THE NEUTRALIZATION OF THE VOICE QUALITY DISTINCTION IN DINKA SONGS: A PRODUCTION AND PERCEPTION STUDY

Luca Rognoni

University of Padova, Italy
luca.rognoni@studenti.unipd.it

ABSTRACT

The purpose of this study is to determine if the phonemic voice quality distinction of Dinka is also conveyed in songs. Acoustic measurements (formant tracking and spectral regression) were applied for the first time to a song data set. The results clearly pointed towards a neutralization of the voice quality distinction in songs.

A perception experiment was performed to back up the results of the production study. Listeners were first trained to distinguish between modal and breathy vowels in speech, and then tested on speech and song data. The experimental results corroborate the claim that no voice quality distinction is conveyed in singing.

Given the absence of any significant acoustic cue shown by the acoustic analysis or detected by the listeners, the author concludes that context seems to be the only means to disambiguate homophones in songs.

Keywords: voice quality, breathy vowels, phonation types, Dinka, songs

1. INTRODUCTION

Many languages present a phonological distinction of voice quality distinguishing between modal and non-modal phonation. Modal voice refers to the default mode used as a term of comparison to describe the other modes of voice quality, such as breathy voice. This has been defined as the impression of an audible escape of air through the glottis, resulting in an h-like sound [8]. From a physiological point of view, voice quality depends on laryngeal settings [3].

In terms of acoustics, voice quality is a complex phenomenon and it is difficult to measure. A variety of measurements have been proposed in the literature [4, 5, 6, 7, 16]: all these methods are connected with the analysis of spectral tilt, i.e. with aspects connected to the slope of the spectrum of vowels. However, so far none of them has proved entirely satisfactory, cf. [12]. Even perceptually, the identification of breathy voice is difficult, due to its high dependency on extra-linguistic idiosyncrasies and gender of the speakers [5;8].

The main research question driving this study is to determine if the phonemic voice quality distinction between the modal and breathy vowels of Dinka is also conveyed in songs. The study of voice quality in singing is interesting, because songs represent a parallel phonetic environment where to test methods to quantify and describe prosodic phenomena characterizing speech. This approach is even more useful for the dimensions that are not fully understood and explained in speech, such as voice quality.

This work is structured in two phases. The first phase is a production study involving two acoustic measurements. The second phase is a perception experiment based on the evaluation of voice quality in speech versus song data.

2. THE VOICE QUALITY DISTINCTION IN DINKA

Dinka is a language belonging to the Nilo-Saharan family, spoken in Southern Sudan by over two million speakers [9]. In recent years, this language has been studied with a growing interest for its unusually rich suprasegmental system [11, 14, 15]. Dinka words are mainly based on CVC monosyllabic stems. The distinctive power of the seven vowels (/a/, /ɛ/, /e/, /i/, /ɔ/, /o/, /u/) is enhanced by the three independent suprasegmental dimensions of length, tone and voice quality, which all carry both lexical and morphological functions [14]. As for voice quality, the vowels are characterized by two distinctive phonation types: modal and breathy.

Voice quality in Dinka speech has been quantified via exploratory phonetic measurements [15]. The results suggested that breathy vowels have lower F1 values. Breathy vowels also present more energy at higher frequencies than at lower frequencies in relation to their modal counterparts. These results were the reference for this study, where acoustic measurements were applied to song data for the first time.

The main hypothesis of this study was that the phonological opposition is conveyed in songs in the same way it is in speech. The results of the acoustic measurements were therefore expected to be in line with the ones found in speech.

3. THE STUDY

3.1. Production study

3.1.1. Singers

The researchers of the "Metre and Melody in Dinka Speech and Songs" (MMDSS) project of the University of Edinburgh (Department of English and Linguistics) have created an extensive online database of speech and song recordings in .wav format [11]. For the purpose of this work, the database included enough annotated songs to collect data for six different singers, four men and two women.

3.1.2. Materials and method

The songs were recorded in fieldwork sessions in South Sudan using a Zoom H4N solid-state recorder and an external headset-mounted directional microphone (Shure SM10A).

Praat [2] was used to collect four modal and four breathy realizations per vowel for each speaker. Since the back high vowel /u/ is always breathy [15], it was excluded from the data set. Occurrences of vowels within repetitions of the words in the songs were taken into account as valid tokens. Since the distinction between Dinka short vowels and vowels involved in diphthongs and triphthongs is often unclear [1], the samples were collected from medium and long vowel realizations in monophthongs.

Following these selecting criteria, 48 sections of vowels per speaker were obtained. The duration of these sections ranged from 60 to 100 ms. The tokens were acoustically measured using Praat, applying formant tracking and spectral regression via scripts [13]. The latter method is a spectral tilt measure consisting of a regression line fitted to the harmonic peaks in the vowel spectrum that yields parametric values of “slope” and “intercept” [7]. The scripts automatically processed the signal considering a window of 30 ms at midpoint of the token, and then yielded numeric values in a tab-delimited text file. These values were then statistically analysed.

3.1.3. Results of formant tracking

The mean formant values showed that the sizeable difference in the energy produced at higher frequencies reported for speech data [10, 15] could not be detected in the song data. The differences between the second and third formant mean values were irrelevant and the lowering of F1 values in breathy vowels was the only expectation based on speech data that resulted somewhat noticeable in the song data. The mean values of F1 in modal vowels was 652.95 Hz (standard deviation = 44.77), while their breathy counterparts presented a mean F1 value of 630.51 Hz (standard deviation = 49.62). However, when the difference was verified in a matched pair t-test, the results showed that there was no statistical significance even for that spectral cue ($p = 0.4302$, $t = -0.8225$).

3.1.4. Results of spectral regression

The results corresponding to modal and breathy vowels were almost identical across the two parametric values yielded by the measurement. The mean slope values were -143.83 (standard deviation = 7.29) for modal vowels and -143.78 (standard deviation = 6.45) for the breathy ones. The intercept mean values were 313.71 (standard deviation = 15.90) for modal vowels and 313.57 (standard deviation = 14.09) for the breathy ones. A matched pair t-test comparing the mean values of slope for modal and breathy vowels yielded a $p = 0.989$ ($t = 0.0142$) and the same test comparing the mean values of intercept for breathy and modal vowels resulted in $p = 0.9876$ ($t = -0.0159$). With such values, there is no evidence on which to reject the null hypothesis, that is, that the voice quality distinction is neutralized in songs.

Since the acoustic analysis yielded negative results, clearly suggesting a neutralization of the voice quality distinction, a perception experiment was designed to verify the results of the production study.

3.2. Perception experiment

3.2.1. Subjects

The subjects tested were 16 postgraduate students, with no specific training in linguistics and of various linguistic backgrounds. Only native speakers of tone languages from South East Asia were excluded, because of the strong interaction between tone and voice quality present in their first languages.

3.2.2. Materials

Four speakers and four singers were selected from the MMDSS database, each group including four male voices and a female one. The tokens consisted of CVC monosyllabic words.

Four different sets of stimuli were created. The first included six minimal pairs of words showing the opposition between breathy and modal vowels. The second and the third sets consisted of 48 tokens, composed of spoken CVC monosyllabic words with one sample of modal or breathy vowel per speaker. Two different 48-item sets of words were selected in order to rule out any possible impact of acoustic learning. The fourth set presented 48 CVC syllables extracted from the song data following the same criteria.

3.2.3. Experiment design and procedure

The experiment was structured in four parts. The first two parts represented training, while the third and fourth phases were the crucial sections of the experiment, allowing the comparison between the subjects' performances on speech and song data.
In the first part (passive training on minimal pairs), the subjects were presented with four minimal pairs, each one uttered by a different voice. The voice quality mode was shown on screen together with a transcription of the Dinka word and the corresponding English translation.

In the second part (forced choice with feedback on spoken data), the training proceeded with an interactive identification test. Here the subjects were asked to choose between the values "modal" and "breathy" after listening once to a spoken stimulus. After each response, the subjects were presented with a feedback message keeping track of their progress.

The third part (forced choice without feedback on spoken data) tested the subjects' ability of identifying voice quality in speech data in an unsupervised replication of the previous phase.
Finally, in the fourth phase (forced choice without feedback on song data) the subjects were asked to identify voice quality in song data without feedback.

3.2.4. Results

The subjects were generally able to recognize voice quality in the speech tokens, but they clearly failed in doing so with sung stimuli. Although the accuracy scores were not very high in absolute terms, the identification tasks on speech data showed values that were generally above 60%. By contrast, there is an evident drop in the accuracy percentage on song data, reflecting the fact that the answers were almost completely given by chance (see Table 1).

Table 1: Accuracy scores in percentage values.

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
S. D.	8.801515	10.4427	5.998611

It is interesting to remark that the subjects had shown an improvement in their accuracy from the forced choice task with feedback to the unsupervised one. The results of a matched pair Student's t-test confirmed the hypothesis that the performance of the subjects was more accurate in the unsupervised section of the experiment (mean = 66.375 %) than in the section with feedback (mean = 60.5 %; $t = -1.7207$, $p = 0.0959$). This suggested that the subjects had successfully learnt how to distinguish voice quality in speech, before dramatically dropping their accuracy when tested on song data.

Informal comments of many of the subjects are consistent with the results of the statistical analysis. Many participants reported that despite feeling progressively more confident in distinguishing voice quality in speech, the identification of the sung syllables was unexpectedly difficult and frustrating.

4. DISCUSSION AND CONCLUSIONS

My hypothesis was not confirmed by the results of the acoustic measurements. The formant tracking analysis only showed a mild lowering of the F1 values in breathy vowels, but this difference proved not to be statistically significant. The

results of the spectral regression measurement also failed to capture any clear distinction in voice quality.

In accordance with the acoustical results, the outcome of the perception experiment suggested that the listeners cannot rely on any spectral cues connected to vowel quality (e.g., formant values). The experimental results also lead me to reject the hypothesis of a dominant voice quality mode in the sung realizations: the subjects' responses varied freely in identifying modal or breathy vowels, only showing random preferences.

The claim that voice quality distinction is neutralized in songs raises crucial questions in terms of the phonetic and phonological study of Dinka songs, as tone does in [17]: how is it possible to disambiguate the words that are distinct by voice quality in spoken language, if the phonemic distinction is not conveyed in singing? Is context enough to compensate for the missing suprasegmental information in conveying the exact meaning of an otherwise perfect homophone? Yet Dinka singers and listeners continue to sing, exchange and enjoy their songs: if context is not enough to make up for the missing prosodic information, then we suspect that there must still be some kind of acoustic cue enabling a correct parsing of the information carried by the songs.

A further step in this study would be the presentation of the same perception experiment to Dinka native speakers, who may rely on some acoustic cues that remained unnoticed by the speakers of other languages [8]. In contrast, if their accuracy scores were comparable with the ones reported in this study, this would represent crucial evidence to support the compensating function of the context alone and the neutralization of the voice quality phonemic distinction in Dinka songs.

5. ACKNOWLEDGEMENTS

This work was based on the research I carried out for my MSc dissertation at the University of Edinburgh under the joint supervision of Bob Ladd and Bert Remijsen. I would like to express my deepest gratitude to both for the inspiration and guidance they have given to my first steps into academic research.

6. REFERENCES

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

[2] Boersma, P., Weenink, D. 2010. Praat (Version 5.1.32). *http://www.fon.hum.uva.nl/praat/*.

[3] 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.

[4] Gordon, M., Ladefoged, P. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29, 383-406.

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

[6] 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.

[7] Kochanski, G., Grabe, E., Coleman, J., Rosner, B. 2005. Loudness predicts prominence; fundamental frequency lends little. *Journal of the Acoustical Society of America* 118(2), 1038-1054

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

[9] Lewis, M. P. (ed.) 2009. *Ethnologue: Languages of the World (16th edition)*. Dallas: SIL International. online version: http://www.ethnologue.com/

[10] Malou, J. 1988. *Dinka Vowel System*. Dallas: SIL & University of Texas at Arlington.

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

[12] Mills, T. 2009. *Speech Motor Control Variables in the Production of Voicing Contrasts and Emphatic Accent*. PhD thesis. University of Edinburgh.

[13] Mills, T. 2010. Spectral tilt analysis scripts (Version 0.0.5). *http://www.ling.ed.ac.uk/~tmills/resources.shtml*

[14] 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.

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

[16] 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.

[17] Wong, P.C.M., Diehl, R.L. 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music* 30, 202-209.