Singing in tone languages: Phonetic and structural effects

D. R. Ladd, University of Edinburgh (bob.ladd@ed.ac.uk)

Text of talk presented at the 27th annual meeting of the Phonetic Society of Japan, Kanazawa, September 2013.

My talk today deals with my on-going work on the ways in which linguistic tone is manifested in singing. As you will see, I'm going to talk about two different kinds of effects, which I call **phonetic** and **structural**.

Let's begin with the question that ultimately motivates this research. Singing in tone languages seems to make contradictory demands on voice pitch. On the one hand, in any language, a song is an arrangement of pitches ('notes') based on musical considerations (melody, harmony, verse structure, etc.). On the other hand, in a tone language, song texts ('words' or 'lyrics') include pitch specifications that are purely linguistic (lexical and/or grammatical). Modifications to the linguistically specified pitch may change the meaning of the words. So this leads to an obvious question: **How does the use of voice pitch for musical purposes coexist with the use of voice pitch for phonological purposes in a tone language?** Much more straightforwardly, we might want to phrase the question as it was expressed in the title of an influential article by Wong and Diehl (2002): **How can the lyrics of a song in a tone language be understood?** These questions are the starting point for everything I'm going to say today.

The first thing to say is that, to some extent, these may be the wrong questions. Most importantly, language is redundant, and there are lots of 'top-down' cues to meaning, not just purely phonetic ones. This also applies to song texts. Moreover, song texts often reinforce this redundancy by including extensive repetition, traditional formulaic expressions, and so on. So the fact that musical melody is competing with tone may not make any practical difference. If you think about it, there are lots of other uses of speech that make contradictory demands on the voice, without seeming to create a mystery like singing in a tone language. For example, if you whisper in a language that has voicing distinctions, you would seem to be overriding those

distinctions, yet we find it perfectly normal that people can be understood when they whisper. So it's not quite clear why we feel that tone language singing poses a special problem.

Nevertheless, there is recent research suggesting that top-down redundancy is not the whole story. That is, there seem to be more specific reasons why singing in a tone language is possible without causing problems for the meaning. Broadly speaking, they fall into the two categories suggested in my title, namely phonetic effects and structural effects.

The phonetic effects may not be very surprising. They arise from the fact that there are multiple interacting phonetic cues to most phonological distinctions. This is true of tonal distinctions just as it is true of segmental distinctions. If the pitch cue to a tonal distinction is overridden or modified by the musical demands of singing, other phonetic cues (such as allophonic vowel quality or durational details) may serve to identify the intended tonal category. We know that this kind of compensation applies to whispering in a tone language (Liu and Samuels 2004) and it seems likely that such effects are relevant in singing as well. We can call these phonetic effects 'residual' cues to phonological categories. In addition, it is possible that in some cases phonetic cues can be enhanced or exaggerated in singing, in order to compensate for the modifications of normal speech pitch imposed by the music; this kind of effect is attested in Cantonese opera (Yung 1989). We can refer to these as 'enhancing' cues. Together, residual cues and enhancing cues make up the 'phonetic effects' of my title.

However, we also need to consider what I'm calling structural effects. These may be less expected than the phonetic effects, but it's beginning to appear that they are actually more important. A number of studies are converging on the conclusion that in many tone languages, it is important for **the musical melody and the linguistically specified pitch sequences to match** in certain ways. That is, there are **text-setting constraints** that limit the possible combinations of text and melody. These have been found in a number of languages, including Cantonese, Shona, Dinka and Vietnamese. In these languages, roughly speaking, it appears that **if the linguistically specified pitch goes up** from one syllable to the next, then across those two syllables **the musical melody must not go down** (and vice-versa). Looked at the

other way, if the musical melody goes down from one note to the next, the words must be chosen so that the linguistically specified pitch sequence does not go up on the two corresponding syllables – and vice-versa. I'll come back to talk about the details of this case a little later, but the important point now is that constraints like this seem to be an important part of the solution to the problem of singing in a tone language.

***

With the scene set in this way, I want to spend the rest of the talk discussing tone language singing in more depth. Specifically, I want to talk about both phonetic effects and structural text-setting constraints. First, I'll discuss phonetic effects on the basis of findings from our large project on speech and song in Dinka, a major language of South Sudan. After that, I'll switch topics, and focus on the question of text-setting constraints, in a range of languages.

First let's look at songs in Dinka. Without going into a lot of detail, the relevant linguistic fact about Dinka for the topic today is that it has a unusually rich set of suprasegmental distinctions: a three-way distinction of quantity (vowel length can be short, long, or overlong), a voice quality distinction (modal or breathy), and tone (high, low, and falling in most dialects, plus a rising tone in some varieties)[1]. The other relevant feature of Dinka society for our purposes is that the Dinkas have a lively song tradition (Deng 1973). Individuals compose and 'own' songs praising family members or cattle, insulting rivals or enemies, commemorating important events, and so on. There are also specialists who compose songs for hire. The importance of songs in Dinka society can be seen from the fact that Dinka people who have taken refuge abroad as a result of the civil war that ravaged southern Sudan in the 1990s keep in touch with each others' songs by exchanging cassette tapes (Impey 2013). Musically, Dinka melodies are rhythmically and harmonically simple and based on a strictly pentatonic scale. They are normally performed solo or in unison, with no instrumental accompaniment except, occasionally, clapping or simple

[1] The version of this talk as presented in Kanazawa included a number of sound files of examples of the Dinka suprasegmental contrasts. Here the reader is referred to Remijsen and Ladd 2008 and Remijsen and Manyang 2009 for more detail.

drumming. There is generally a simple rhythmic pulse, and phrases are of variable length, with no 'verses' or overarching metrical structure.

For the last several years I've been working on Dinka and Dinka songs with a couple of colleagues, and one of our central research questions has been to discover how all the suprasegmental distinctions of the Dinka language are manifested in Dinka singing. Right now I'm going to concentrate on the phonetic effects we've found – and also on the phonetic effects we haven't found. Briefly, we have found some subtle phonetic effects for quantity and voice quality, but not for tone.

First, let's look at **voice quality**. You could imagine various kinds of phonetic effects that might reflect the voice quality distinction. First, you might imagine that the singer would somehow use two distinct voice qualities in singing to maintain the lexical distinction, or would enhance the differences found in normal speech. Alternatively or additionally, you might imagine that the allophonic differences that normally accompany the voice quality difference would be preserved as what I referred to above as residual phonetic cues. These allophonic differences include small differences in F0 (breathy vowels are slightly higher, other things being equal) and/or small differences in F1 (breathy vowels have slightly lower vowel height, i.e. higher F1, again, other things being equal). Our investigations have provided weak evidence of residual differences like these, which were explored in a Master's thesis a couple of years ago (Rognoni 2011). Specifically, the F1 differences – the allophonic differences of vowel height – tend to be maintained in singing, though these are not usually big enough to give the voice quality away on their own. However, the differences of voice quality itself do not seem to be maintained. It appears that the voice quality involved in singing largely overrides the distinction found in speech.

Now let's consider **quantity**. Here again we could imagine various phonetic effects: maybe the basic note values are actually held longer for long vowels and cut short for short vowels. Or maybe there's some subtle manipulation of the **ratio** of vowel to consonant in each word. Our preliminary results suggest that there do seem to be small but measurable effects on note duration. Notes of the same rhythmic quantity do seem to be held slightly longer if the vowel being sung is long or overlong than if it is short. There also seems to be an effect on the ratio of vowel to coda consonant:

long vowels are held the longest at the expense of the coda consonant, while short vowels are cut short leaving the consonant duration to fill out the full musical note value. I emphasise that these are not final results, and that we will publish a more complete report in due course. So again, there do seem to be subtle phonetic manifestations of the quantity distinction – certainly not enough to signal the suprasegmental distinctions unambiguously, but perhaps enough to provide a hint of redundancy for the intended meaning.

Now, finally, what about **tone**? Here again you could imagine a couple of different phonetic cues to the tonal categories, cues that would be enough to signal the intended tone. These might take the form of modification of musical notes – performance ornaments that would enhance the difference between one syllable contour and another. Ornaments for this purpose are reported in Cantonese opera by Yung 1989. Alternatively or additionally, we might find that there are allophonic differences of duration or vowel quality that redundantly cue the intended tonal category in speech and can survive the modification of the pitch for musical purposes.

However, we have not found any evidence of these kinds of things in Dinka. There are performance ornaments, especially at the ends of phrases, but they don't seem to be related to tone. As for preserving redundant allophonic differences, there are no obvious allophonic differences of this kind anyway, so we shouldn't expect to find them preserved in song. In short, there do not seem to be any **phonetic** cues to tone in Dinka singing at all. So once more we come back to the apparent mystery of how tone languages manage to convey tonal distinctions in singing.

***

This brings us to the second kind of manifestation of tone in song, namely text-setting constraints. This is what I want to discuss for the rest of the talk. Much of the current work on this topic is based on popular music in Cantonese. The first work I know of along these lines was a paper by Marjorie Chan (1987). By looking at popular songs with multiple verses, Chan showed that Cantonese text-setting appears to be very systematic. Specifically, she showed that the texts set to the same melody in different verses generally have the same linguistic tone sequences even though the words are

different from verse to verse. However, she did not give a general statement of the principles underlying the correspondence between melody and linguistic tone sequence. It wasn't until Wong and Diehl's paper experimental paper on tone in Cantonese songs (2002) that a general rule was proposed.

Wong and Diehl's proposal is essentially the very basic text-setting constraint I mentioned earlier: **if the musical melody goes *up* from one note to the next, the linguistic tone on the corresponding words should not go *down* (and vice-versa).** To apply this constraint, you consider pairs of successive syllables. For each pair of syllables, you see whether the linguistic tone sequence goes up or down in pitch. The hypothesis is that the musical melody across those two syllables should match the direction of the linguistic tone sequence. You also have to make provision for cases where the tone sequence, or the musical melody, doesn't go up or down but stays level. This complicates things slightly, but for the moment let's assume that level should be matched by level, just as up is matched by up and down by down. Oversimplifying, then, we should find the following correspondences[2]:

Tone sequence	**Ideal musical sequence**
L H	**Up**
H L	**Down**
H H	**Level**
L L	**Level**

Here's a different way of representing the same constraint, that allows us to focus more clearly on the predictions it makes about correspondences between linguistic tone and musical melody:

		Melody		
		Up	**Down**	**Level**
Tones	**Up**	*Good*	Bad	Bad
	Down	Bad	*Good*	Bad
	Level	Bad	Bad	*Good*

[2] For simplicity, this example is based on a tone language with only two tones, high and low, as in Schellenberg's work on tonal text-setting in Shona (2009).

Wong and Diehl tested these predictions on a small corpus of Cantonese pop songs. They found that matches in the cells marked 'Bad' were infrequent, meaning that the predictions covered most of the pairs of notes in their corpus. With one adjustment of the constraint on level sequences, they found that they could cover 92% of all cases:

		Melody		
		Up	**Down**	**Level**
	Up	*Good*	Bad	Bad
Tones	**Down**	Bad	*Good*	Bad
	Level	OK	Bad	*Good*

More recent work suggests that it may be possible to refine Wong and Diehl's analysis. Cantonese is normally described as having six tones, and Wong and Diehl's basic constraint actually needs to be supplemented by a classification of the six tones that makes it possible to decide whether a given tone sequence counts as up, down, or level. Specifically, Wong and Diehl grouped the six tones of Cantonese into three classes – high, mid, and low – and defined 'up', 'level', and 'down' on the basis of those classes. More recently, Vincie Ho (2007) proposed a four-way classification – high, mid, low, and extra-low. Last year my student Albert Lo (2013) applied this analysis to a larger corpus of eleven Cantonese pop songs (over 2000 note pairs), and showed that it improves the predictions from 75% with Wong and Diehl's classification to 83% with Ho's. His analysis also suggests a slightly different adjustment for 'Level' sequences:

		Melody		
		Up	**Down**	**Level**
	Up	*Good*	Bad	Bad
Tones	**Down**	Bad	*Good*	OK*
	Level	Bad	Bad	*Good*

*OK only with sequence of high tones

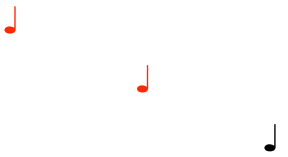
All this shows that a lot of details remain to be worked out, but it suggests that the general principle is valid.

The general principle also appears to apply, to some extent, in other languages. Schellenberg (2009) tested the constraint on a small corpus of songs in Shona. He showed that more than half the pairs of notes in his corpus matched the tone sequence. We find something similar in our corpus of Dinka songs. Nevertheless, he also found (as we do in Dinka) that there are still plenty of cases where the constraint does not seem to hold – that is, there are plenty of pairs of notes that do not match the hypothesized ideal. Putting it somewhat differently, the Wong and Diehl constraint makes better-than-chance predictions, but it is only a statistical tendency. This may reflect the fact that the functional load of tone is considerably lower in Shona and Dinka than in Cantonese.

Alternatively, it may suggest that there are other constraints at work in Shona and Dinka that we haven't figured out yet. It seems pretty clear that other constraints may be at work in other languages. For example, a discussion of Mandarin songs by Lian-Hee Wee (2007) suggests that in Mandarin songs the constraints apply especially to metrically prominent notes or musical downbeats. Similarly, my own current work with my colleague James Kirby suggests that Vietnamese has a system that's broadly like the one in Cantonese, but that for certain purposes we need to consider sequences of **three** syllables, not just two as in Cantonese. In some Vietnamese songs, we find a difference in the way the middle note of three-note sequences like these is treated:

(a) down-up sequence:

(b) downward sequence:



In Cantonese, the first two notes of these sequences (in red) would count as 'Down' and should accordingly allow any 'Down' tone sequence. In Vietnamese, however, the second note of sequences like (a) tends to be set with a low-falling (huyèn) tone, whereas the second note of sequences like (b) tends to be set with a mid (ngang) tone.

Before I finish, I'd like to consider a few broader implications of the text-setting constraints I've been discussing. First, I think it's very significant that the constraints seem to work on **sequences** of syllables or notes. This suggests that the pitch

movement from one syllable to the next is an extremely important aspect of the perception of tone in tone languages, more than the pitch characteristics of individual syllables considered on their own. The aim of the text-setting constraints is to make the musical melodic **contour** match the pitch contour that is created by the sequence of linguistic tones. This is relevant to various issues in tonal phonology and in the psychology of music (e.g. Patel 2008). Of course I haven't got time to talk about these issues now.

The second general implication concerns the relation between tonal text-setting constraints and the kinds of constraints we see in other languages. It's important not to think of such constraints as being unique to tone languages. Specifically, text-setting constraints like this are familiar in European traditions of **metrics** going back at least to Classical Greek and Latin. European systems are based not on pitch but on quantity and stress, but they have the same effect of limiting the choice of which syllables (long/short, stressed/unstressed, etc.) can occur in specific positions in a musical or poetic line. For example, every English speaker knows that a limerick could begin *There was a young lady from Dallas* but not *There once was a young lady from Guadeloupe.* This shows us that many languages, not only tone languages, have a problem matching up the non-segmental aspects of words to the rhythm and melody of music, and it shows us that different languages solve this problem differently depending on their musical traditions and depending on the specific non-segmental aspects found in their language.

***

So let me conclude.

First, we have some kind of answer to the research question we started with. The use of voice pitch for musical purposes coexists with the phonological uses of pitch in a tone language in two main ways: in mostly residual phonetic detail, and in constraints on text-setting. Second, we also have a set of very specific new research questions that could lead us to detailed investigations of how tone fits into music in a variety of different languages. What is the range of possible tonal text-setting constraints? How widely do such constraints differ from language to language? Detailed descriptive work will be necessary to answer these questions satisfactorily.

Finally, we have new, much more general research question that touches on the whole problem of the relationship between language and music: can the structural constraints on tone and melody be integrated with standard theories of poetic metre in a universal account of text-setting? There is clearly a lot of interesting research still to be done.

References

Chan, Marjorie K. M. (1987). Tone and melody interaction in Cantonese and Mandarin songs. *UCLA Working Papers in Phonetics* 68: 132-169.

Deng, Francis Mading (1973). *The Dinka and their Songs.* Oxford University Press.

Ho, Wing See Vincie (2006). The tone-melody interface of popular songs written in tone languages. *Proceedings of 9th International Conference on Music Perception and Cognition. Bologna,* pp. 1414-1422.

Impey, Angela (2013). Keeping in Touch via Cassette: Tracing Dinka Songs from Cattle Camp to Transnational Audio-Letter. *Journal of African Cultural Studies* 25: 197-210.

Liu, Siyun and Samuel, Arthur G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech* 47: 109-138.

Lo, Albert Tsz Chun (2013). *Correspondences between Lexical Tone and Music Transitions in Cantonese Pop Songs: A Quantitative and Analytic Approach*. Honours dissertation, University of Edinburgh.

Patel, Aniruddh (2008). *Music, Language, and the Brain.* Oxford University Press.

Remijsen, Bert and Ladd, D. R. (2008). The tone system of the Luanyjang dialect of Dinka. *Journal of African Languages and Linguistics*, 29: 173-213.

Remijsen, Bert and Adong Manyang, Caguor (2009). Luanyjang Dinka. *Journal of the International Phonetic Association* 39: 113-124.

Rognoni, Luca (2011). The neutralization of the voice quality distinction in Dinka songs: a production and perception study. In *Proceedings of the 17th International Congress of Phonetic Sciences*, vol. 3, pp. 1702-1705.

Schellenberg, Murray. 2009. Singing in a Tone Language: Shona. In *Selected Proceedings of the 39th Annual Conference on African Linguistics,* ed. Akinloye Ojo and Lioba Moshi, (Somerville, MA: Cascadilla Proceedings Project), 137-144.

Wee, Lian Hee (2007). Unraveling the Relation between Mandarin Tones and Musical Melody. *Journal of Chinese Linguistics* 35:128-144.

Wong, Patrick and Diehl, Randy. 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music* 30: 202-209.

Yung, Bell (1989). *Cantonese opera: performance as creative process.* Cambridge University Press.