On tune-text relations

Carlos Gussenhoven

Centre for Language Studies, University of Nijmegen
c.gussenhoven@let.kun.nl

ABSTRACT

Relations between tune and text are to be separated into phonological relations and phonetic factors affecting the timing of pitch event. Two phonological relations are 'association', a relation between tones and syllables or moras, and (phonological) 'alignment', the right-adjustment or left-adjustment of a tone with a prosodic constituent or another tone. Phonetic relations are best divided into within-speaker and across-speaker variation in the timing of pitch events. Across-speaker variation includes phonetic differences found between languages and varieties of the same language. Within-speaker effects can be divided into three groups, mechanical effects of the speech production mechanism, ergonomic effects, which are due to context, and communicative effects, which are due to paralinguistic communication.

1 INTRODUCTION

Phonological relations are part of the programme of instructions presented to the speaker's phonetic implementation module, to be processed by the articulators with the aim of creating a suitable acoustic signal. These phonological relations include Goldsmith's 'association' as well as the 'alignment' of Optimality Theory. By contrast, the phonetic relations arise as a result of the way in which the implementation module works. In particular, they depend on what articulatory anchorings exist, on physical limits to the speed at which an intended gesture can be started and executed, and on ergonomically or communicatively inspired adjustments made by speakers.

2 PHONOLOGICAL RELATIONS

2.1 Association

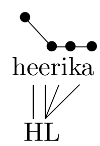
The concept of 'association' arose as a way of stating where autosegmentalized tones were pronounced and was motivated by the independent behaviour of tones and vowels [1]. Instead of the one-to-one relation between tones and vowels that was implicit in a segment's representation as a list of features, as in (1), relations could, in addition to one-to-one (2a) be many-to-one (2b) and one-to-many (2c).

(1) [+hightone, −contour, −cons, +son, −back, ...]

(2) a. V V | | T T   b. V /\ T T   c. V V \/ T

Unlike (1), the representations in (2) predict that deletion of vowels with retention of their tones ('tone stability') or deletion of tones with retention of vowels ('tone deletion') should be common, and that morphemes that consist solely of tones ('floating tone', a usage which is now obsolete) or morphemes that consist solely of vowels and consonants ('toneless morphemes') are to be expected. Tone deletion occurs in the second constituents of compounds in Barasana, causing the last tone of the first constituent to associate with the vowels of the second [2]. Thus, if héè and rìkáà come together to form a compound 'tree fruits', only the HL-tones of the first constituent remain, with the L spreading right, as in (3).

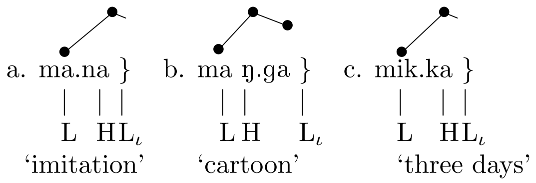
(3)



The element in the segmental string with which tones associate, the Tone Bearing Unit (TBU), is the syllable or the mora. Japanese uses the sonorant mora. (4) shows citation pronunciations of three unaccented words, which begin with an initial LH and end with a final L-tone. In (4a), the H-tone associates with the second syllable, as the single sonorant mora of the first syllable is occupied by L, but in (4b), the H associates in the first syllable, which contains two moras. The evidence that it must be the *sonorant* mora is (4c), where the second mora of the first syllable is obstruent [k], the first half of a geminate. Although its first syllable is bimoraic, (4c) behaves like (4a), not (4b). As a result, the pitch rises from first to second syllable in (4a,c), but falls in (4b). For lack of TBUs in (4a,c), the final L-tone cannot be realized and may at best be observed as a truncated downward movement. (The monosyllabic rise in (4b) is pronounced as a high level pitch by many speakers [3].)

ISBN 1-876346-48-5 © 2003 UAB

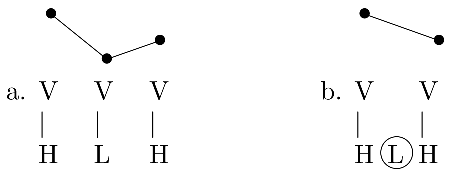
(4)



Representations do not require full association between all tones and all TBU's. It has been argued on the one hand that TBU's need not be associated with tone ('phonetic underspecification'), and on the other that tones may float, i.e., exist in the representation without association ('floating tone', in current usage). The *locus classicus* for phonetic underspecification is [3] (p.263). Japanese Accentual Phrases, which may be accented or unaccented as determined by the morphology, typically begin with a pitch rise. An accented Accentual Phrases has a subsequent sharp fall from one of its syllables, while the post-rise pitch of an unaccented Accentual Phrase slowly descends, towards the end of the utterance or the beginning of the next Accentual Phrase. The crucial point is that the slope of this fall is determined by the length of its Accentual Phrase. For this reason, it cannot be the case that the H-tone of the rise associates with following TBU's, since such multiple association would predict that the pitch remains high, or is at least independent of how many TBU's there are. Rather, the slope results from an interpolation between the target of the H-tone and the target of the following L-tone.

Foating tones have been used to account for two phenomena. In one case, the effect of a tone is observable in the behaviour of following tones, but for lack of a target is not itself observable as a tone. This is shown in (5).

(5)



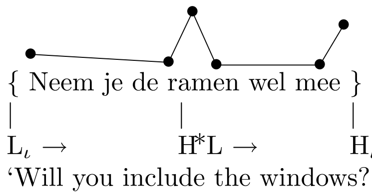
In (5a), the second H-tone is lowered as a result of the preceding HL ('automatic downstep' in Stewart's terminology [4]), while in (5b) the second H-tone is scaled identically to the second H-tone of (5a) ('non-automatic downstep'). Moreover, the effect on the remaining H-tones in the utterance is identical, too, in that once a H-tone is lowered, it imposes an upper limit on the pitch of following H-tones. The postulation of a 'floating' L, as in (5b), explains these facts without further ado [4, 5]. Downstep in English intonation has been described in just this way [6].

The second use to which floating tones have been put is for tones whose targets are particularly variable, as is often the case in intonation languages. The timing of the target of the L* of a pitch accent like L*+H is more reliably described with reference to the segmental string than that of H, whose location may be in the same syllable or in the third syllable after the stressed syllable, depending of the structure of the text [7]. Thus, assuming it is followed by final low pitch, H will create a peak in the middle of the syllable *JOHN!*, but in *CINnamon flavour!* it will be located in *-na-* or *-mon-*. Similar facts are documented for Swedish [8].

2.2 OT Alignment

If floating tones receive a target there must be a representation to determine their location. In Optimality Theory, 'alignment' is used to locate phonological or morphological elements in the linguistic structure. When an element aligns with some other element, its right or left edge coincides with the right or left edge of the constituent it aligns with [9]. Thus, a right-hand boundary tone of the intonational phrase (ι) aligns its right edge with the right edge of the ι. Whether or not aligned tones associate in their location is an independent issue. In (6), from Dutch, the initial L$_\iota$ is left-aligned with the left edge of the ι, the final H$_\iota$ with the right edge, while the left edge of the 'trailing' L is left-aligned with L*.

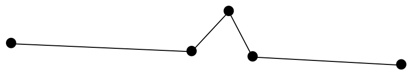
(6)



As illustrated by (6), tones may align *both* right and left. The initial boundary tone creates two targets, while also the trailing L-tone has an early and a late target. If these are far apart, they may show the effect of declination. Alignment with opposite edges accounts for the way tones fill up the space made available by the segmental string in tonally sparse representations [10]. It is distinct from multiple association ('spreading'), which typically leads to sustained F_0 target. Either or both alignments may have an association. The difference is illustrated by the dialect of Roermond, which has a privative lexical tone contrast. Syllables with two sonorant moras without lexical tone (7a) (= Accent 1), those with two with a lexical tone on the second mora (7b) (= Accent 2), and those with one sonorant mora (7c) behave differently. In declarative intonation, H* associates with the first mora of the stressed syllable of the focused word, and L$_\iota$ closes the ι. The fall for Accent 1 is completed inside the Accent-1 syllable in (7a), due to the association of L$_\iota$ with the second TBU in the accented syllable. However, neither in (7a) nor in (7c) is there a free TBU available, and as a result association fails and the falls are slower. Boundary tone association in various languages is discussed in [11], who assigned the term 'phrase accent' to them.

ISBN 1-876346-48-5 © 2003 UAB

(7) a.

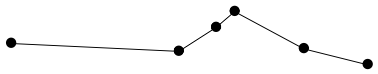


{ Hae waas vorige maondj in Milaan }

$L_\iota \rightarrow$ H* L_ι

'He was in Milaan last MONTH'

b.

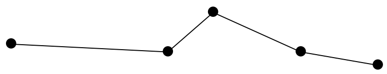


{ Hae waas vorige waek in Milaan }

$L_\iota \rightarrow$ H*H $\leftarrow L_\iota$

'He was in Milaan last WEEK'

c.



{ Hae waas vorige nach in Milaan }

$L_\iota \rightarrow$ H* $\leftarrow L_\iota$

'He was in Milaan last NIGHT'

3 PHONETIC RELATIONS

Each tone is translated into one (or two, see above) instructions to reach a time-specified ('aligned') F0-value, or range of values (the latter being a 'dynamic' as opposed to a 'static' target [12]), and to interpolate between these across the voiced sections of the signal. There is a multitude of factors that have been shown to affect the timing of pitch events, also known as 'alignment'. (Both meanings of 'alignment' are current). First, there is variation across speakers with different language backgrounds (section 3.1). Second, there are three types of within-speaker factors, automatic effects, ergonomic effects, and communicative effects (sections 3.2-3.4).

3.1 Across-speaker variation

The same phonological representation will have different implementations in different languages [13]. Ladd (this symposium) reports on detailed timing of pitch events that are likely to be language or variety specific. Thus, the duration of prenuclear rises in Greek is co-extensive with the accented syllable plus the onset of the next. This is quite different from the situation described for English and Swedish at the end of section 2.1. Also, different varieties of German may time the 'same' pitch events differently. Such data suggest that in the phonetic implementation, pitch gestures are timed with reference to various elements in the phonology, but that there is some freedom in choosing these synchronizations, or 'landmarks'. It may not always be clear whether contours in different varieties differ enough to have different phonological representations [14].

3.2 Mechanical effects

Mechanical effects are inherent to the execution of the instruction to reach a target. I am aware of two aspects, mainly thanks to Yi Xu's work (cf. this symposium). One concerns the presumed need to anchor the execution of a muscular gesture intended to create an F_0 effect to other muscular gestures involved in speech production, which might either be due to a requirement for phased muscular activity or phased innervation. The second concerns temporal restrictions on the execution of the gesture. Such restrictions might firstly be due to a non-negligible time lag between the innervation of the muscle(s) and the execution of the gesture, and secondly may concern the acceleration or deceleration of vocal cord vibration rates, or their reversal, which must be limited by inertia.

3.3 Ergonomic effects

Ergonomic effects are due to the speaker's monitoring of the quality of the signal and its production efficiency. The desire to place targets safely within sonorant stretches of speech is responsible for the effects of segmental structure on pitch alignment, observed in English production data [15, 16] as well as in perception research with Dutch listeners [17]. Switching round sonorant and obstruent consonants in onset and coda is enough to cause a shift in perception from downstepped (early fall) to non-downstepped (late fall) pitch accents. Conversely, perception of the syllable boundary in English *Ella Norman/Ellen Orman* is influenced by the timing of the F_0 valley between the accents [18].

Taking a larger perspective, the timing of peaks may be sensitive to the number of unaccented syllables that remain till the phrase end or the next accent, suggesting speakers allow the target to be delayed relative to the segmental production when the opportunity arises [19, 20]. Moreover, the presence of another pitch accent within the same intonational phrase may precipitate the occurrence of an accent peak [20]. Phrase-final syllables need not invite speakers to delay the target, who would rather execute the full movement within the syllable, while phrase-internal ones do in Hamburg German [21]. Target alignment is one aspect of ergonomic behaviour in pitch production. A functional account of F_0 contours that charts the multiple effects of downstep, H-raising, utterance length and following tone on the maintenance of a three-way surface tone contrast in Yoruba is [22].

3.4 Communicative effects

Speakers employ pitch variation to signal universal meanings which are metaphorical interpretations of factors conditioning F_0 variation [23]. Just as John Ohala's Frequency Code associates high pitch with such 'small' meanings as vulnerability and submissiveness and low pitch with protectiveness and authority,

ISBN 1-876346-48-5 © 2003 UAB

the Production Code associates high initial pitch with topic beginning and low initial pitch with continuation, as well as low final pitch with the speaker's end of turn and high final pitch with continuation, while the Effort Code associates wide excursion with expenditure of effort, and hence emphasis and agitation, and narrow excursion with withdrawal and lack of interest. Timing of pitch movements comes into this picture because high pitch can be enhanced or replaced by late F_0 peaks. As a result, later initial peaks can signal newness of topic (e.g. [24], later peaks can signal interrogative meaning (e.g. [25]), while equally later peaks can signal greater emphasis (e.g. [26]). It may, again, be hard to assign such communicative uses of F_0 timing either to speaker control during the phonetic implementation or to representational differences. Peak height and timing differences between final peaks signalling (narrow focus) statements and questions in Bengali are unambiguously representational, and so is the peak timing difference between wide and narrow focus in European Portuguese, while peak timing differences between statements and questions for Dutch is very almost certainly phonetic. But if form-function relations go against the predictions of the biological metaphors, they must be grammatical. The timing of the high peaks in West Swedish that are found on the word *after* the focus constituent, which form would appear to contradict the Effort Code (Bruce, this symposium), must therefore be specified in the phonology.

REFERENCES

[1] John A. Goldsmith, *Autosegmental Phonology*, Ph.D. thesis, MIT, 1976, Distributed by IULC, 1976, Published by Garland Press, New York, 1979.

[2] Elsa Gomez-Imbert, "More on the tone versus pitch accent typology: Evidence from Barasana and other Eastern Tukanoan languages," in *Cross-Linguistic Studies of Tonal Phenomena*, Shigeki Kaji, Ed., pp. 369–412. Tokyo University of Foreign Studies, Tokyo, 2001.

[3] Janet B. Pierrehumbert and Mary E. Beckman, *Japanese Tone Structure*, MIT Press, Cambridge MA, 1988.

[4] John M. Stewart, "The typology of the Twi tone system," *Bulletin of the Institute of African Studies*, vol. 1, 1966, Preprint.

[5] George N. Clements and Kevin C. Ford, "On the phonological status of downstep in Kikuyu," in *Phonology in the 80s*, D.L Goyvaerts, Ed., pp. 309–357. Story Scientia, Gent, 1980.

[6] Janet B. Pierrehumbert, *The Phonetics and Phonology of English Intonation*, Ph.D. thesis, MIT, 1980, Published by Garland Press, New York, 1990.

[7] Mary E. Beckman and Janet B. Pierrehumbert, "Intonational structure in English and Japanese," *Phonology Yearbook*, vol. 3, pp. 255–309, 1986.

[8] Gösta Bruce, "How floating is focal accent," in *Nordic Prosody IV*, K. Gregersen and Hans Basbøll, Eds., pp. 41–49. Odense University Press, Odense, 1987.

[9] John J. McCarthy and Alan Prince, "Generalized alignment," in *Yearbook of Morphology*, Geert Booij and Jaap van Marle, Eds., pp. 79–153. Kluwer, Dordrecht, 1993.

[10] Carlos Gussenhoven, "The lexical tone contrast of Roermond Dutch in Optimality Theory," in *Prosody: Theory and Experiment*, Merle Horne, Ed., pp. 129–167. Kluwer, Dordrecht, 2000.

[11] Martine Grice, D. Robert Ladd, and Amalia Arvaniti, "On the place of phrase accents in intonational phonology," *Phonology*, vol. 17, pp. 143–185, 2000.

[12] Yi Xu and Q. Emily Wang, "Pitch targets and their realization: Evidence from Mandarin Chinese," *Speech Communication*, vol. 33, pp. 1399–1413, 2001.

[13] Janet B. Pierrehumbert, "Phonological and phonetic representations," *Journal of Phonetics*, vol. 18, pp. 375–394, 1990.

[14] Gösta Bruce, "Intonational prominence in varieties of Swedish revisited," in *Prosodic Typology and Transcription: A Unified Approach*, Sun-Ah Jun, Ed., Oxford, forthcoming, Oxford University Press.

[15] Jan van Santen and Julia Hirschberg, "Segmental effects on timing and height of pitch contours," *ICSLP3*, pp. 719–722, 1994.

[16] Pilar Prieto, Jan van Santen, and Julia Hirschberg, "Tonal alignment patterns in Spanish," *Journal of Phonetics*, vol. 23, pp. 429–451, 1995.

[17] Toni Rietveld and Carlos Gussenhoven, "Aligning pitch targets in speech synthesis: Effects of syllable structure," *Journal of Phonetics*, vol. 23, pp. 375–385, 1995.

[18] D. Robert Ladd and Astrid Schepman, "'Sagging transitions' between high pitch accents in English: Experimental evidence," *Journal of Phonetics*, forthcoming.

[19] Janet B. Pierrehumbert and S. Steele, "Categories of tonal alignment in English," *Phonetica*, vol. 46, pp. 181–196, 1989.

[20] Kim E.A. Silverman and Janet B. Pierrehumbert, "The timing of prenuclear high accents in English," in *Laboratory Phonology I*, John Kingston and Mary E. Beckman, Eds., pp. 72–106. Cambridge University Press, Cambridge, 1990.

[21] Jörg Peters, "Intonation und Fokus im Hamburgischen," *Linguistische Berichte*, vol. 189, pp. 27–57, 2002.

[22] Yetunde Laniran and Clements G.N., "Downstep and high raising: Interacting factors in Yoruba tone production," *Journal of Phonetics*, vol. 31, forthcoming.

[23] Carlos Gussenhoven, "Intonation and interpretation: Phonetics and phonology," in *Speech Prosody 2002. An International Conference*, Bernard Bel and Isabelle Marlien, Eds., Aix-en-Provence, 2002, pp. 47–57, CNRS and Université de Provence.

[24] Anne Wichmann, *Intonation in Text and Discourse. Beginnings, Middles and Ends*, Longman, Harlow, 2000.

[25] Mariapaola D'Imperio and David House, "Perception of questions and statements in Neapolitan Italian," in *Proceedings of EUROSPEECH '97*, 1997, vol. 1, pp. 251–254.

[26] D Robert Ladd and Rachel Morton, "The perception of intonational emphasis: Continuous or categorical?," *Journal of Phonetics*, vol. 25, pp. 313–342, 1997.