety of other types of phonation. From a physiological perspective, voice quality can be defined as a further characterisation of phonation, in which the quality of voice is affected by some changes in the state of the space between the vocal folds, known as the glottis. The majority of literature uses quite freely the expressions “type of phonation” and “voice quality” to express the same concept (Esling 2006); the interesting expression “manner of vibration” has been recently introduced to depict the same concept (Blankenship 2002).

By modal voice is meant the standard or neutral mode used as a term of comparison to describe contrastively the other modes of voice quality. It is hard to find a short and absolute definition for modal voice, being basically a negative concept, which deals more with what is missing than with what is realised; summarising, modal voice can be considered as the neutral or normal production of a vowel, realised when the vocal folds are pulled close together (adducted) and vibrating periodically so that air can flow through efficiently and without audible friction and no specific feature is added or changed (Laver 1980). Conversely, the expression “nonmodal” is an all encompassing term including any voice quality contrasting with modal (Berry 2001).

A variety of nonmodal phonation types are realised as controllable variations in the state of the glottis (Esling 2006bis): there is a certain amount of confusion in the terminology used to define the different kinds of voice quality, mainly because of the different perspectives from which it has been studied. Indeed, the study of voice quality spans many disciplines, from linguistics to medicine, and from physics to music theory and appreciation, resulting in a rich, although somewhat confused panorama that is well far from offering a unified description and terminology (Gerrat & Kreiman 2001). Many phonation types and voice qualities are described impressionistically from an auditory point of view: words like “harsh”, “hollow”, “murmur” are frequently used freely to name vocal phenomena without any agreement on what they actually mean, resulting in overlapping and often confusing descriptions.

Moreover, the uncertainties regarding the characterisation of voice quality do not reside only in terminology, but also in the physical description of how it is produced. The traditional vision was based on what is known as the phonation continuum hypothesis (Ladefoged 1971), which proposes the different types of phonation as a succession of different modes of vocal fold vibration on a line which ideally connects the most open state of the glottis to the most closed (or constricted) glottal state, as depicted in the figure below.



However, this model is only partially able to account for the vocal phenomena connected with voice quality. The advancements in technology and the application of aerodynamic measuring techniques to the field of acoustic phonetics have shown how phonation and voice quality do not only depend on the state of the glottis, but also on the action of the surrounding tissues and cartilages. In this sense the study published by Edmondson & Esling (2006) is emblematic: in their article the authors explain and document the complex manoeuvres necessary for phonation with the aid of colour pictures and videos obtained with the application of the techniques of modern fibreoptic laryngoscopy and the use of transnasal endoscopes with high-quality video recording capabilities and cold-light sources. What emerges is a far more complex range of states of the glottis and of connected muscular manoeuvres than what was encompassed by the mono-valve and mono-dimensional continuum model. This is especially true for languages with a complex prosodic system like Dinka, where the independent use of contrastive voice quality and tone requires extremely precise and controlled laryngeal manoeuvres, as both suprasegmental dimensions are “controlled and implemented by the same anatomical instruments, namely the larynx” (Avelino 2010).

Although most languages do not present any kind of phonetic difference in voice quality, a fair number of non-European languages use at least one phonation type among creakiness or breathiness to contrast phonologically with modal voice (Ladefoged &

Maddieson 1996). In the complex and evolving context regarding the study of voice quality, there seems to be agreement at least on the definition on these two most frequent and most studied non-modal voice qualities. Creaky voice, indicated with the diacritic [̰] and also known as laryngealised voice (Ladefoged 2006) or vocal fry (Laver 1980), is produced with "a low frequency periodic vibration of a small section of the vocal folds" (Catford 1964) and the correlate auditory effect is that of a series of taps, impressionistically reminding of "running a fingernail across the teeth of a comb" (Gussenhoven & Jacobs 2005). It is not infrequent to notice this voice quality at the end of falling intonations in some British speakers (Ladefoged 2006), in the response token "yeah" (Grivicic & Nilep 2004) or in other fillers expressing indecision or acknowledgement (Horne 2009). For the purpose of this study, it will be worthwhile to describe in more depth the voice quality of breathiness, as opposed to modal voice.

Breathy voice, represented with the IPA symbol [̤], is usually defined as the impression of a turbulent amount of airflow and of an audible escape of air through the glottis, due to a lessened muscular effort and the consequent imperfect closure of the glottis (Gerrat & Kreiman 2001). By comparison to modal quality, the modality of vibration of the vocal folds is therefore inefficient and accompanied by audible friction (Laver 1980) and the result of this partial closure is an h-like sound. Breathiness has therefore been impressionistically described as "voice mixed with breath" (Catford 1977), and, more colourfully, as the kind of voice used in European languages to convey confidentiality (Laver 1994) or even to create the jocular effect of a sensual voice (Gussenhoven & Jacobs 2005). This flexible use of breathiness to characterise personal voice quality has been considered as a proof of its status of continuous variable. This implies that there are various degrees of breathiness, as opposed to other types of phonation that vary categorically, in the sense that they form a coherent perceptual category where listeners can tell whether a phonation is realised or not (Gerrat & Kreiman 2001). Moreover, breathiness can also vary a great deal from one speaker to another, partly depending on the gender and on the unique physiological features of each individual (Hanson et al. 2001, Blankenship 2002, Ladefoged 1983). This variability is also what makes breathiness more difficult to identify for untrained listeners, especially in linguistic contexts where it has no phonological contrast (Kreiman & Gerrat 1997; 1999).

2.1.1 Measuring voice quality: state of the art

As already mentioned in the first chapter, the acoustic characterisation of voice quality is not straightforward. There is not a direct auditory correlate that accounts for it, as happens for example with pitch for fundamental frequency and with formant values and vowel quality. Voice quality is particularly difficult to distinguish also because at the same time it is a carrier of socio-cultural characteristics (such as the accent or the regional variety of the language used by the speaker) and extralinguistic idiosyncrasies, mainly anatomical, uniquely corresponding to single speakers, which can vary a great deal from an individual to another (Ladefoged & Broadbent 1957, Esling 2006, Avelino 2010). What's more, phonation types and their acoustic characteristics are highly affected by the gender of the speaker, as pointed out in many recent studies (Hanson 1997, Blankenship 2002, Simpson 2009).

Another serious issue to be considered is that voice quality is also very difficult to characterise cross-linguistically. As it will be presented in more detail in the following section of this chapter, Gordon & Ladefoged (2001) have found a number of similarities across many languages of the world and they have listed a series of phonetic properties that can be generally associated with voice quality across the languages, including, among the others, periodicity, overall intensity, formant frequencies and spectral tilt. However, the same authors admitted that these correlates can vary a great deal from language to language and a generalisation can have only an approximate value.

Finally, the elusive nature of voice quality is also influenced by its interaction with other suprasegmental dimensions such as intonation and tone, which complicates greatly any attempt to measure voice quality alone.

As mentioned before, in recent years voice quality has been studied from a variety of perspectives. In this section, the main measuring tools and approaches will be briefly presented as a background of the main objectives of this work: the acoustic and perceptual definition of the voice quality distinctions.

The physiological correlates of voice quality have been studied with increasing precision and detail thanks to the steady advancement in technology, especially in the field of biomedicine; this resulted in a redefinition of the valves and tissues involved in conveying voice quality, as reported in the cited study by Edmondson & Esling (2006). In addition, aerodynamic investigations, mainly aimed at determining the pressure and flow of air through the glottis, have recently profited from the adoption of non-invasive procedures that have successfully replaced older invasive techniques such as tracheal puncture (Ladefoged 2003).

Electroglottography (EGG) is one of the techniques that have been more frequently used in this sense: using this method, observations of the characteristics of the vibration of the vocal folds are performed by measuring the electrical resistance/conductance between two electrodes located on the neck of a subject in proximity of the larynx (Avelino 2010).

If the study of the kinematics and of the physiological correlates of voice quality is a thriving field, it is still rather difficult to obtain a completely satisfactory procedure to determine voice quality in terms of acoustic measurements. Although an increasing number of researchers have started to seek for reliable ways to measure voice quality and therefore to single out the acoustic correlates and characteristics which could phonetically define voice quality, no definitive results have been provided yet. In addition, the acoustic measurement of voice quality has been rarely the main or single focus of in-depth studies. Indeed, the most influential studies and methods applied to the measurement of voice quality have often been directed to the definition of other phonological, prosodic or even pathological phenomena and rarely aimed at the definition of an effective way to determine acoustically the types of phonation.

In this section a brief literature review will present the main methods proposed, with a specific attention on the ones implemented by Mills in his Praat scripts and applied by the same author together with Remijsen (Mills & Remijsen, unpublished work) to the analysis of Dinka speech data. Two of these methods will be also adopted in this project to measure phonetically the song data in this project (see Chapter 3).

A certain amount of acoustic information can be obtained from the impressionistic observation of spectrograms and waveforms, as performed in works like Gordon & Ladefoged (2001) and Ladefoged (2003) and as it will be done here in the next section of this chapter. With this kind of approach variations such as the intensity of the formants or the presence of aperiodic bursts of noise should be immediately noticeable. However, a more efficient way to measure fine-grained differences in phonation and voice quality is the analysis on the FFT spectrum of the vowel sounds, which is the base of all the acoustic measurements proposed in the literature. Also in this case, trends can already be observed impressionistically when one analyses the spectral contour of vowels and compares it among the different voice quality realisations. However, a deeper inspection of the spectra is possible: the patterns connecting the spectral values can be implemented as functions, so that numeric values can be analysed in depth and transferred into mathematical terms. The analysis of the spectra is in the backbone of the spectral measurements and the acoustic models proposed in the literature. For example, as it will be shown below in this section, the relatively higher energy in the fundamental frequency (or first harmonic) that can be observed in a breathy vowel is used as a starting point to measure the amount of breathiness. This is possible performing a comparison of the amount of energy produced in that frequency with the rest of the spectrum, measuring what is called the "spectral tilt" of the vowel.

The expression spectral tilt refers to "the overall distribution of energy in the spectrum" or to "the rate at which intensity falls off as frequency increases" (Mills 2009). Spectral tilt is involved in the perception and production of lexical stress (Sluijter & Van Heuven 1996, Remijsen 2001), but it can also account for information related to phonation in general and voice quality in particular: for example, differences between the modal and breathy types of phonation are recorded in the spectral tilt, because a relatively more high-frequency energy is clearly produced by a breathy excitation than by modal voice (Hanson 1997, Mills 2003). Moreover, spectral tilt has been associated with a number of physiological features, correlating for example with the degree of openness of the glottis (Gordon & Ladefoged 2001). However, it has to be borne in mind that the differences in phonation are not the only factors affecting the relative intensities of spectral frequencies: phonetic variables, such as vowel quality, and other prosodic phenomena, like variations in pitch and lexical stress, may as well interact with the energy distribution in the spectrum and therefore with the measures of the difference in phonation types (Ladefoged 2003, Mills 2009).

Two widely adopted methods for measuring voice quality and, in particular, to compare breathy and modal voice are presented in Ladefoged (2003) and Gordon & Ladefoged (2001). Both these measures are aimed at determining the relative amplitude of the first harmonic (H1): the first technique consists in a comparison with the second harmonic (H1- H2), while the second consists in a more spaced out difference along the spectral slope, comparing the first harmonic and the third harmonic (H1-H3). These methods have already been successfully applied to Luanyjang Dinka speech data by Remijsen & Manyang (2009) and they will be discussed in more detail in the second section of this chapter.

In the study cited above, Sluijter and van Heuven (1996) propose a measurement called "spectral balance", defined as the intensity in critical contiguous "spectral bands" or "energy bands". This method consists of the identification of four frequency bands and the subsequent measurement of the average energy within the range of each of these bands. A univariate analysis of the single bands follows and the results are compared to determine what areas of the spectrum are affected by lexical stress. Although easily computable and implementable as a voice quality measure, this method has two main disadvantages: it interacts with the formant values that are too close to the boundaries of the spectral bands and it is specifically effective on the analysis of single vowel qualities, therefore being unsuitable in a cross-vowel perspective (Mills 2009).

Another spectral tilt measurement proposed in the literature is known as "spectral regression" or "regression lines" (Kochanski et al. 2005). After converting both frequency and power into Bark scale, a regression line is fit to the spectrum in order to obtain a function from which values of "slope" and "intercept" can be extracted. Although this measure proves a good tool to straightforwardly compare values for different vowel qualities, the slope and intercept values do not represent clearly the spectral properties of the source and are affected by variations in the supraglottal filter. The glottal properties specifically involved in phonation are therefore downplayed.

In response to this kind of problem, Hanson (1997, Hanson et al. 2001) based her method on an acoustic model where the supra-laryngeal influences can be ruled out and the effect of formant variation reduced, with the aim of effectively measuring glottal events at source level. The method consists in comparing the amplitudes of the first harmonic and the energy produced by the third formant (H1-A3). These values are corrected for the boosting effects of the first and second formants with the application of a series of equations, reported in Hanson (1997) and Mills (2009), and the impact of nearby formants is therefore reduced. This stratagem solves one of the major problems affecting the spectral bands measurement by Sluijter and Van Heuven (1996) and the spectral regression proposed by Kochanski et al. (2005). The effectiveness of the application of this technique to the measurement of voice quality depends necessarily on the measurement of H1, which is not always possible, as for example in whispered voice (Mills 2003); nevertheless, in a study concerned with other types of phonation it proves a very reliable method of measuring events at the level of phonation.

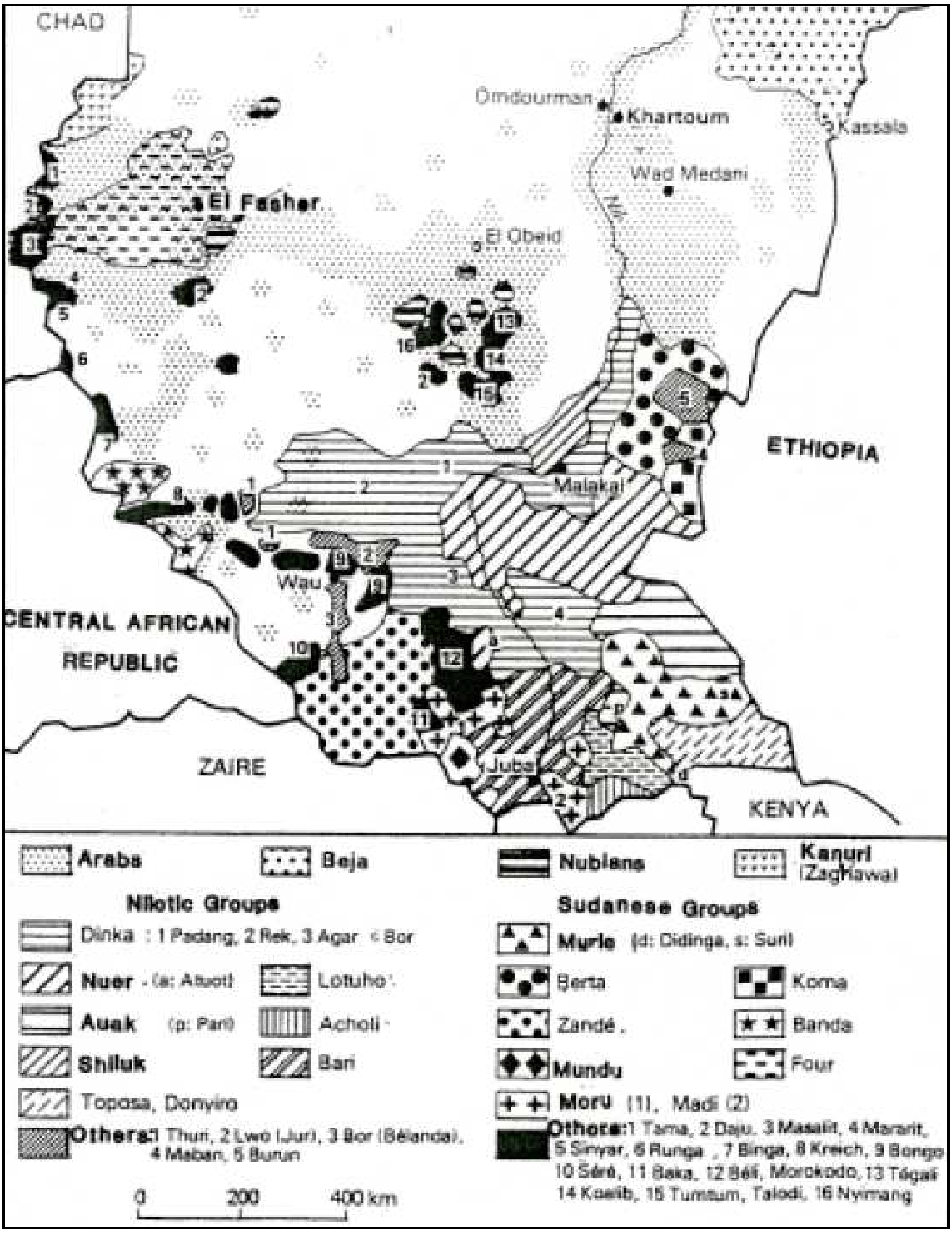
Another complex model, based on the mathematic modelling proposed by Fant (1960), is used by Fulop et al. (1998), who take into account a variety of parameters, such as the values of the first three formants (computed individually), the resting formants (calculated altogether using a catch-all formula), the source spectrum and the radiation of the sound from the mouth. The final spectral tilt measure corresponds to the comparison of the peak amplitudes of the first and the second formants (A1-A2). The explicit way in which this method accounts for formants, and source spectrum, makes it a very effective tool to describe clearly glottal and supraglottal events, also in a cross-vowel perspective.

One last spectral tilt measure applied to the study of voice quality is the one proposed by Heldner (2001; 2003) and known as "spectral emphasis". Heldner defines this expression as the relative intensity of sound in the higher-frequency bands of a spectrum, as opposed to the overall intensity of the whole spectrum. The author contrasts it to the expression “spectral tilt”, which should be reserved “for measures explicitly representing the slope” (Heldner 2003); in fact, spectral tilt and spectral emphasis can be considered in a relation of opposition, in the sense that increasing spectral emphasis corresponds to decreasing spectral tilt. Several spectral measures are proposed, although the main one consists in calculating the difference in dB between the overall intensity and the intensity of a low-pass-filtered signal. One advantage of this method lies in the fact that the dynamic filtering technique reduces the impact of formants, making this method partially insensitive to variations in formant frequencies (Mills 2009).

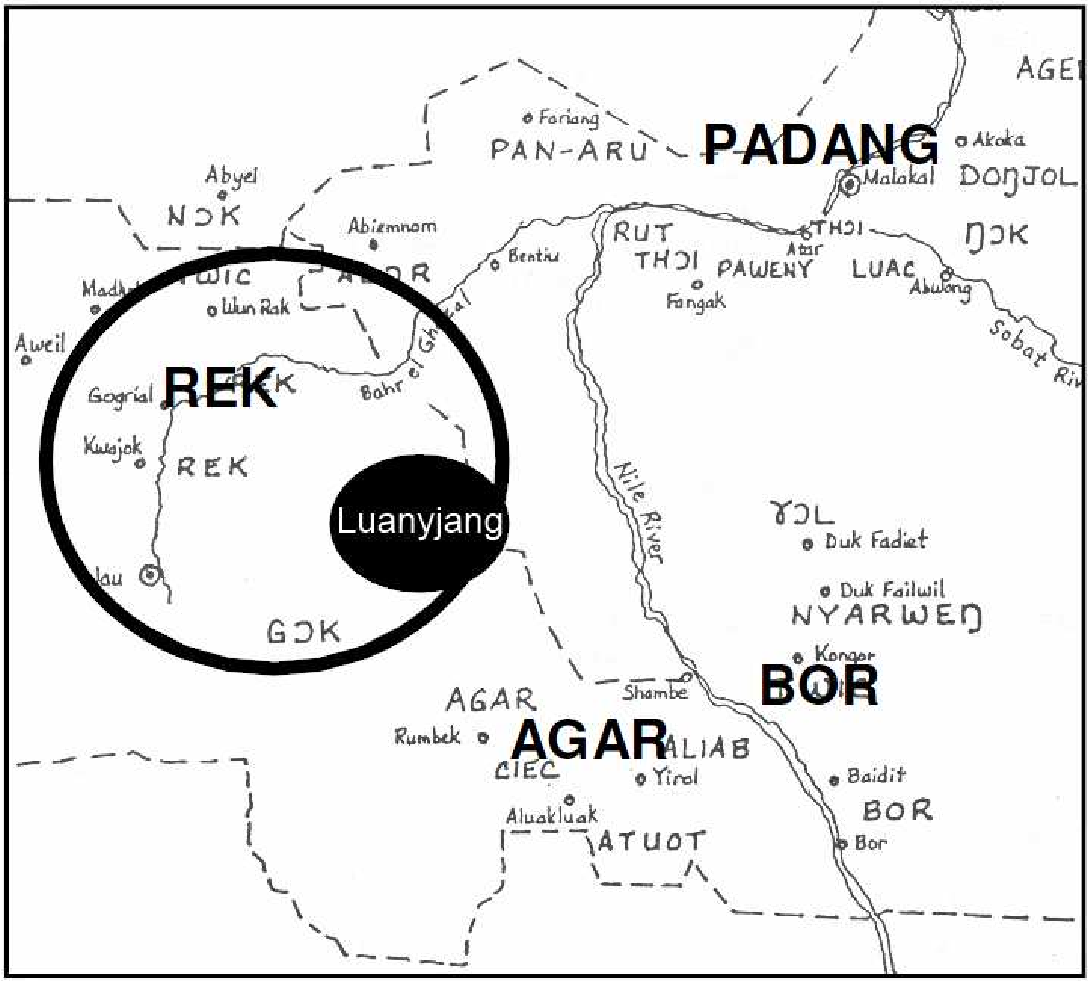
All these methods have been applied with more or less success to the measurement of voice quality, and they all have been implemented as Praat scripts by Mills. However, it has to be borne in mind that one of the most effective and widely used acoustic measurements for voice quality, which was also implemented in Mill’s scripts, is tracking the formant values and analysing the relative energy distribution at low and high frequencies of the spectra. In contrast with some of the methods presented here, this measurement focuses on the supraglottal filter properties, downplaying the role of the events at the source (i.e. glottal) level. In a general study aimed at measuring voice quality as a whole, this quick and easy procedure still proves meaningful and valid.

2.2 Dinka language: typological and geographical information

Dinka is a Western Nilotic language, belonging to the Nilo-Saharan family. It is spoken in extensive areas of Southern Sudan by over two million speakers (Lewis 2009). The Dinka people are concentrated in the area of the White Nile and its tributaries, but there are also smaller cultural enclaves in urban areas in the rest of Sudan, in particular in Khartoum, and abroad. Dinka is differentiated in four main dialect areas: Padang, Rek, Agar and Bor. The local variety described in this section is part of the South-Western dialect group called Rek and is known as Luanyjang or Luac (as in Roettger & Roettger 1989). It is spoken by at least 15.000 speakers (Lewis 2009), the majority of whom live in the town of Wuncuei, located approximately 170 km East of the city of Wau (in the State of West Bahr al Ghazal), and in the surrounding area. The maps below were adapted respectively from Abu-Manga & Miller (1991, see Idris 2004) and Roettger & Roettger (1989). Although they might be somewhat outdated, they give a basic idea of the distribution of the four major Dinka dialects within the wider linguistic context of Southern Sudan (Map 1) and of Luanyjang Dinka within the Rek dialect area (Map 2).



Map 1



Map 2

In recent years, Dinka has been studied with a growing interest mainly for its unusually rich suprasegmental system, which presents independent phonological distinctions of tone, vowel length and voice quality. Starting from the research published by Andersen (1987) on Agar Dinka and by Malou (1988), specifically on the Dinka vowel system, studies on the phonology of Dinka have been carried out by an increasing number of scholars. In particular, most of the current members of the “Metre and Melody in Dinka Speech and Songs” project have focused their research on Luanyjang Dinka in several recently published papers and articles (Remijsen & Gilley 2008, Remijsen & Manyang 2009, Ladd et al. 2009, Reid 2009).

2.1.2 The sound system of Dinka

The description of the phonology of Luanyjang Dinka which follows is mainly based on the extensive article published by Remijsen & Mayang (2009) for the Illustrations of the

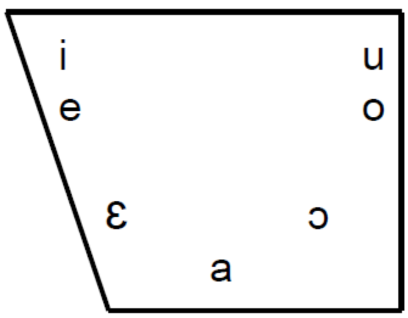
IPA. For the purpose of this study, the phonology of Dinka will be briefly presented with a special focus on the phonological value of voice quality.

From the morpho-phonological point of view, Dinka words are mainly based on monosyllabic stems bearing on a consonant-vowel-consonant (CVC) structure, where the distinctive power of the vowels is enhanced by the three independent suprasegmental dimensions of vowel length, tone and voice quality, which all carry both lexical and morphological functions.

Luanyjang Dinka presents a consonant system consisting of a set of 20 phonemes appearing in five different places of articulation, featuring voiced and voiceless stops and nasals. No fricative phonemes are present.

	Labial	Dental	Alveolar	Palatal	Velar
Plosive	p b	t̪ d̪	t d	c ɟ	k g
Nasal	m	n̪	n	ɲ	ŋ
Trill			r		
Approximant	w			j	ɰ
Lateral approximant			l		

For the purpose of this study, it is worthwhile to focus in more detail on the vowel system. This is structured in seven different vowel qualities, thoroughly described in their distribution by Remijsen and Gilley (2008), which can be visualised in the vowel chart reproduced below.



A rather peculiar characteristic of Dinka is the presence of three levels of vowel length: short (/V/), medium (/VV/) or long (/VVV/), corresponding to average durations of 70, 100 and 150 milliseconds respectively (Remijsen & Gilley 2008). This peculiarity is shared by only a few other languages in the world, some of which belong to the same Western

Nilotic family, such as Shilluk (Remijsen 2008, Reid 2009) and Thok Reel (Reid, forthcoming). Further acoustic measurements have shown how short vowels tend to be centralised towards a schwa sound, as a result of the limited time to be realised (Remijsen & Manyang 2009). In this regard, it is interesting to point out the non-occurrence of the front low mid vowel /ɛ/ in the shortest level of vowel length, resulting in a merge with short /a/. A similar blurring over clear vowel quality distinctions can be noted also in the frequent cases of diphthongs and triphthongs, where the vowel qualities tend to interact and assimilate one with the other (Andersen 1987).

The tone system of Luanyjang Dinka is particularly rich, even compared to other dialectal varieties of Dinka. Here the four tonemes Low, High, Rise and Fall are present, while the tone repertoires of other dialects are not so complex: Agar Dinka, for example, presents only three tonemes, with the absence of the Rise pattern (Andersen 1987). No interaction between tone and other suprasegmental features is detectable, although the more complex tone patterns usually occur in the presence of syllables presenting medium or long vowel lengths (Remijsen & Manyang 2009).

Dinka vowels are also characterised by two phonemically distinctive voice qualities, modal and breathy, with the exception of the back high vowel /u/, which is always breathy. The binary opposition between modal and breathy phonation types is particularly neat in Luanyjang Dinka, where no other type of phonation is produced. In contrast, other local varieties of Dinka present an even more complex phonetic reality. For instance, Denning (1989) has detected up to four distinct voice qualities (modal, breathy, creaky and hollow) in the Bor Dinka dialect and Andersen (1987) describes the phonological distinction of voice quality in the Agar Dinka dialect as an opposition between creakiness (instead of modal voice) and breathiness. However, it is clear that, across the many local varieties of Dinka, the only distinction in voice quality that carries phonological value remains the one between breathy and non-breathy, while the other types of phonation can only be detected phonetically, with no variation in the meaning of the words they are in; phonologically, these supplementary voice qualities represent allophones of the modal voice quality.

The phonological distinction between modal and breathy vowels can be demonstrated with series of minimal pairs, as reported in the examples below, transcribed according to the conventions adopted by Remijsen & Manyang (2009) to account for the longest level of vowel length (/VVV/) and used the IPA diacritics to signal breathiness:

kiiir (big river)	**ki̤iir** (kind of tree)
leeel (challenge)	**le̤eel** (small hoe)
bɛɛɛr (regurgitate)	**bɛ̤ɛɛr** (answer)
maaan (hate)	**ma̤aan** (harelip)
rooor (forest)	**ro̤oor** (men)
tɔɔɔŋ (spear)	**tɔ̤ɔɔŋ** (say goodbye)

Exploratory phonetic measurements have been successfully applied to Luanyjang Dinka speech data, as reported and illustrated in Remijsen & Manyang (2009). The results suggest that breathy vowels in Dinka have lower F1 values, in accordance with previous studies (Malou 1988, Edmondson & Esling 2006), and that breathy vowels present more energy at higher frequencies than at lower frequencies related to their modal counterparts. These phenomena are clearly visible in the variations of the spectral slope, in particular in the difference between the values of the first two harmonics (H1-H2), which appear very close to each other in the upper part of the spectrum, or in the difference between the first harmonic and the energy produced by F2 (H1-A2), as shown in Ladefoged (2003). In the case of modal voice the second harmonic is sizably higher than H1, resulting in a steep rise in intensity, while the H2-A2 function presents a fairly level slope. In contrast, breathy voice presents a more level slope between the first and the second harmonics, which are almost at the same level, while the H2-A2 function presents a steeper slope than in the modal counterpart.

2.2.2 Dinka songs

In the Dinka socio-economical reality, where literacy is incomplete and the lifestyle is highly dependent on agriculture and cattle raising and trading, songs play a very important role. From the point of view of the contents, they are carriers of cultural memory, both at a private and at a collective level. The themes of the songs range from the celebration of the quality and quantity of cattle belonging to the singers or to their family, to the chronicle of family micro-histories and to macro-historical events such as episodes of the so-called Second Sudanese Civil War (1983-2005) or other events of public interest, such as the recent death of the former leader of the Sudan People's Liberation Army/Movement (SPLA/M) John Garang. One peculiar characteristic regarding Dinka songs is the concept of ownership: Dinka people can sing songs they have composed themselves or they can actually commission and buy songs from other composers, who create and sell songs composed *ad hoc* on the topics requested by the buyer. From the moment of the end of this kind of transaction the buyer becomes the owner of the song. The processes of composing, singing and trading songs are therefore vibrant components of the Dinka culture, and they find their ideal context in cattle camps, where the business transactions are sided by all kind of trading activities, including buying and selling songs. As reported in the corresponding metadata files (see Chapter 3), many of the songs that are stored in the database of the “Metre and Melody in Dinka Speech and Songs” project and that have been used in this study were recorded in some of these cattle camps.

A discussion about the musical and melodic features of the songs is beyond the scope of this study, so here it will be enough to say that Dinka songs are sung in unison and they are based on a regular texture of connected song segments following a regular pulse, where the usual accompanying sounds are limited to the clapping of the hands or to the sound of a drum. This plain style is in contrast with the far more complex singing traditions typical of other Sub-Saharan languages, usually based on multi-part vocals and instrumental rhythmic interactions (Remijsen 2010) and makes Dinka songs particularly suitable for a phonetic and phonological study comparing them to speech data.

As mentioned in the introductory section of this work, the importance of studying songs in a language like Dinka lies in the fact that it represents another phonetic environment where to test theories and hypothesis on the suprasegmental system. In the song environment the three dimensions of tone, vowel length and voice quality compete with the rules and constraints of melody (Wong & Diehl 2002), and the study of these interaction beyond speech could lead to interesting results that can contribute to explain, quantify and describe prosodic phenomena which so far have not been satisfactorily or completely understood in their complexity.

2.3. Preliminary analysis of spectrograms

Before dealing with the description of the methodology adopted to measure acoustically the voice quality distinction in Dinka songs and to evaluate experimentally its perceptual reception by human listeners, it is interesting to carry out a preliminary, informal, analysis of the data. In this section, the inspection of spectrograms showing the difference between modal and breathy voice and generated from the speech and song data sets used in this project, will give some useful information which will be used as a background for the deeper acoustic analyses presented in the chapters to follow. This will be also the ideal starting point to get a first idea on the acoustic differences between the two types of phonation across the two channels of speech and song.

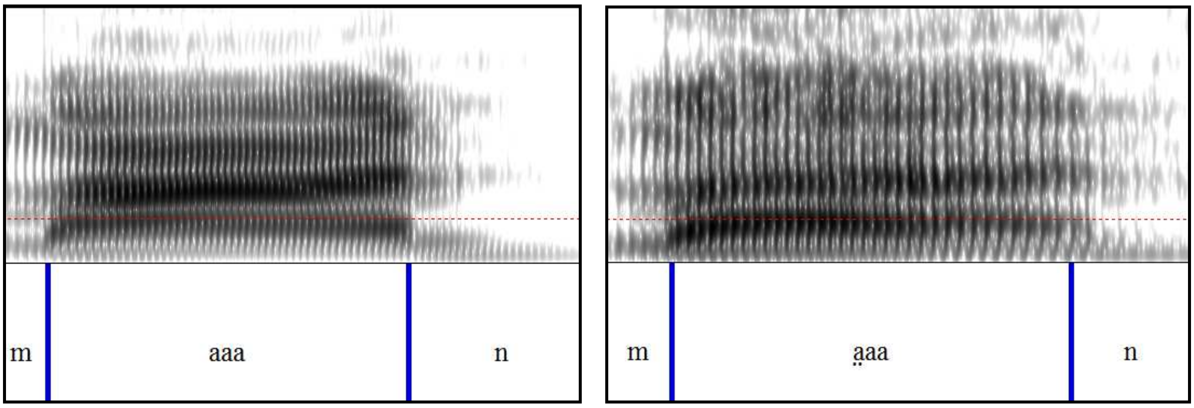
Several scholars have adopted this kind of informal analytical approach (Gordon and Ladefoged 2001, Ladefoged 2003, Avelino 2010) as a preliminary step before going in-depth with the acoustic analysis, although this is the first time that this has been done specifically on Dinka data, not only for songs, but also for speech.

Here only two examples, one for speech data and one for songs, are presented visually, although the discussion is about trends and properties that can be found across all the data used in this work. The data used in this preliminary analysis of speech and song data were all drawn from the data set extracted from the database of the "Metre and Melody in Dinka speech and songs" project and used in the perceptual experiment that is described in Chapter 3.

The reference study for this informal analysis is the already mentioned cross-linguistic overview on voice quality distinctions carried out by Gordon & Ladefoged (2001). In this work, the authors presented a set of the phonetic properties usually associated with the different nonmodal phonation types across languages, namely: spectral tilt, periodicity, acoustic intensity, formant frequencies, fundamental frequency and duration. Spectral tilt has already been explained in detail in the previous section, also in respect with Dinka language, and it is therefore excluded from this brief overview; as for the other general features identified by the authors of the article, some of these describe well voice quality distinctions in Luanyjang Dinka, while others are definitely less relevant. The model results even more difficult to apply to song data, where some proposed correlates are immediately ruled out by the nature of singing *per se*. However, it is interesting to check the Dinka data against these guidelines in order to obtain a first, general background for a better understanding and contextualisation of the results of the acoustic analysis and perception experiment at the centre of this work.

The speech data that were analysed consisted in a set of spectrograms generated from the recordings of monosyllabic minimal pairs that show the phonemic opposition between modal and breathy vowels. In order to test the efficiency of the set of phonetic properties defined by Gordon & Ladefoged (2001) in a language with a complex suprasegmental system like Dinka, it was necessary to consider words that matched with each other along the double dimension of tone and vowel length, otherwise the respective acoustic correlates of fundamental frequency and duration would have been impossible to be compared.

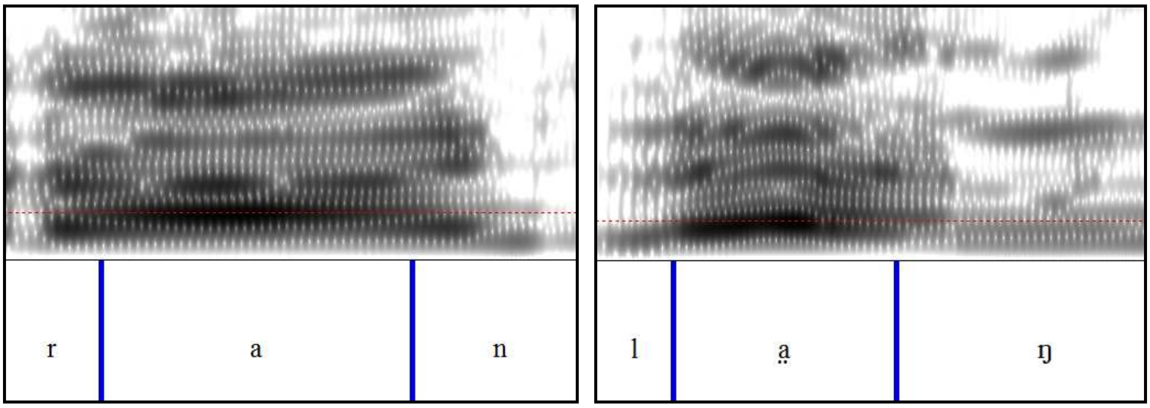
One of the first phonetic properties listed in Gordon & Ladefoged's article is periodicity. The authors note how breathy realisations are characterised by a less periodical noise, clearly visible in the corresponding spectrograms, where an increased spectral noise can be observed especially at high frequencies, reflecting the aspiration or turbulence that is present in breathy voice. This happens clearly also in Dinka spoken data, as it can be seen in the figure below, where the spectrogram of the breathy vowel /a/, extracted from the word /maaan/ (tobacco) shows a sizable difference in terms of noise in respect to its modal counterpart present in the word /m̤aaan/ (men).



The same significance could not be observed for the phonetic property of overall acoustic intensity. In our examples from Luanyjang Dinka, the expected decrease in overall intensity for breathy vowels was only partly visible and certainly not significant in itself to be detected as a distinguishing feature, neither in speech nor in song data. The already commented figure shows well how significant differences cannot be detected at first sight.

The properties of fundamental frequency and duration could be compared only in speech data, because in songs it would be necessary to find vowel tokens presenting the same tonality, note and length. Even with enough material, finding the right samples that could match with each other would be a virtually impossible work, which would lie well beyond the scopes of a phonetic analysis and which would belong more to the field of musicology. Interestingly enough, the expected differences in fundamental frequency and duration did not result particularly relevant even in speech. As for duration, in particular, the speech data showed no difference between modal and breathy vowels. The latter were expected to be relatively longer than their modal counterparts, but this could not be verified in any analysed sample.

As for formant frequencies, the expected results should show a lowering of the first formant and a higher relative intensity at higher frequencies for breathy vowels, in line with the trends that have already been tracked and reported for Luanyjang Dinka by Remijsen & Manyang (2009). Nevertheless, the data analysed in this work only showed the expected lowering of F1 values, while the energy distribution at higher frequencies did not present any sizable difference. This was true both for speech and song data (see figures below, reporting the CVC syllables /ran/ and /la̤ŋ).



One last interesting feature highlighted in Gordon & Ladefoged (2001), although not explicitly listed in their series of phonetic features, is the issue of timing in nonmodal phonation: in particular, the authors report how this kind of phonation is usually confined to a limited portion of the vowels. Nevertheless, this interesting characteristic could not be detected in any of the examples from Dinka, both in speech and in songs.

Chapter 3.
Methodology

3.1. Acoustic analysis

3.1.2 Data set

One of the aims of the “Metre and Melody in Dinka Speech and Song” project is the creation and maintenance of a constantly growing database of annotated speech and song data. The phonetic data are mainly collected in fieldwork sessions in Sudan by Bert Remijsen and other members of the project team, who include Dinka native speakers living in Southern Sudan and trained to record and annotate the phonetic data after collecting them in the field. In particular, the song database includes digital audio recordings obtained from several male and female singers, often recorded in cattle camps, where the songs are often performed and exchanged as a side activity accompanying the main business of cattle trading (see Chapter 2).

The directories hosting the Dinka songs are accessible on-line on the web pages of the Nilotic Prosody research unit and they include the audio recordings along with a variety of supplementary information. The first sources of extra information provided are the annotation tiers, encoded in the format of Praat TextGrid files, which include up to five levels of information: transcription in Dinka orthography, English translation, phonetic transcription in IPA, morpheme gloss and commentary on the topics of the song when necessary. The annotations are constantly updated, mainly thanks to the collaboration of the trained native Dinka speakers collaborating to the project.

Another useful source of information provided in the project directories is represented by the metadata files. These documents include valuable information about the linguistic consultants (personal data, such as name, age, birthplace, home town etc.), their linguistic background (what other languages they speak and how often and for what purposes they speak Dinka) and their dialectal variety of Dinka. In addition, the metadata files present information about the songs, such as genre and accompanying instruments (if there are any), the summary of the thematic contents and the ownership status (self-composed and owned by the singer, only owned by the singer, that is, commissioned and bought from another composer, or simply borrowed from another person) of each song. The piece of information which is certainly most useful for the aims of this study is the regional variety of Dinka used in the song and the place where the song was recorded.

For the purpose of this work, the database offers enough annotated song data to collect a satisfactory amount of data for six different singers, four of whom are men, two women. Four of the six singers presented a dialectal variety belonging to the Rek group, which is the regional variety connected with Luanyjang Dinka, while the remaining two used the North-Eastern Dinka dialect known as Ageer.

3.1.3 Selecting criteria

The back high vowel /u/, as stated in the previous chapter, is always breathy (Remijsen & Manyang, 2009) and it is therefore left out from this study. Four tokens of the remaining six vowels, which can all be realised with modal and breathy voice, were collected. The data set therefore resulted in a total of 48 tokens per singer.

The main selecting criterion was the perceptual clarity of the sound; the phonological constraints, which are frequently violated in favour of the rules of melody (Wong & Diehl 2002), were considered only as a secondary factor. For example, if a sound that was expected to be a long vowel was realised as a short vowel in order to fit with the melody, the corresponding realisation was not selected. In this strictly phonetic perspective, the occurrences of vowels within repetitions of the same words, a typical phenomenon in Dinka songs, were considered as different actual realizations and therefore taken into account as different valid tokens to be selected and processed in the analysis. The duration of the samples ranged from 60 to 100 ms; this was considered an acceptable time span, on the one hand, for the scripts to set a window which can process a significant amount of voice signal at the midpoint and, on the other hand, to give the researcher the chance to adjust some parameters in case the script failed to provide completely satisfactory results (see section 3.1.5).

As mentioned above, there is often unclear distinction between short vowels in Dinka, which tend to centralise and to blur their phonological boundaries in their actual realisations, and similar complications arise when the short and long vowels are part of diphthongs (Remijsen & Manyang 2009, Andersen 1987). Considering these limits in phonetic clarity and distinction, the samples were collected only from medium and long vowel realizations. Short vowels or vowels produced within diphthongs were not used.

3.1.4 Code scheming

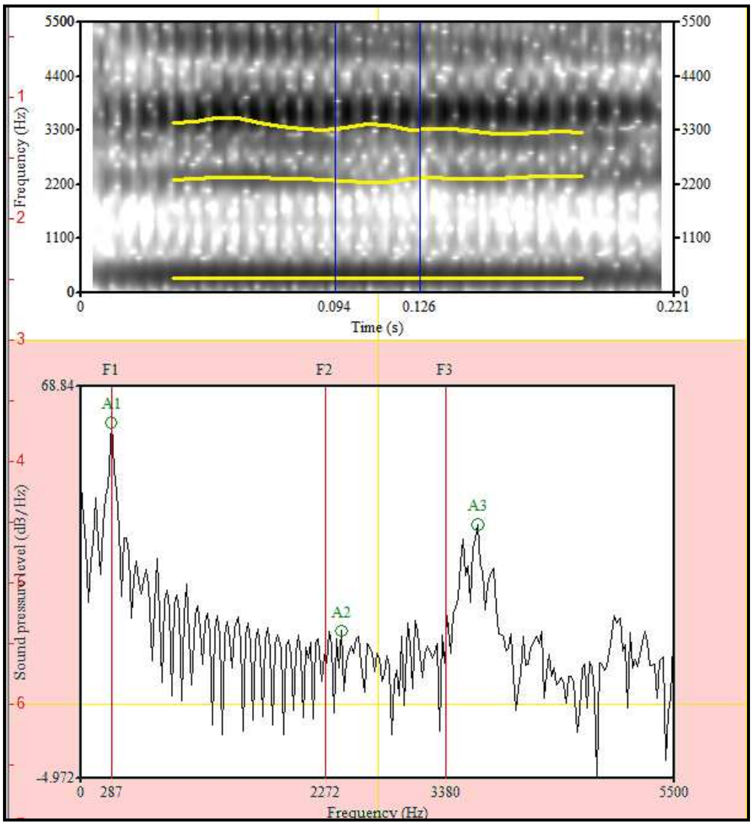
For the sake of an effective and transparent storage, a transparent file naming coding scheme was devised, where the file name structure presented clearly singer (coded with a number from 01 to 06), vowel quality, voice quality, an indication of the time point where the token was extracted, and name of the original song file. This was a necessary step for clear book-keeping, resulting in a data collection which could be easily accessible and consultable for further reference and research by other researchers as well. Moreover, this was crucial in order to obtain files that could be successfully parsed by the Praat scripts measuring voice quality.

3.1.5 Praat and Praat scripts

The signal processing package Praat was the software tool used to work on the recordings and to produce the data set analysed in this project. The core functions of the program allowed working in an environment that simultaneously presented spectrograms, waveforms and the different tiers of annotations provided in the "Metre and Melody of Dinka Speech and Songs" database, that are specifically meant to be run in the program. Moreover, the same software tool could be used to select and cut off the segments corresponding to breathy and modal vowel sounds to be stored and eventually analysed with the scripts.

Mills' Praat scripts replicate all the methods to measure spectral tilt presented and discussed in the previous chapter, together with auxiliary subscripts to track formants and pitch values. A master script is set up with a series of forms, where the user can select the measurements to be performed and adjust the default values to the needs of the acoustic analysis. Once started, the scripts can run the measurements in two ways: automatically, directly yielding the values in a tab-delimited text file, or semi-automatically, leaving to the user the choice to accept the results or to adjust the default parameters in order to get a satisfactory output. For example, the spectral regression method is run automatically, while methods like formant tracking and A1-A2 require the user's visual check for every token.

The formant tracking subscript is a good example to describe how the scripts work: after selecting the preferred measuring method using a graphic interface, the user is presented with visual feedback consisting of two parts: a wideband spectrogram in the upper part of the working window, reporting overlaid formant tracks in yellow and an average FFT power spectrum of the window processed, where vertical red lines correspond to the formant values and labelled green circles highlight the respective peak amplitudes (see figure below). If the users detect visible inconsistencies, they can fix them by modifying a series of parameters including the position of the window and its length.



In this study, two of the spectral measures implemented in the scripts were applied, namely, formant tracking and spectral regression, which where chosen because of their effectiveness and the significance of their results in measuring voice quality in speech data, already proved on Dinka by Mills & Remijsen (unpublished data). It is important to underline the fact that these measurements were tested for the first time on song data.

3.1.6 Analysis of the results

The statistical analysis of the results was performed using the statistical software packages R, while histograms and box plots were generated processing the output of the Praat scripts using Microsoft Excel and SPSS. The data collected and processed were then compared with the results obtained by Mills & Remijsen (unpublished data) regarding the Dinka speech data in order to find similarities or significant discrepancies in voice quality distinction between the spoken and sung data.

3.2. Experiment design

3.2.1. Aims

A perceptual experiment was set up to contribute to answering the main research question from a perceptual point of view. Its main purpose was therefore to determine whether the phonological voice quality distinction between modal and breathy vowels is perceived with the same accuracy by human listeners in songs and in speech. The experimental results are meant to be analysed and compared with the results of the acoustic analysis.

3.2.2. Stimuli

Four speakers and four singers were selected from the “Metre and Melody in Dinka Speech and Songs” project database. Each set included four male voices and a female one. All the audio files were recorded in waveform audio format (wav).

The speech tokens consisted of monosyllabic words presenting a CVC structure, which were preferred to tokens corresponding to vowels alone. This preference was motivated by the fact that the task of identifying the vowels in some kind of context would result easier and more natural for the human listeners involved. In this way, events that could interfere with the intelligibility and naturalness of the stimuli (such as the presence of disruptions or clicks in the recordings) were avoided.

Three different sets were created: the first one included six minimal pairs showing the opposition between breathy and modal voice in the six Dinka vowels that can be affected by the voice quality distinction: /i/, /e/, /ɛ/, /a/, /ɔ/, /o/. The second and the third sets consisted in 48 tokens each, presenting again CVC monosyllabic words carrying one sample of each vowel in its breathy and modal qualities per speaker. It has to be noted that the stimuli in the second and third sets, although uttered by the same four speakers, consisted in completely different words, in order to rule out the impact of acoustic learning and memory effects on the overall accuracy (Ladefoged and Broadbent 1957).

In order to keep consistency between the number of voices used for the identification of the speech and song data, only four of the six singers studied in the acoustic analysis were selected. The four singers were also selected to reflect the same gender distribution of the speech data set: one female and three male voices. In absence of enough material to collect an adequate set of meaningful monosyllabic CVC words, syllables presenting a CVC structure were cropped out from the songs, with the aim of obtaining song samples as similar to the speech ones and as intelligible as possible. As already mentioned, the intelligibility of the data was a priority to be maintained throughout all the phases of the experiment: an identification task of the voice (and often also vowel) quality in the same tokens used in the automated acoustic analysis would have been a virtually impossible job for human listeners, mainly because of the short production time and the presence of clicks where the sounds had been cut abruptly from their surrounding context.

3.2.3. Listeners and procedure

The experiment was designed, implemented and presented to the subjects using the e-Prime software suite.

The subjects tested were 16 postgraduate students belonging to the School of Philosophy, Psychology and Language Science of the University of Edinburgh, with no particular linguistic training and with no regard to their linguistic background. Only native speakers of tone languages from South East Asia such as Vietnamese and Southern Chinese were excluded, because the strong interaction between tone and voice quality present in their first language could have affected their judgment in determining the types of phonation.

The experiment was run in the quiet environment of the computing laboratory of the Department of Linguistics and English Language of the University of Edinburgh. Upon the successful completion of the experiment the participants were paid a small sum of money granted by the University for taking part in the study. Informed consent forms were distributed to the subjects, signed and collected for safe keeping.

3.2.4. Structure

The experiment was structured in four phases:

1) Training on minimal pairs;
2) forced choice with feedback on spoken data;
3) forced choice without feedback on spoken data;
4) forced choice without feedback on song data.

The first two parts were designed as training, while the third and fourth phases were the crucial sections of the experiment, as they allowed the comparison of the perception accuracy scores between speech and song data.

Part 1:

The subjects were presented with four minimal pairs of Dinka monosyllabic words with the typical CVC structure, presenting the phonological opposition between modal (for the sake of clarity presented as "normal" in the instructions presented on screen to the subjects) and breathy vowels. Each pair of words was spoken once by one of the four speakers selected, in order to highlight the variations in voice quality from speaker to speaker and to train the listeners to get used to the impact of these idiosyncratic differences among the various voices. The phonological value of the distinction was underlined with the on-screen presentation of the transcription of the word using the traditional Dinka orthography and with the simultaneous presentation of the corresponding English translation. An indication of voice quality was always clearly presented in capital letters below the words and their English gloss.

Part 2:

The training proceeded with an interactive identification test in which the subjects were asked to choose between the values "normal" and "breathy" after listening once to a stimulus. Stimuli were randomly extracted and presented from a data set of 48 vowel realisations. As already mentioned, the tokens consisted of CVC monosyllabic words containing 6 modal vowels + 6 breathy vowels each for four speakers, for a total of 48 tokens. In this part of the experiment, the subjects were presented with a feedback message after every choice, reporting a short message letting them know if their answer was right or wrong and their progress in terms of overall accuracy score in identifying voice quality.

Part 3:

This phase tested the subjects' ability of identifying voice quality in speech data with a replication of Part 2, with a significant difference: the elimination of the feedback messages. In this section, the performance of the subjects should prove that the previous training phase, first passive and then active, had helped them to learn how to identify the voice quality distinction. A completely different 48-item set of stimuli was used in order to avoid too much auditory influence on the performance from the previous phase. Finally, the stimuli were selected with the same criteria adopted in Part 2 and using speech material produced by the same four speakers. Therefore, the only change in the set of stimuli consisted in the words presented.

Part 4:

In the last phase, subjects were asked to identify voice quality in song data, again without receiving any feedback. This section was then similar to Part 3, with the difference that the stimuli presented were extracted from the sung material instead of being drawn from the speech data set.

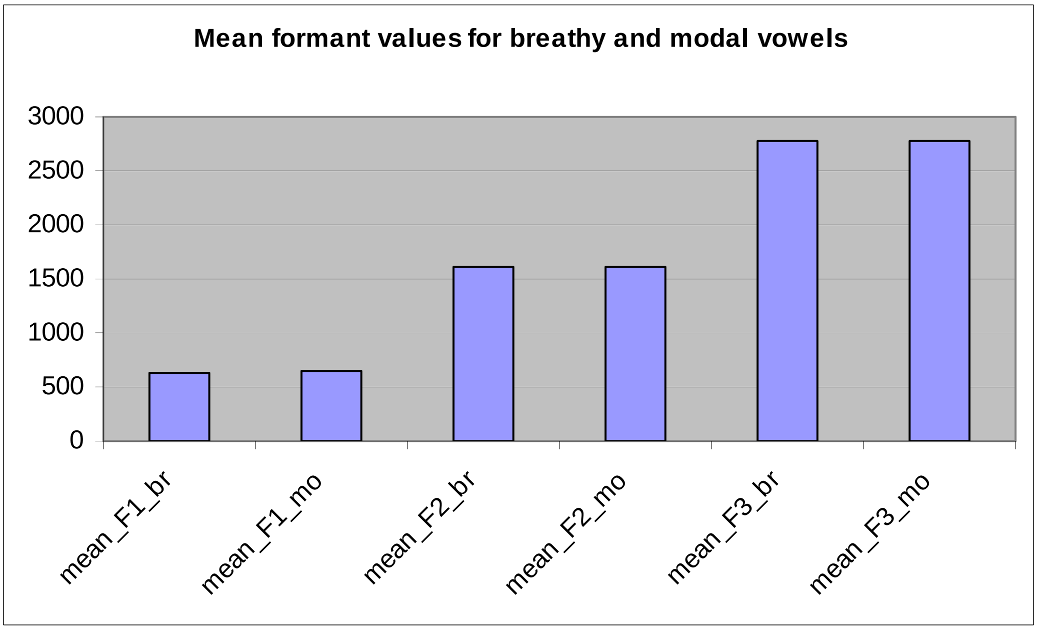
Chapter 4.
Analysis of the results

4.1 Results of the acoustic measurements

4.1.1. Formant tracking

The formant tracking Praat script described in the previous chapter was run to measure the vowel tokens, resulting in a graphic and numeric presentation of the values of the first three formants and their corresponding amplitude peaks.

Even after a first informal inspection of the mean values of the formant values, plotted in a histogram, reproduced in the figure below, it was evident that the expected sizeable difference in the energy produced at higher frequencies, which had been found and reported in Luanyjang Dinka speech data by Remijsen & Manyang (2009), could not be detected in the song data. Indeed, the difference between the second and third formant mean values resulted immediately very small to be of any significance.



After a closer analysis of the mean formant values organized by speaker (see table below), the claim that the values of the first formants are lower in breathy vowels than in their modal counterpart was the only expectation which resulted at least partially fulfilled by the song data.

Singer	F2 breathy vowels	F2 modal vowels	F2 breathy vowels	F2 modal vowels	F3 breathy vowels	F3 modal vowels
1	638.663778	640.8414883	1918.781149	1857.666468	3255.287178	3260.841456
2	625.2931857	663.6332074	1418.13726	1442.636961	2679.137799	2591.88161
3	721.4059763	628.4879623	1538.860534	1598.650175	2516.19155	2544.392601
4	624.3045337	720.9570312	1706.597026	1630.10573	2831.46704	2916.879279
5	580.4379514	589.1696388	1557.776597	1613.718556	2816.642134	2771.805015
6	592.9736022	674.6336711	1551.493683	1531.408052	2521.311275	2526.178809
Mean	630.5131712	652.9538332	1615.274375	1612.364324	2770.006163	2768.663128
S.Dev,	49.61950567	44.77099627	174.6609726	138.5922847	274.218191	284.4521136

At this point, the lowering of F1 in breathy vowels in respect to modal vowels seemed to remain the only acoustic candidate to account for the voice quality distinction in songs, although the difference between the F1 mean values did not seem to be very high. The significance of the difference between the mean F1 values in breathy and modal vowels was verified running a matched pair t-test, while the values corresponding to the higher formants were left aside because evidently too similar to each other. The outcome of the test on the F1 mean values (p-value = 0.4302, t = - 0.8225, d.f. = 9.896) showed that the difference between the two mean values was not statistically significant.

Nevertheless, before drawing a conclusion in favour of the neutralisation of voice quality in Dinka songs, a more detailed analysis of the F1 values was required. In particular, the data needed be optimised in order to convey a more precise vision by reducing the impact of the extralinguistic factors that could have influenced the outcome of the analysis. As it has already been mentioned, voice quality in particular and formant values in general are highly dependent on gender differences and on personal idiosyncratic features (Hanson 1997, Blankenship 2002, Avelino 2010) and for this reason it was necessary to adopt some procedure that could make the data comparable with each other with a smaller amount of variation. First, the song data were separated in two groups corresponding to the sex of the singers: the vowel tokens were then collapsed across the four male singers on one side (32 tokens, four per voice quality) and across the two female voices on the other (16 tokens). Second, in order to reduce the influence of the idiosyncratic and personal features that could have affected the results, the scores were z-transformed, following the standardising procedure adopted by Mills and Remijsen (unpublished data) for the illustration of the Dinka speech data, to be finally organized in box plots that could straightforwardly illustrate the differences in F1 values for modal and breathy vowels.

After this treatment, the data showed that only two male singers consistently presented lower F1 values in their breathy realisations, while all the other singers did not seem to follow any particular pattern. These results seemed to confirm quite strongly that neither the lowering of F1 is a distinctive property in determining the voice quality in Dinka songs. More statistical tests (matched pair t-tests) were performed to compare the mean values of the opposite pairs of phonation types across the female singers and across the male ones. The outcome of the tests confirmed the absence of any statistical significance in the difference between breathy and modal vowels in the song data.

The fact that only two of the six singers analysed were consistent with the expected lowering of F1 in breathy vowels deserved some more attention. If the dialectal variety of Dinka was the same for the two singers, being at the same time different from the one characterising the productions of the other singers, the data could have shown an interesting result in terms of variation within the complex Dinka linguistic macro-reality. In fact, the information incorporated in the metadata files in the "Metre and Melody in Dinka Speech and Songs" database confirmed that the two singers had indeed used the same variety of Dinka, that is, the one spoken in the Twic county, in the State of Warrap, where people speak the main dialect of Rek Dinka. As mentioned in Chapter 2, Luanyjang Dinka belongs to this group. The potential importance of this discovery, however, was soon mitigated by the fact that only two of the other four singers had used a different variety of Dinka (the North-Eastern Dinka Ageer dialect), while the remaining two had used the same Rek Dinka variety spoken in the Twic county and the surrounding area. The fact that they belonged to the same dialect area of the two singers presenting constantly lower F1 values in breathy realisations virtually neutralised the idea that this could be a salient characteristic of the Luanyjang, or at least of the Rek regional variety.

To sum up, the results yielded by the formant tracking script and by the following analyses, after a thorough analysis which has included statistical tests and which has even taken into account some socio-linguistic factors, clearly suggest that the null hypothesis, predicting a neutralisation of voice quality distinction, is not to be discarded.

4.1.2. Spectral regression

The Praat script, which implemented the measuring method reported in Kochanski et al. (2005), worked comparing the functions obtained from spectral tilt for the various vowel tokens, yielding a series of parametric values for slope and intercept of every function.

A first look at the numeric data and a preliminary plotting of the results into a histogram (see figure below) impressionistically showed that the results corresponding to modal and breathy vowels were almost identical.

This observation, which could also be drawn straightforwardly from the numerical results reported in the table below, was statistically confirmed both for the slope and for the intercept values: a matched pair t-test comparing the mean values of slope for modal and breathy vowels yielded a p-value = 0.989 (t = 0.0142, d.f. = 9.856) and the same test comparing the mean values of intercept for breathy and modal vowels resulted in a p-value = 0.9876 (t = -0.0159, d.f. = 9.858). With such values, there is no evidence on which to reject the null hypothesis: consistently with the conclusions drawn upon the results of the formant tracking measurement, the data obtained from the spectral regression measurement strongly suggest that there is no detectable spectral difference in conveying voice quality in Dinka songs.

Singer number	Slope (modal vowels)	Slope (breathy vowels)	Intercept (modal vowels)	Intercept (breathy vowels)
1	-135.939	-136.005	296.4996	296.6427
2	-145.639	-145.581	317.6561	317.529
3	-153.332	-152.42	334.4371	332.4481
4	-138.83	-141.052	302.7963	307.6442
5	-138.148	-138.195	301.3193	301.3295
6	-151.085	-149.382	329.5331	325.8197
Mean	-143.829	-143.772	313.7069	313.5688

4.1.3. Discussion of the results of the acoustic analysis

The formant tracking analysis has showed that breathy vowels produced in songs presented mildly lower F1 values, to a certain extent in accordance with the expectations based on studies like Malou (1988) for Dinka in general, Remijsen and Manyang (2009) for Luanyjang Dinka and Gordon & Ladefoged (2001) from a cross-linguistic perspective. This was also what has been observed in the ongoing study on speech data carried out by Mills and Remijsen (unpublished data). However, the difference detected in the song data is definitely smaller than the one measured in speech data, and was proved to be not statistically significant, thus impeding a clear characterisation of voice quality in the song data.

In addition to this, the results obtained from the spectral regression measurement strongly support the claim that no distinction in voice quality is conveyed in Dinka songs, thus answering negatively to the main research question of the work.

4.2. Results of the experiment

The results of the experiment clearly confirmed the conclusion which was drawn from the results obtained from the acoustic measurements: the voice quality distinction seems to be really neutralised in Dinka songs. The listeners taking part in the experiment could not recognise voice quality in songs, although they were effectively able to do it when presented with speech data. Indeed, as it was expected, the listeners showed fairly good performances in identifying normal or breathy vowels for the speech tokens, but they presented a clear drop in accuracy when facing the same task applied to song data. The table below shows the accuracy scores of the subjects in Part 2 (forced choice with feedback on speech data), Part 3 (forced choice without feedback) and Part 4 (forced choice without feedback on song data). Although the accuracy scores were not very high, the identification tasks on speech data showed scores that were always above 60%, while the accuracy percentage on song data showed how the answers were given almost completely by chance.

Subject number	**Accuracy % Part 2**	**Accuracy % Part 3**	**Accuracy % Part 4**
1	62	60	48
2	65	48	40
3	54	64	56
4	60	81	43
5	69	67	54
6	73	73	58
7	69	73	50
8	58	71	56
9	60	69	50
10	54	67	54
11	73	77	46
12	46	48	65
13	52	54	54
14	48	67	50
15	54	60	52
16	71	83	50
Mean	60.5	66.375	51.625
Standard Deviation	8.801515	10.4427	5.998611

What was particularly surprising is that before tackling the identification tasks on song data, the subjects had shown an improvement in their performances, giving the impression that some learning process was activated. Indeed, the accuracy scores showed that the subjects could have really taken advantage of the training phase of the experiment, corresponding to Part 1 (passive training) and Part 2 (active training), resulting in an evident improvement from the forced choice task with feedback to the unsupervised one.

A statistical analysis of the results was therefore performed to back up the supposition that subjects had effectively been trained and learned to a certain extent how to distinguish voice quality differences in speech. The statistical significance of this difference was tested comparing the mean accuracy percentages scored by the subjects in Part 2 and Part 3 performing a matched pair Student's t-test. The results of the test confirmed the hypothesis that the performance of the subjects was significantly more accurate in the unsupervised section of the experiment (mean = 66.375 %) than in the section with feedback (mean = 60.5 %; $t = -1.7207$, p-value = 0.0959, d.f. = 29.164).

The results yielded by the statistical analysis are consistent with what could be observed in the informal comments and reactions of the subjects after taking part in the experiment. Many participants expressed their difficulties and, in some cases, their frustration in respect with the last part of the experiment on song data. They had the impression that they had been really learning how to distinguish the different types of phonation and that they had progressively become more confident in their own ability and more accurate in identifying voice quality until the beginning of Part 4, when many of them reported how they suddenly felt lost and confused when presented with the sung syllables.

As reported in the last chapter of this work, this drop in accuracy for song data could be motivated by a series of causes. However, the experimental results suggested that the listeners did not seem to tune on the spectral cues connected to vowel quality (the formant values), which are the only ones that could be still acoustically detected in Dinka songs, although to a lesser extent than in speech data.

In addition to these findings, which suggested once more that the distinction in voice quality modes was virtually neutralised in songs, this study tried to answer the side research question raised by these negative results: is there a dominant voice quality mode in the sung realisations? The experimental results showed that the choices of the subjects varied freely in identifying modal or breathy voice quality, and that they showed no particular preference towards a voice quality rather than another, presenting once more random selections that could not pattern in favour of one or the other possible alternative.

Chapter 5.
Conclusion

"Is the phonemic distinction between modal and breathy vowel conveyed in Dinka songs?" This was the principal research question asked at the beginning of this study and driving the entire project. After a close analysis of the evidence obtained one the one hand from the acoustic measurements, and on the other hand from the perception experiment, the answer to the question clearly seems to be negative.

It could be possible that the acoustic measurements applied, namely formant tracking and spectral regression, could have not been effective or sensitive enough to account for the phonetic characteristics of songs. However, there are two good reasons to rely on their results on Dinka song data. Firstly, one must consider the fact that the measuring methods had previously worked well and yielded significant results in representing the voice quality distinction in Dinka speech data, as it is demonstrated by the data obtained by Remijsen and Mills with the application of the same scripts on spoken Dinka (unpublished data). Secondly, the outcome of the perception experiment pointed clearly in the direction of a neutralisation of the distinction in voice quality, clearly confirming the results of the acoustic analysis.

Regarding the experiment, there might have been a number of interferences or factual problems affecting its results: the qualitative difference between the spoken audio recordings and the song data was probably one of them. While the spoken data set included very clearly uttered samples, the song data presented many broken syllables, often too short or deformed by the melody to be clearly parsed and understood by human ears. This could have contributed to the difficulties that the subjects encountered in identifying the song data after performing well on the identification tasks based on speech data.

This potential problem could be solved with a greater amount of song data. A more abundant data set would allow finding vowel realisations in minimal pairs of meaningful words and comparing them both acoustically and perceptually in the same way it was done for the speech data. The availability of more data would also give the opportunity to select only very clearly intelligible and natural sounding samples to be more successfully processed by human listeners.

Other factors that could have affected the voice quality identification process are the idiosyncratic features and the individual voice differences across the speakers. As seen in Chapter 2 and 3, these personal idiosyncrasies often influence a great deal the identification of voice quality and they could have misled the subjects in their judgments, especially when they were not familiar with phonological distinctions based on voice quality (Kreiman & Gerrat 1997; 1999). Also in this case, the presence of a larger song data set would help in selecting clearer and unambiguous tokens to be presented to the listeners.

The problem of determining which acoustic correlates characterise voice quality still remains unsolved. The spectral cues that seem to be effectively used to acoustically define and perceive the voice quality distinction in speech seem to be absent or neutralised in Dinka songs. There might be some residual distinctive force in the lowering of F1 in breathy vowels, although the statistical tests performed on the results of the formant tracking analysis showed that the difference was not significant, in accordance with the inconsistent performances in the perception experiment. Moreover, in the identification task on song data, the choices made by the subjects did not show any pattern, thus answering negatively also to the side question regarding a possible predominance of one type of phonation (hypothetically the modal quality) in songs.

The claim that voice quality distinction is neutralised in songs raises a crucial question in terms of the phonetic and phonological study of Dinka songs: how will it be possible to disambiguate the homophones that are distinct by voice quality in spoken language, if the voice quality distinction is not conveyed? Is the context enough to compensate for the missing suprasegmental information in conveying the exact meaning of an ambiguous word? The picture gets even more complex if we consider that the tone system is also competing with the rules of the songs (Wong & Diehl 2002) and that even the rules of vowel length seem to be often violated in favour of the needs of the melody (see Chapter 3): all the three independent levels of the Dinka suprasegmental system seem to interact with the musical constraints. Yet Dinka singers and listeners continue to sing, trade and enjoy their songs: if the context is not enough to make up for the missing prosodic information, then there must be some kind of acoustic cue enabling the correct parsing of the information carried by the songs.

All these considerations lead to what could be a very interesting development of this study: the presentation of the perception experiment to Dinka native speakers. The native speakers could tune on acoustic cues different from the cues used by speakers of other languages in making voice quality judgments, overcoming the difficulties in distinguishing voice quality often reported for the speakers of languages in which voice quality is not phonemically distinctive (Kreiman & Gerrat 1997; 1999). Therefore, comparing their performances with the ones scored by the subjects involved in this study could give more insight on the nature of the acoustic cues that may or may not still be present in songs. For instance, the mild lowering of F1 detected in songs could be still enough for a Dinka native speaker to parse correctly the information conveyed by a song. On the contrary, if their accuracy scores were comparable with the ones recorded in this study, this could represent strong evidence to support the compensating function of the context alone. As it was suggested before, the availability of tokens including sung minimal pairs that could replace the often meaningless syllables used in the last part of the experiment could contribute a great deal to answer to this question.

Another important factor that should be considered for further developments of the research carried out in this study is language variation, both within the Dinka speaking enclave and in relation with the neighbouring languages. The study of variation across Dinka dialects is one of the main research lines currently pursued by the members of the "Metre and Melody in Dinka Speech and Songs" project team; in particular, fieldwork activity recently carried out by Bert Remijsen was mainly addressed to the documentation and study of the Dinka dialectal variation (Remijsen, personal communication, June 2010). As suggested in Chapter 4, the linguistic variation across the different Dinka dialects is an important phenomenon that should be taken into consideration in further phonetic studies on Dinka songs. Working on large data sets that consistently present a single dialectal variety, or, in contrast, that systematically compare materials extracted from different dialects would greatly contribute to verify whether phenomena such as the neutralisation of the voice quality distinction and its connected acoustic cues are detectable only in some dialects or if they are features shared by the whole Dinka linguistic community.

As for a comparative perspective that takes into account the variation and the contacts between Dinka and the surrounding linguistic realities, very interesting results have been obtained by another current member of the "Metre and Melody in Dinka Speech and Songs" project: acoustic measurements performed by Tatiana Reid on the neighbouring Western Nilotic language Thok Reel have shown the expected stronger energy amount at higher frequencies, but no lowering of F1 (Reid, personal communication, July 2010). These observations once again strongly support the existence of significant differences in conveying the voice quality distinction across the languages. It is worth repeating that the lowering of F1 was the only expected event to be, at least mildly, shown in the acoustic measurements applied in this study, and that it has been documented for any variety of Dinka that has been acoustically measured so far (Malou 1988, Edmonson & Esling 2006, Remijsen & Manyang 2009). It is quite surprising to see how this feature is not present at all in a kin language spoken in a very close geographical area.

In the words written by Gordon & Ladefoged (2001), “it is unlikely that future research will yield many truly universal observations about the range and realisation of phonation types in languages of the world”. However, the steady advancements in the technology applied to the study of voice quality in other disciplines, especially in the field of biomedicine and in the study of pathological speech, have already contributed a great deal to the study of the physiological correlates of voice quality. In particular, more and more scholars are following the advice suggested by Gerrat and Kreiman (2001) and they are working hard in order to come to a unified description of voice quality and phonation types that could include the findings and the terminology of the different disciplines involved (Esling 2006; 2006bis, Edmonson & Esling 2006, Avelino 2010). This study is to be considered one more attempt to move the discussion on voice quality in the direction of this unifying perspective.

This work, characterised by the application to the study of songs of phonetic measurements and experimental procedures that are usually limited to the analysis of speech data has hopefully contributed to give some more insight on the complex problem of explaining voice quality. Although at the present state the answer to the main research question is clearly negative, further research based on the results and the suggestions reported in this section could lead to a finer understanding of how voice quality is conveyed in Dinka songs. In particular, the author wishes that this work could be read by members of the Dinka speaking community and that the perception experiment could be soon presented again to Dinka native speakers, seeing in this further development the ideal follow-up of his research in order to understand and explain the mechanism the voice quality distinction in Dinka songs.

References

Andersen, T. (1987). The phonetic system of Agar Dinka, Journal of African Languages and Linguistics, 9, 1-27.

Avelino, H. (2010). Acoustic and Electroglottographic Analyses of Nonpathological, Nonmodal Phonation, Journal of Voice, 24:3, 270-280.

Berry, D.A. (2001). Mechanisms of modal and nonmodal phonation. *Journal of Phonetics*, 29, 431-450.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, 30:2, 163-191.

Catford, J.C. (1964). Phonation types: the classification of some laryngeal components of speech production. In D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scout & L.M. Trim (eds.). *In honour of Daniel Jones*. London: Longman, 26-37.

Catford, J.C. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Catford, J.C. & Esling, J.H. (2006), *Articulatory phonetics*. In K. Brown (ed.), *Encyclopedia of language and linguistics*. 2nd edition. Oxford: Elsevier.

Edmondson, J.A. & Esling, J.H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. Phonology 23, 157–191.

Esling, J.H. (2006), States of the glottis. In K. Brown (ed.), *Encyclopedia of language and linguistics*. 2nd edition. Oxford: Elsevier.

Esling, J.H. (2006bis), Voice quality. In K. Brown (ed.), *Encyclopedia of language and linguistics*. 2nd edition. Oxford: Elsevier.

Fant, G. (1960). *Acoustic Theory of Speech Production*, Mouton & Co, The Hague.

Fulop, S.A., Kari, E., Ladefoged, P. (1998). An Acoustic Study of the Tongue Root Contrast in Degema Vowels. *Phonetica* 55, 80-98.

Gerrat, B.R. & Kreiman, J. (2001). Toward a taxonomy oh non-modal phonation. *Journal of Phonetics*, 29, 365-381.

Gordon, M. & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29, 383-406.

Grivicic, T. & Nilep, C. (2004). When Phonation Matters: The Use and Function of Yeah and Creaky Voice. *Colorado Research in Linguistics*, 17:1.

Gussenhoven, C. & Jacobs, H. (2005). *Understanding Phonology*, 2nd edition. London: Hodder Education.

Horne, M. (2009). Creaky fillers and speaker attitude: data from Swedish. In D. Barth-Weingarten & N. Dehé (eds.), *Where Prosody Meets Pragmatics*. Bingley: Emerald Group Publishing.

Hanson, H.M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustic Society of America*, 101:1, 466-481.

Hanson, H.M., Stevens, K.N., Je, H.K., Chen, M.Y., Slifka, J. (2001). Towards models of phonation. *Journal of Phonetics*, 29, 451-480.

Heldner, M. (2001). Spectral emphasis as an additional source of information in accent detection. In *Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*. Red Bank, NJ: ISCA.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* 31, 39-62.

Idris, H.F. (2004). Modern developments in Dinka language. *Goeteborg Africana Informal Series*, 3.

Kreiman, J. & Gerrat B.R. (1997). Validity of rating scale measures of voice quality. *Journal of the Acoustic Society of America*, 104:3,1, 1598-1608.

Kreiman, J. & Gerrat B.R. (1999). Measuring voice quality. In *Handbook of voice quality measurement* (R. Kent & M.J. Ball, editors). San Diego: Singular.

Ladd, D.R., Remijsen, B., Manyang C.A. (2009). On the distinction between regular and irregular inflectional morphology: evidence from Dinka. *Language*, 85(3), 659-670.

Ladefoged, P. and Broadbent, M.E. (1957), Information conveyed by vowels. *The Journal of Acoustical Society of America* 29: 98-104.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: University of Chicago Press.

Ladefoged, P. & Matthieson, I. (1996). *Sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, P. (2003). *Phonetic data analysis*. Oxford: Blackwell.

Ladefoged, P. (2006), *A course in phonetics*. 5th edition. Boston: Wadsworth.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lewis, M. P. (ed.) (2009). *Ethnologue: Languages of the World*, 16th edition. Dallas, SIL International [online version: http://www.ethnologue.com, Ethnologue report for Dinka http://www.ethnologue.com/show_language.asp?code=din].

Malou, J. (1988). *Dinka vowel system*. Dallas: SIL & University of Texas at Arlington.

Metre and Prosody in Dinka Speech and Songs. http://projects.beyondtext.ac.uk/dinkaspeech/index.php

Mills, T. (2003). *Cues to voicing contrasts in whispered Scottish obstruents*. MSc dissertation. University of Edinburgh.

Mills, T. (2009). *Speech motor control variables in the production of voicing contrasts and emphatic accent*. PhD thesis. University of Edinburgh.

Nilotic Prosody. http://www.ling.ed.ac.uk/nilotic/index.shtml

Reid, T. (2009). *Suprasegmentals in Shilluk nominal morphophonology. Undergraduate Honours dissertation*. University of Edinburgh.

Remijsen, B. (2001). *Word-prosodic systems of Raja Ampat languages*. PhD thesis, Universiteit Leiden Centre for Linguistics.

Remijsen, B. (2009). Stem-*internal morphology in Shilluk transitive verbs*. Paper presented at the Morphology of the World's Languages (MOWL) Conference, University of Lepzig.

Remijsen, B. (2010). *Metre and Melody in Dinka Speech and Song*. Poster presented at the Beyond Text Award Holders Event, 10th May 2010, London: Mermaid Conference Centre. [available online at http://www.beyondtext.ac.uk/posters/may2010/dinkaspeech.shtml; last access 9th August 2010]

Remijsen, B. & Gilley, L. (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics*, 36(2), 318-344.

Remijsen, B. & Manyang, C.A. (2009) Luanyjang Dinka – Illustration of the IPA. *Journal of the International Phonetic Association*, 39, 113-124.

Roettger, L. & Roettger, L. (1989). A Dinka dialect study. Occasional Papers in the Study if Sudanese Languages, 6, 1-64.

Simpson, A. (2009). Breathiness differences in male and female speech. Is H1-H2 an appropriate measure? In *Proceedings, FONETIK 2009, Department of Linguistics, Stockholm University*.

Sluijter, A.M.C. & van Heuven, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100, 2471-2485.

Titze, I.R. (2008). Nonlinear source-filter coupling in phonation: theory. *Journal of the Acoustic Society of America*, 123:5, 2733-49.

Wong, P.C.M. & Diehl, R.L. (2002). How can the lyrics of a song in a tone language be understood? *Pyschology of Music*, 30, 202-209.