Gerda Lechleitner (Herausgeber)
Christian Liebl (Herausgeber)

Jahrbuch des Phonogrammarchivs der Österreichischen Akademie der Wissenschaften 4

https://cuvillier.de/de/shop/publications/6785

Beiträge des internationalen Workshops „Relationships of Speech Tone and Music“

Introduction (Jürgen Schöpf)

Background

Relationships of speech tone and music have been intriguing me since my early studies in ethnomusicology and linguistics. Conducting fieldwork in Botswana in 1997 brought me in contact with a Bantu tone language of two tones, but in those days I had neither a matching methodology nor linguistic records that would have allowed me to focus on their relationship to singing and music – a recurring situation that is mentioned, among others, in Catherine Ingram’s contribution. The chance to participate in a DoBeS project in Upper Assam, India, documenting endangered Tai and Tibeto-Burman languages (isolating tone languages with contour tone) with project leader linguist Stephen Morey was a welcome challenge that taught me about the spectrum of possible relationships between speech tone and musics across different cultures. The “Workshop Relationships of Speech Tone and Music” in Vienna in 2012 developed out of the discourse between Stephen Morey and myself, and has become an important mile stone in what I see as an interdisciplinary basic research endeavour. Volumes dealing with this topic in a comparative way have previously been published, for example Thomas A. Sebeok & Donna Jean Umiker-Sebeok (1976) or Bonnie Wade (1993), but none of them targeted it as directly as the present one. The most valuable bibliography on the relation of speech tone and music is currently maintained online by Murray Schellenberg (2013); I had the pleasure of contributing a few items irregularly in the past.

Workshop “Relationships of Speech Tone and Music” (WRSTM)

The “Workshop Relationships of Speech Tone and Music” aimed at scholars from both linguistics and (ethno)musicology to contribute their methodologies, results and theoretical thinking. We had made it a point to provide travel money for students to attend. But in fact only one person in a pre-doc career status applied. In hindsight I believe that this was due to the relationship of speech and music being indeed a very demanding subject, requiring a certain level of expertise in both fields that is usually only developed in post-doctoral careers. The workshop from July 5th to 7th 2012 was rather small with two keynotes and a dozen papers, and an audience of thirty. But this manageable

size contributed to develop a laboratory atmosphere that allowed for an intensive exchange and mutual learning. At the end, I felt that all participants had profited from the workshop – and I would be pleased if this could spread through the present volume.

Review process

Each paper has been reviewed by both a musicologist and a linguist. Linguists were chosen by Stephen Morey, musicologists by myself. The pivotal criterion we had imposed on us was their familiarity with the language or area in question. Given that some contributions are dealing with very small minority languages it will be understood that not for all languages/areas complete matches were available. It is also obvious that in a small field full anonymity is impossible.

It is indeed a challenge working across the two disciplines, musicology and linguistics. Even basic terms like "tone", "pitch", "syllable", "rhyme", "intonation" and more have their specific discourses in both disciplines and they only rarely, or with long delays, inform each other (a major reason to hold the WRSTM). It is also unavoidable that a researcher's methodology is much more reflected, explicit and mature in his or her own discipline than in the respective neighbouring one. We have asked the reviewers to take this into account when reviewing the papers; yet it will always remain disputable what a useful balance should look like in that respect. In one case, a musicologist reviewer clearly recommended not to publish the paper of a linguist, for methodological reasons. We have, however, decided to include this contribution against this recommendation because we value it as a useful standpoint in the discourse with the other papers.

For lack of more compelling criteria I have arranged the papers geographically, starting in the West of Africa and ending in Japan, preceded by Bob Ladd's wonderful keynote. This keynote explains a number of phenomena in the music-speech relationship and their interplay, such as pitch, tone, intonation, contrast of tone and intonation, of grammatical versus lexical tone, and the important notion that tone in speech is not pitch alone, but includes other timbral parameters (very strikingly exemplified later on by Morey). Ladd's twofold concept of "top-down cues" and "phonetic residue" indeed provides a helpful tool to approach any tone language music relationship. A key to its understanding is that both provide redundancies in different ways to keep up intelligibility even though certain features of language are not represented in the singing. Moreover, I see the principles known from Gestalttheorie to be at work here for maintaining meaningfulness, such as the principle of "proximity", "similarity", and "common fate". It might be appropriate to add that music also comes with constraints that project onto language, e.g. by concepts of scales, an idea I have not yet seen reflected in linguistic studies.

“Mind the gap”

An interesting point becomes visible when arranging the articles geographically: the larger areas where tone languages occur (Africa, South East Asia, Middle America) appear mutually exclusive with areas where microtonal scale systems have developed (the Islamic world between Morocco and India). Although this is a very superficial observation I believe that tone languages profit from pitch contrasts beyond microtonic intervals to function efficiently (compare Vidal, quoted after Villepastour in this volume, saying “that spoken Yoruba would not normally place neighbouring speech tones less than a whole-step apart”). Whereas seen from the music end, a concept of musical scale that provides ample opportunities for microtonic expression in performance (systematically as a scale, not as ornaments) may not easily use the same perceptory parameter of pitch to develop the categorical (binary) contrasts needed to disambiguate meaning in language.

Amanda **Villepastour** has recently (2010) published an impressive work on Bàtá drumming, and in her present paper discusses methodological questions that I found particularly interesting, especially how the discourse in Yorùbá interacts with researchers. The concept of “do-re-mi-consciousness” is indeed an inspiration and calls for a subtle critique of our sources in this complex topic. She also reminds me that a plausible explanation is not necessarily a true one. Also, the often quoted continuum of speech over chant to song (quoting List, Bright and Agawu, see for example Ingram in this volume) is challenged by her data rather suggesting a continuum between speech and song, the latter exaggerating language features, whereas chant represents a performance practice less connected to both others. **Schellenberg,** at the far (eastern) end, has also recently been active in this area of research – see *Ethnomusicology* 56/2 (2012) – with a thesis in sharp contrast to Villepastour’s, which triggered very useful discussions at the WRSTM, to the end – as I see it – that the cultural relative as a research paradigm in ethnomusicology can make it difficult to argue with linguistics that is used to working with much more universal claims in its paradigm.

David **Locke**’s long research experience in West Africa allows for a broader and insightful overview more in qualitative terms than in detailed examples, including an interesting literature overview on the topic of West African music traditions. Concepts of the music of language (=drumming) and the language of music (= vocables) beautifully complement each other. That his analysis of drumming reveals three tone levels whereas the phonetics of the language distinguishes two should be of interest to linguists. Vanna **Crupi**, working in Uganda, brings us back to detailed analysis and field experiments showing current analytical methods as well as their limitations. Connecting “downdrift” and “descending melodies”, two old concepts of

linguistics and musicology respectively, with a conceptual background of neuroscience may be as instructive about the history of disciplinary gaps as it helps to understand her research data.

Jumping the (microtonal?) gap to Asia, **Morey** reveals detailed conceptions of a tone system that includes at least as many timbral qualities as actual pitches, and brings the historical depth of tonal changes into view. Other important lessons which I myself, as a musicologist, learned from him and in the Tai Phake community are that their speech tone system in its application to singing is hierarchical, resulting in the observable fact that syllables in rhyming position are treated differently than in non-rhyming position. In addition, a representation of contour tones appears more important than that of level tones. And not to forget: all these factors can lead to very different results for different genres within the same culture. **Schwörer**'s rich field work experience in south-east Asia allows to present an example of a very close relationship of text and instrumental music, including the perhaps surprising fact that this surrogate speech is not comprehensible to all members of the speech community. This finding is supported by **Ingram**'s data. Her analysis reveals the huge task that emerges once a language is not well documented, and, indeed, the methodologies available for researching speech tone music relationships are scarce. In her case, the pedagogical process is a key to access this relationship for Kam, which I believe is an important contribution to the methodology.

The collaborative work of **Karlsson, Lundström, Svantesson and Tuttle** is, in small, what the whole volume attempts: searching interdisciplinary methods and issues in the speech tone music relationships, and bridging linguistics and musicology. A tonal and a non-tonal dialect of Kammu are analysed in contrast, and Siri Tuttle, an authority on Athabascan, brings in her expertise on this native American language, the only example from the Americas in this volume. A problem that occurs across tone language cultures is how to separate tone from intonation, a problem already addressed by Svantesson in the past.

Jähnichen digs through amounts of data revealing a number of mechanisms in place between language and music in Vietnam, and is repeating Schwörer's find of surrogating speech of a musical instrument, this time a lute.

Schellenberg explains his experimental methodology in more detail than in previous publications, contrasting Mandarin and Cantonese in laboratory experiments, using specifically composed songs, and on the way explaining a couple of linguistic concepts such as intrinsic pitch of vowels, influences of consonants in front of vowels, and intrinsic vowel duration. The journey is completed by **Fujita**, working on Japanese where the rhythmic

component in language is of major interest. This parameter, although not directly discussed during the WRSTM, of course plays an important role, which is also touched on by Karlsson, Lundström, Svantesson and Tuttle.

Recurring issues

Overall, a number of recurring issues emerge. In my reading it is obvious that it is still a big methodological challenge to research the speech tone music relationship. This is explicit or implicit in most contributions, and due to different reasons. A major and structural one is the disciplinary gap between musicology and linguistics (I imagine that this gap would not exist had the humanities been developed by native speakers of Yoruba or Vietnamese). The theoretical paradigms of more universalist tendencies in phonetics and phonology on the one hand (claims like the "International Phonetic Alphabet", or the simplifying assumption that all speakers of a language share the same knowledge of that language), and a rather cultural relativist paradigm in ethnomusicology on the other, take their toll, making it difficult to understand each other over this disciplinary gap. A second challenge is that linguistic data on tone systems of languages is still scarce or difficult to access for musicologists, and sometimes it is even controversial; and a third one is that linguists have not yet discovered musicological studies as sources of information that can help them support their findings, e.g. how scale definitions and tone systems might connect. As long as only one side sees the opportunity it is difficult to develop cooperative project proposals.

In the present volume, I believe, the enthusiastic laboratory atmosphere from the WRSTM has matured into a batch of contributions that collectively improve our methodologies considerably. I hope that this topic will become increasingly addressed in both disciplines as methods advance and get exchanged more frequently. I would be pleased to be able to further this through present and future work.

Acknowledgements

A big word of thanks is appropriate to a number of institutions and people: the Phonogrammarchiv and the Institute for Corpus Linguistics and Text Technology, both part of the Austrian Academy of Sciences; the VolkswagenStiftung, and the Österreichische Forschungsgemeinschaft, who all contributed infrastructure and finances to the Workshop Relationships of Speech Tone and Music. Thanks to (now former) Präsident Helmut Denk of the Austrian Academy of Sciences, who inaugurated the WRSTM, and to the staff of the Phonogrammarchiv for their assistance during the workshop; a special thanks to former director Rudolf Brandl for supporting my

interest in this subject early on, and my personal congratulations on his 70th anniversary; thanks also to his successor, current director Helmut Kowar. Big thanks to Stephen Morey, who from the beginning developed the idea of the workshop with me and assisted with his experience in designing and running an appropriate review process. Thanks to my co-editors Gerda Lechleitner and Christian Liebl, especially for their patience in an endeavour continuously exceeding every deadline. Thanks to Elke Salzer for the layout. And, of course, to all authors (listed at the end of the volume) and the reviewers listed below:

Wolfgang Bender, Center for World Music, Hildesheim, Germany
Roger Blench, Kay Williamson Educational Foundation, Cambridge, UK
Peter Cooke, Professor Emeritus, University of Birmingham, UK
Jerold Edmondson, University of Texas at Arlington, USA
Catherine Falk, Melbourne Conservatory of Music, Australia
Ingrid Fritsch, Universität zu Köln, Germany
Nerida Jarkey, Sydney University, Australia
Randy La Polla, Nanyang University, Singapore
Debbie Loakes, University of Melbourne, Australia
Andreas Meyer, Folkwang Hochschule, Essen, Germany
Terry E. Miller, Professor Emeritus, Kent State University, Ohio, USA
Martin Mous, Leiden University, The Netherlands
Knut Olawsky, James Cook University, Townsville, Australia
Sophie Salffner, SOAS, London, UK
Jürgen Schöpf, Phonogrammarchiv, Austria
Paul Sidwell, Australian National University, Canberra, Australia
Scott Swan, Florida State University, USA
Marija Tabain, La Trobe University, Melbourne, Australia
Mike Tian Qiao Lu, James Cook University, Townsville, Australia
Trang Quang Hai, CNRS, Musée de l'Homme, Paris, France
Bell Yung, Professor Emeritus of Music, University of Pittsburgh, Pennsylvania, USA

References

Schellenberg, Murray. 2013. "Singing in Tone Languages - Partial Bibliography". Online: <http://www.linguistics.ubc.ca/singing_tones_biblio> (20/12/2013).

Sebeok, Thomas A. & Donna Jean Umiker-Sebeok. 1976. *Speech Surrogates: Drum and Whistle Systems*. 2 vols. The Hague & Paris: Mouton.

Wade, Bonnie C. 1993. *Text, Tone, and Tune*. New Delhi etc.: Oxford & IBH Publ.

2. Keynote

Singing in tone languages: An introduction to the kinds of things we might expect to find

(D. Robert Ladd)

This paper has two main goals. First, it gives a brief introduction to tone languages, aimed in particular at ethnomusicologists who are aware of the existence of such languages but are uncertain of exactly what it means to describe a language as tonal. Second, it sketches the kinds of interactions we might expect to find between musical pitch and linguistic pitch when people make music in a tone language. In addressing this second goal I draw extensively on our ongoing study of song in Dinka, a major language of South Sudan (<http://www.lel.ed.ac.uk/nilotic>).

The use of pitch in language

All spoken language, unless it is whispered, has pitch. Roughly speaking, the main physical basis of what we perceive as pitch is the fundamental frequency of regular vibrations, and it is physically impossible to have voice (vibrations of the vocal cords) without having a fundamental frequency (the rate at which the vocal cords are vibrating). In speech, the vocal cords vibrate at some (usually rapidly varying) frequency, and at the same time the tongue and lips and other articulators make rapid movements to produce vowels and consonants. The two systems operate fairly independently, so that a speaker can produce pretty much any syllable with any pitch and vice versa. This gives us some basis for thinking about the stream of vowels and consonants as being something separate from the pitch of the voice, and to some extent that separateness applies in all languages. For example, a voice raised in anger (where, among other things, the high pitches are higher) sounds very similar in any language.

However, all languages also use pitch for specifically linguistic purposes as well. For example, many languages distinguish questions from statements through pitch alone. The Italian sentence

(1) *Maria non viene* 'Maria not come-(3rd-person-singular)'

can be either a question meaning 'Isn't Maria coming?' or a statement meaning 'Maria isn't coming', depending almost exclusively on the pitch contour that accompanies the words. Such uses of pitch at the sentence level are usually called **intonation** (obviously, this term means different things in linguistics and in music). It is important to emphasise that intonation is *not* like the voice raised in

anger – intonation definitely varies from one language to another, although there are certain widespread properties. For example, consider the Hungarian sentence

(2) *József mérnök*, lit. 'Joseph engineer'

If it is spoken with a rise in pitch on *József* and a fall in pitch on *mérnök*, it means 'Is *Joseph* an engineer?' – that is, it is a question with the emphasis on *József*. To speakers of English or German, however, it sounds like a statement with the emphasis on *mérnök*, something like 'Joseph is an *engineer*'. Unless you know Hungarian, it is difficult to know which combination of emphasis and questioning or asserting is being expressed by the intonation here. For a general introduction to intonation see e.g. Cruttenden (1997) or Ladd (2008: chapters 1 and 2).

It is fairly common to categorise languages according to whether they use pitch at the sentence level only – as in the Italian and Hungarian examples we have just looked at – or also at the word level. The best-known example of a language that uses pitch at the word level is Chinese. Consider the following pair of Chinese[1] words:

(3) a. *tāng* 'soup' [moderately high level pitch]
b. *táng* 'sugar, sweets' [mid-to-high rising pitch]

Phonetically the difference between *tāng* and *táng* is just a matter of pitch, like the difference between a question and a statement in Italian or Hungarian. But functionally, it is obviously very different: in Chinese, the difference in pitch changes the meaning of the word, not just the general force of the sentence as a whole.

Pitch used at the word level is generally called **tone**, and languages that use pitch at the word level are generally called tone languages. Broadly speaking, there are three major areas of the world where tone languages are found: East and Southeast Asia, including New Guinea but not Australia; pretty much all of Sub-Saharan Africa; and various parts of the Americas, especially southern Mexico and the Amazon. Something like half the languages in the world are tone languages, and something like half the people in the world speak a tone language, so this is hardly a marginal phenomenon. But since the very idea of such a language strikes most speakers of European languages as outlandish, it is important to spend some time explaining a bit more about how they work. Further detail can be found in such works as Gussenhoven (2004) and Yip (2002).

1 Here and throughout the paper the examples from 'Chinese' are drawn from the standard language, often called Mandarin.

Basic structural properties of tone languages

It is surprisingly difficult to come up with a definition of a tone language that commands general agreement. On the other hand, it is not hard to identify a set of languages that everyone will agree on classifying as tone languages, a set of languages that everyone will agree are non-tonal, plus a set of languages that, for one reason or another, fall somewhere on a scale between definitely tonal and definitely non-tonal. I begin by describing the properties of languages that are definitely tonal.

The most important point is that pitch in tone languages is not some kind of overlay or musical supplement, but *an integral part of the word*. In order to talk about word forms in a tone language, it is not enough just to talk about the consonants and vowels; it is necessary to talk about the pitch and sometimes the voice quality as well. Let us go back to the Chinese example in (3). For a speaker of a European language, it is very tempting to talk about 'the word *tang*' and say that if you pronounce it one way it means 'soup' and if you pronounce it another way it means 'sugar'. But that is similar to looking at a pair like English *sweets* and *tweets* and concluding that the way you pronounce the first consonant makes a difference to the meaning of the word. Most English speakers would say 'No, *sweets* and *tweets* are different words', and that is exactly what Chinese speakers would say about *tāng* and *táng*. It is important to get used to the idea that in tone languages, pitch can work *just like* consonants and vowels for distinguishing one word from another.[2]

Tonal distinctions can signal various kinds of meaning differences. The most basic kind of difference is the **lexical** difference just illustrated: *tāng* and *táng* have a different tone and are different words. But tone can also make a **grammatical** difference. In many languages tonal differences are used to signal the difference between singular and plural, or between present and past, or other such grammatical distinctions. In Dinka, for example, we get singular-plural pairs like the following:[3]

(4) a. *kët* [L] 'shoulder'; *kët* [H] 'shoulders'
b. *joth* [L] 'ring'; *joth* [H] 'rings'

On the basis of these two cases only, it might seem that we can talk about 'the word *kët*' or 'the word *joth*' and treat the pitch difference – low in the singular, high in the plural – as something added. But if we look at Dinka grammar as

2 It is not just pitch that can work this way; in some languages voice quality can have the same kind of effect. In Dinka, *roor* (with modal or creaky voice) means 'forest' and *röör* (with breathy voice, indicated by the dieresis on the vowel letter) means 'men'. However, in this paper I concentrate on pitch.

3 The examples are given in Dinka orthography, which does not indicate tone. Tone is specified here as low [L] or high [H]. The dieresis on a vowel indicates breathy voice quality; cf. the previous footnote.

a whole, we see that the change in these forms is more like the difference in English between *woman* and *women* or *mouse* and *mice* – in both cases what we are doing is *modifying* the word, and the *modification itself* is the signal of the grammatical category. Plural marking in Dinka is almost always a question of changing some part of the word, and it can be almost any part, not just the tone. This can be seen in these further examples:

(4) c. *dhɔk* [L] 'boy'; *dhäk* [L] 'boys'
d. *meth* [L] 'child'; *mïth* [L] 'children'
e. *nɔk* [H] 'feather'; *nak* [L] 'feathers'
f. *jïc* [H] 'ear'; *jïth* [L] 'ears'

Here we see: pairs where the vowel is different, like (4c); pairs where the vowel and the voice quality are different, like (4d); pairs where both the vowel and the tone are different, like (4e); and pairs where the final consonant and the tone are different, like (4f). Once again, in short, tone is behaving exactly like vowels and consonants.

Now let us consider the kinds of **phonetic** distinctions that are involved in tone. A very common traditional distinction is the difference between **level** tones and **contour** tones. In the case of level tones, the idea is that the pitch on each syllable is relatively unchanging, so that the linguistic distinctions are between categories like high and low – as in the Dinka examples just given. With contour tones, the distinctions are between categories like rise and fall. As a very broad generalisation, the tone languages of East and Southeast Asia involve contour tones, whereas the languages of sub-Saharan Africa involve level tones. As a result of this broad typological difference, the descriptive and notational conventions of linguists and anthropologists working in those two areas are also rather different.

A brief illustration will suffice. Here are the tones of Chinese, along with the standard tone diacritics.

(5) a. *fū* [high level] 'hatch [verb]'
b. *fú* [mid-high rising] 'submit'
c. *fǔ* [low dipping-rising] 'rot, ferment'
d. *fù* [high-low falling] 'pair, set'

What can be seen here is that the diacritics are a rough visual indication of the pitch pattern across the syllable. This is fairly typical of notation schemes for the contour tones of the tone languages of East and Southeast Asia. It is also fairly typical that among the contours there are one or more levels; the Chinese tonal inventory, for example, includes a high level tone in addition to a rise and a fall and a sort of low dip. The Chinese examples can be compared with a lexical set from Yoruba, a major language of Nigeria.

(6) a. *īgbā* [mid – mid] ‘200’
b. *īgbá* [mid – high] ‘calabash’
c. *ìgbà* [low – low] ‘time’
d. *ìgbá* [low – high] ‘[type of tree]’

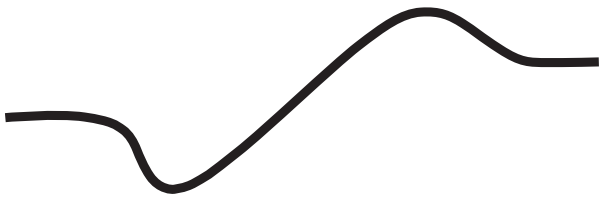
In writing and transcribing African languages the acute and grave accents tend to be used to indicate high and low, and if a language also has a mid tone, which Yoruba does, it may be marked like in the way illustrated in (6a) and (6b), or just left unmarked. Ethnomusicologists working in different parts of the world may need to be aware that these different descriptive conventions exist; tone marks may mean different things in descriptions of different languages.

It is also worth mentioning that in many African and American tone languages, especially in West Africa and Mexico, even though the system is essentially based on level tones, there are many cases where contour tones occur phonetically but count as a combination of two levels. For example, consider the following Yoruba sentence:

(7) a. Ayọ̀ lọ ‘Ayo [man’s name] is coming’

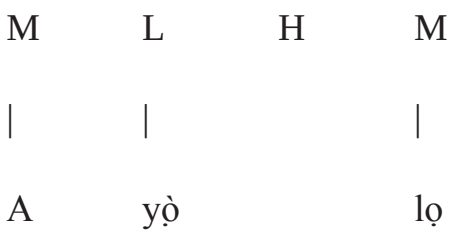
The name *Ayọ̀* has a sequence of a mid and a low syllable, while the verb *lọ* has a mid tone. However, the grammar requires a high tone that (very approximately) marks the end of a subject noun phrase, and this high tone attaches itself to the second syllable of *Ayọ̀*. This gives a low-high sequence on that syllable, which is pronounced as a steep rise before going back down to the mid-level for the verb *lọ*, so that the pitch contour on the whole utterance looks roughly like this:

(7) b.



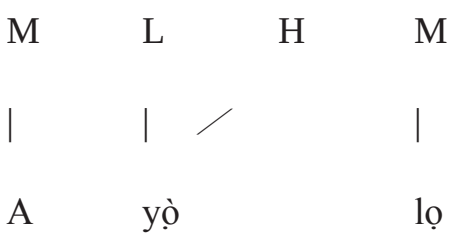
Tonal phenomena like this are the basis of so-called autosegmental notation that is often seen in linguistic descriptions. In an autosegmental analysis, the tones form one sequence of units and the syllables form another sequence. Some of the tones come preattached to syllables, and some do not; the tones that do not come preattached, like the H tone that marks the end of the subject in Yoruba, have to line themselves up according to various principles. That is, we start with a structure like this:

(7) c.



and then the 'post-lexical' attachment of the subject-final H tone gives us this structure:

(7) d.



which yields the pitch rise on *yọ̀*. The details vary to some extent from language to language, and there are lots of theories about the general structural principles that govern the association of tones to syllables. Unfortunately, I can do no more than mention the existence of autosegmental notation here; for further detail the reader may consult Goldsmith (1990).

'Pitch accent' and other problems

There are many other linguistic aspects of tone that could be discussed here – whole books have been written about tonal phonology – but the foregoing brief introduction will have to do. Before moving on to singing, however, it is important to mention why there are borderline cases – what it means for a language to be hard to classify between tonal and non-tonal. The basic source of the uncertainty is the existence of accent.[4]

Many languages, including English, pick out one syllable of a word or phrase for what I will refer to informally as 'special phonetic treatment' of some sort. This is a good rough-and-ready definition of accent. Just like tone, accent can depend on both lexical and grammatical factors. In Italian, for example, we have pairs like the following:

(8) a. ***an****cora* 'anchor'; *an****co****ra* 'again, still'
b. ***prin****cipi* 'princes'; *prin****ci****pi* 'principles'

Here the difference in accent placement (indicated by graphically emphasising the accented syllable) distinguishes two different words. There are also pairs like the following:

4 This section draws especially on ideas from Beckman (1986) and Hyman (2006, 2009); see also Himmelmann & Ladd (2008).

(9) a. ***par****lo* ‘I speak’; *par****lò*** ‘s/he spoke’
b. ***la****vati* ‘wash! [imperative];
*la****va****ti* ‘washed’ [masc. pl. past participle]

In these the different accent placement signals the grammatical difference between two forms of the same verb.[5]

Also, as with the difference between tone and intonation, accent can be used at the sentence level as well as at the word level. This is most easily seen in English:

(10) a. I **thought** he was coming.
b. I thought he was **coming.**
c. I thought **he** was coming.

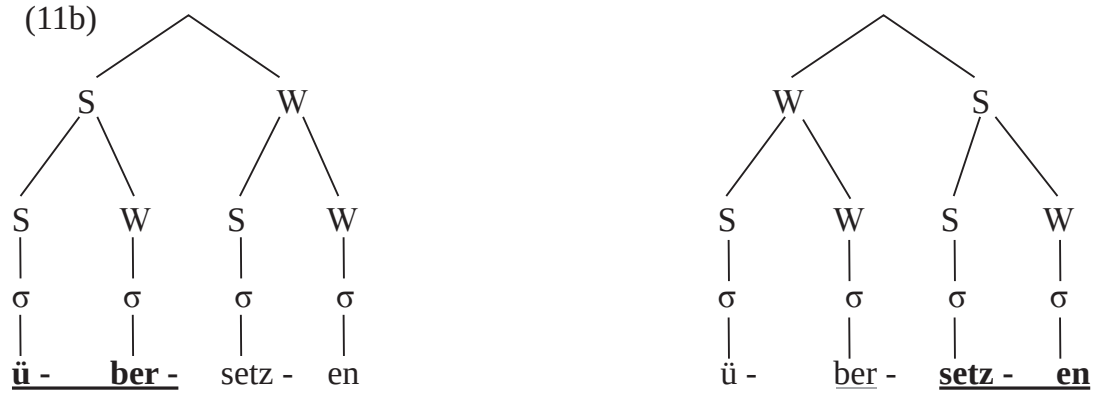
These three sentences involve the same words and in some sense have the same propositional meaning, but would be used in different circumstances: 10a would likely be used when the speaker has just had confirmation that ‘he is coming’, whereas 10b would likely be used when the speaker has just learned that this is not true after all, and 10c when the speaker learns that someone else is coming instead. Finally, in some languages accent is organised hierarchically, so that there are metrical structures that are reminiscent of music. This is very definitely true of English and German. In German for example, there is a distinction in pairs like the following:

(11a) ***über****setzen* ‘take across’; *über****setzen*** ‘translate’

At one level of analysis, this distinction involves the relative prominence of the prefix *über* and the stem *setzen*. But within both of those component parts – at the next lower level, as it were – there are also differences of relative prominence, with the first syllable in each case being more prominent than the second. This nesting of relative prominence can be represented graphically as a ‘metrical tree’, as follows (the symbol σ stands for ‘syllable’; *s* and *w* stand for ‘strong’ and ‘weak’ respectively):

5 For the sake of academic rigour, two things should be pointed out here. In (9)a, the diacritic on the final syllable of *parlò* is part of standard Italian orthography and indicates that the final syllable is accented; accent on other syllables is not normally indicated in spelling. In (9)b, the imperative form actually has a different internal structure from the past participle form, in that it is a sequence of the imperative *lava* and the second person reflexive pronoun *ti*; on the surface, however, the only difference is the placement of the accent.

(11b)



This kind of structure is central to accentual patterns in English and German, but it may be a relatively unusual phenomenon cross-linguistically. The basic metrical idea is due to Liberman (1979); for more on metrical phonology see e.g. Hayes (1995). In any case, the essence of accent is taking a syllable and singling it out for special phonetic treatment – or, if there is hierarchical structure like that shown in (11a), it is a matter of singling out one element in a larger group, at multiple levels. That much seems to be a valid cross-linguistic generalisation. What can vary a lot from one language to another, however, is the actual phonetic nature of the special treatment that signals which element is singled out. In Italian and English and German, the special treatment involves what is usually called **stress** – the accented syllable is pronounced with greater force of articulation, it is often longer in duration, and there are a variety of pitch phenomena connected with the way intonation patterns are associated with words. In other languages the special treatment may be different, and in particular, it may be primarily or exclusively a matter of pitch. Moreover, in languages where the accent is signalled mainly by pitch there is often no hierarchical organisation of the kind we just saw in English or German, but only a syllable-by-syllable specification: accented or unaccented. Languages like this are sometimes called 'pitch accent languages', but there is no general agreement on the boundary between 'pitch accent' and tone.

The problem is easily seen in Standard Japanese. The main phonetic manifestation of accent in Japanese is a drop in pitch from the accented syllable to the following syllable. The following is a standard example:

(12) a. ***ha****si desu* 'they're chopsticks'
b. *has**i** desu* 'it's a bridge'

In ***ha****si* 'chopsticks' the pitch drops from the first syllable to the second, whereas in *has**i*** 'bridge' the pitch drops from the second syllable to the first syllable of the next word. This still seems easy to distinguish from tone – and it still seems a lot like English or German or Italian, except for the

phonetic detail. However, the problem is that Japanese also has words *with no accent*. That is impossible in English or German or Italian – in any word of more than one syllable, one of the syllables has to be more prominent than the others. In a word with two syllables, like English *permit* or Italian *parlo/parlò*, there are only two possibilities – either the first syllable is accented, or the second syllable is accented. But in Japanese there is a third possibility:

(13) c. *hasi desu* 'it's an edge'

Phonetically, this is distinguished from (13) b. by the fact that there is no pitch drop during or after the unaccented word *hasi* 'edge'.

When we consider these two characteristics – the existence of unaccented words, and the fact that accent, when it is present, is manifested mostly by pitch – we suddenly see that Japanese seems rather different from English or German or Italian. There are quite a few languages that have accentual systems of this general sort. In these cases, some people emphasise the similarity to accentual systems as in English or German, and others emphasise the similarity to tone languages. There is definitely a scale or continuum, and this is why a given language may be classified both as 'tonal' and 'non-tonal' by different researchers. There is still no generally accepted theory of how to fit pitch accent systems into a valid typology.

Singing in a tone language

Now let us consider the kinds of things that we might expect to find when people make music in a tone language. People who speak non-tonal languages tend to assume that there is a fundamental problem here, but any problem is entirely analogous to other situations in which language is used in special ways. For example, when people ask me 'how do they sing in a tone language?', I always want to respond with something like 'how do you whisper in a language with voicing distinctions?'. In many languages, including almost all European languages, there are consonant distinctions based on the presence or absence of voicing. The difference between [s] and [z], for example, is a difference of whether the vocal cords are vibrating or not.[6] When we whisper, the vocal cords are not vibrating. Yet it is not hard to understand a whispered sentence like *Sue went to the zoo* or *There were fleas in the fleece.*

6 This difference is readily observed by pronouncing [s] and [z] alternately, while either (a) placing the thumb and forefinger on either side of the Adam's apple to feel the vibration, or (b) blocking the ears to hear the vibration directly through the bones of the neck and head.