Explanations for Lg. Universals
NY: Mouton 1984
eds. Butterworth, Comrie, Dahl
237–247

Some physiological and perceptual constraints on tonal systems[1]

R. COLLIER

Introduction

Looking at the results of phonological or phonetic analyses of tonal phenomena in speech, one gets the impression that natural languages are strikingly similar as to the *kind of pitch attributes* they exploit in order to establish their tonal contrasts; they are also very similar as to the *number of categories* they discriminate along each attributive dimension. These similarities might be explained to some extent by taking into account certain limitations that are inherent in the speech production and perception systems.

In this paper I explore some of the physiological and perceptual constraints that set limits to the communicative use of pitch distinctions in speech. Such constraints define the set of possible 'phonetic features' from which natural languages can select the building blocks for their tonal systems. They also determine how many values each of these features can take.

The primary purpose of this undertaking is to explore what kinds of external restrictions are imposed on possible contrasts between *single pitch movements*. Given this limitation, many fundamental issues regarding the explanation of complete tonal systems will not be dealt with. I hope to show, however, that the approach is sufficiently promising to encourage further research along these lines. In the main body of the paper I will discuss the following topics: (1) the declination line of pitch contours, (2) the variable speed of pitch movements, (3) the variable size (or interval) of pitch movements, and (4) the variable position of pitch movements within the boundaries of the syllable. In each case I will present some observations (mainly related to intonation languages and occasionally to tone languages), offer a physiological or perceptual explanation, and consider some implications for a theory of pitch-movement universals.

The slope of the declination line

In intonation languages, declarative utterances have a lower pitch at the end than at the beginning. In visual displays of the fundamental frequency (F_0) variations of such utterances, it can be seen that the overall contour is somewhat tilted, superimposed as it were on a baseline that gradually slopes downward. For instance, if a contour consists of successive rise–fall combinations of roughly equal magnitude, the peaks and valleys will nevertheless have increasingly lower values (Figure 1).

The lines that can be fitted through the peaks and valleys in Figure 1 are called the upper and lower 'declination' lines (Cohen and 't Hart 1967). The declination phenomenon has been attested in languages such as English, French, Dutch, Swedish, Italian, and Japanese.

A comparable phenomenon can also be observed in tone languages (Meyers 1974). In these languages it is common that after a low tone (L), a high tone (H) does not reach its usual pitch but stops short of it: it is assimilated downward. L following H, however, is generally not assimilated upward. As a result of this asymmetry, a sequence of HLs shows the so-called 'downstep' effect illustrated in Figure 2A. Moreover, in sequences of like tones — where tonal assimilation cannot apply — each tone is produced on a successively lower pitch. This phenomenon is called 'downdrift'; it is illustrated in Figure 2B. Downdrift and downstep have been attested in tone languages with so-called 'level' or 'register' tones, such as Yoruba, Hausa, and Shona.

It thus appears to be a widespread property of (declarative) pitch contours to exhibit a gradual declination tendency. In a sense this tilted baseline seems more natural than a strictly monotonous one. It may reflect an equally natural property of the mechanism that controls the rate of vocal-cord vibration. Our present understanding of this mechanism is far from perfect, especially as far as pitch lowering is concerned. It is not unlikely, however, that declination is primarily caused by the gradual decrease of subglottal air pressure in the course of an utterance. There is still debate over whether the diminishing amount of air in the lungs affects the rate of vocal fold vibration directly — by reducing subglottal pressure (Collier 1975) — or indirectly — through a tracheal pull (Maeda 1976). Furthermore, it is not excluded that the laryngeal musculature, too, contributes to the declination rate, especially if the latter is large (Ohala 1978).

Being a natural consequence of certain properties of the pitch-regulation mechanism, declination explains part of the discrepancy between the phonological and the acoustic-phonetic characterization of tonal features. More importantly, declination appears to be a psychological reality: the

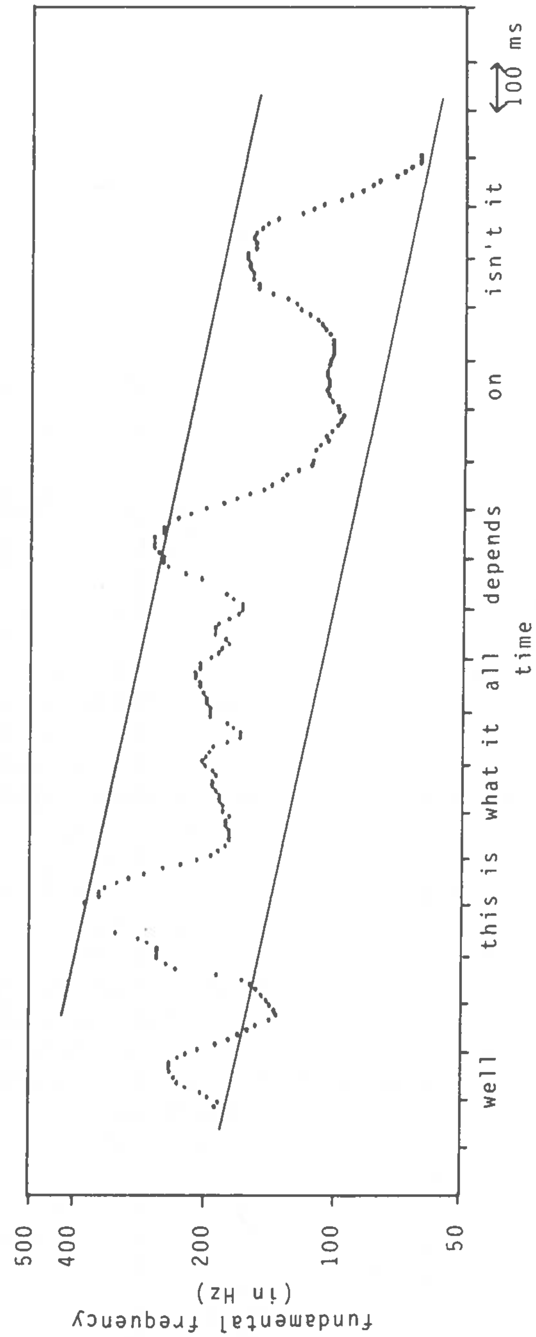

Figure 1. *Fundamental frequency curve (dotted line), with upper and lower declination lines*

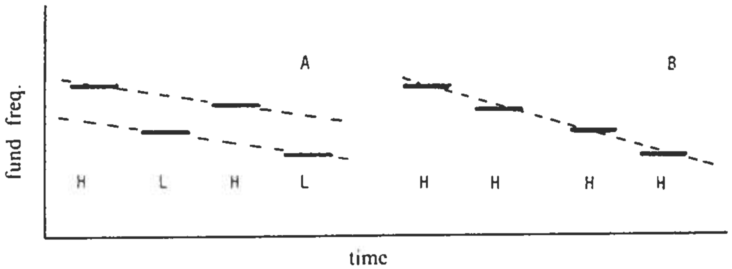

Figure 2. *A. Stylized F_0 curve of a sequence of high (H) and low (L) tones, illustrating 'downstep'*
B. Stylized F_0 curve of successive high (H) tones, illustrating 'downdrift'

speaker is aware of this physiological constraint and the listener actively uses knowledge about it in the perceptual evaluation of pitch phenomena. The following observations support this claim.

First, there is the observation that electronically generated utterances sound very unnatural if declination is omitted in synthesizing their F_0 contours. More direct experimental evidence for the greater naturalness of contours that incorporate declination comes from a study by Abramson (1978). In synthetic approximations of the tones of Thai, he found that the so-called 'midlevel' tone was far better identified if the stimulus was not strictly monotonous (120 Hz), but had a slightly declining pitch (120 to 116 Hz in 350 ms): the identification percentage rose from 71 to 84.

Second, it appears that a listener computes the slope of the declination line and uses it as a reference point when evaluating the endpoint of a rise or fall. Pierrehumbert (1979) reports an experiment in which listeners had to evaluate whether the peaks of two successive rise-falls (in the same contour) were equal in pitch. The second peak was estimated to be equal to the first when, in fact, it was 10 to 15 Hz lower. Apparently, the listener subjectively adds to the objective F_0 values the amount that he thinks has been subtracted by declination. To put it differently: successive peaks are judged to be of equal height if they reach the perceptually computed level of the upper declination line.

The same strategy appears to apply to the situation in which the listener has to evaluate which type of pitch fall is encoded in a decrease of F_0. Indeed, a fall is judged to convey an impression of finality only if it restores the pitch to the level of the lower declination line. In that case it will be considered a full-size pitch movement, even if its interval is objectively smaller than that of a larger fall that does not reach the

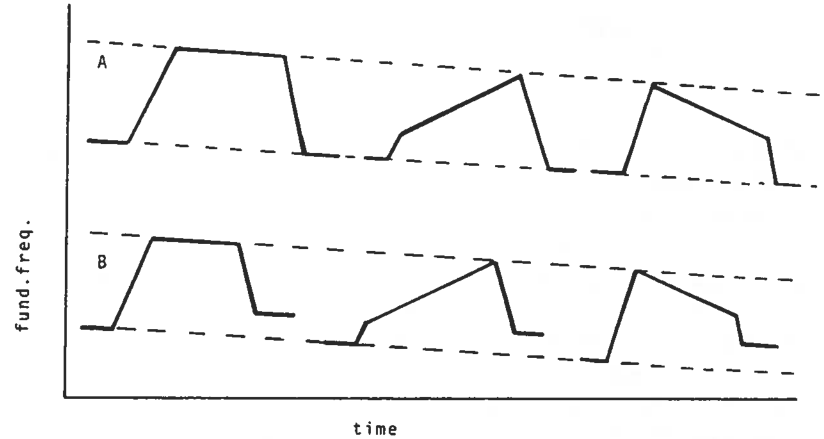

Figure 3. *Difference between final falls (A) that reach the lower declination line and nonfinal falls (B) that do not reach the baseline*

baseline. The latter will be interpreted as half-size and nonfinal (Figure 3).

Third, the perceptual reality of declination appears to be such that longer contour stretches with objectively steady F_0 are not perceived as monotonous, but rather as having gradually increasing pitch ('inclination').

A fourth and final observation concerning the psychological relevance of declination is the fact that marked deviations from its expected course appear to have communicative significance. One kind of deviation is 'resetting', i.e. the start of a new declination ramp. Resetting may serve as a cue that signals the beginning of a new syntactic constituent. If that constituent is a new sentence, resetting may be a rather trivial phenomenon, especially if it results from inhalation. If, however, resetting occurs at phrase boundaries and is caused by some voluntary respiratory or laryngeal maneuver, it may be a helpful cue, among other prosodic markers, for the listener to recover the syntactic organization of the utterance. There is some evidence that declination is reset at certain types of clause boundaries (Cooper and Sorensen 1981). Experiments by Thorsen (1980) indicate that in nonfinal and interrogative utterances the declination ramp is significantly less steep than in declarative ones, and listeners appear to be sensitive to this difference when assessing the degree of finality of an utterance.

In short, declination (or downdrift) is a common property of the speech melody in natural languages. It has a natural physiological cause and it constitutes a psychological reality to both the speaker and the listener. As such it explains part of the discrepancy between phonological and phonetic representations of tonal features.

The variable speed of pitch movements

If one attempts to produce melodic variations in synthetic speech by making abrupt jumps between two fixed pitch levels, as Isačenko and Schädlich (1970) did for German intonation, the result sounds highly unnatural. Even if the phonological analysis of tonal phenomena postulates the existence of 'levels', the phonetic implementation of successive unlike levels always implies some gradual transition in the form of a pitch movement rather than a pitch jump. Thus, for instance, in Yoruba a HL sequence is realized as 'high + falling' and a LH sequence as 'low + rising' (Elimelech 1974). The ubiquity of pitch movements in the melodic variations of languages suggests that the speed with which pitch changes can be effected is determined by limitations in the mechanism that is responsible for their implementation.

From a study by Sundberg (1979) we learn that it is physiologically impossible to produce pitch jumps: pitch changes take a minimum of time and any transition between unlike pitch levels of necessity results in some form of glide. For instance, when requested to bridge an interval of four semitones as quickly as possible, an untrained male subject needs some 85 ms if the direction is rising and some 75 ms if it is falling. To cover a preset interval of one octave takes some 100 ms with rises and some 75 ms with falls. In other words, the maximum speed varies with the interval in the case of rises and is fairly constant with falls, but under all circumstances pitch changes require time.

The asymmetry in the maximum speed of rises as compared to falls suggests that the latter may in some sense be easier to produce. According to Ohala (1978), this relative ease of production may explain why falls are more common than rises in the tonal inventories of languages and why they are less vulnerable to tonal assimilation than rises are. One may add that, conversely, rises are more conspicuous pitch movements because they are a marked departure from the natural tendency for a pitch contour to fall off gradually.

Individual languages may differ as to the speed of their pitch changes. In Dutch, for instance, the average interval covered by a pitch movement is some 4.5 semitones. Its average duration is 100 ms. In British English, on the other hand, pitch movements spanning one octave are very common. They then take an average of 160 ms to be completed (De Pijper 1980). These standard values are the ones used at the Institute for

Perception Research (Eindhoven, The Netherlands) in synthesizing pitch contours in these two languages. They reflect tendencies measured in F_0 curves. Admittedly, a direct comparison with the Sundberg data on singing is not warranted, but this nevertheless suggests that pitch changes in speech are effected more slowly than the maximum speed allowed by the laryngeal control mechanism. If required, they could be made faster or, of course, slower than usual. This raises the question of whether within a single language tonal contrast could be based on a difference in speed among otherwise similar pitch movements. In other words, could there be several categorically different rises (or falls), each having a particular slope?

An answer to this question can be provided by examining the perceptual limitations in the discriminability of different slopes. 't Hart (1976) has combined various psychoacoustic findings regarding this issue into one predictive graph. It appears that the sensitivity to variations in the rate of change of pitch rises is a function of both their duration and their size. More than one slope can be distinguished for the longer durations only: when the interval ranges from 6.5 to 9.5 semitones, two slopes can be distinguished if the duration exceeds 400 ms; when the interval is between 3.5 and 6.5 semitones, two slopes can be discriminated if the duration is at least 250 ms; when the interval is between 0.5 and 3.5 semitones and the duration between 150 and 300 ms, two slopes can be distinguished, and if the duration exceeds 300 ms, three slopes may be discriminated. However, since the duration of medium-sized pitch movements is usually shorter than 250 ms, it follows that differences in the rate of change of pitch rises will remain below threshold and will therefore not be a useful basis for tonal contrast. Only gradual rises extending over several syllables (i.e. inclination) may audibly differ in slope.

In short, the observed speed of pitch changes is slower than the maximally attainable speed. This suggests that speakers do not really program their pitch changes as sudden jumps between levels (jumps which would then happen to result in glides because of the inertia of the laryngeal structures). Listeners, on the other hand, are rather insensitive to differences in speed between pitch movements of standard duration. Only with pitch changes that extend over several syllables does a difference in speed become noticeable. This observation suggests that only a slope difference in declination or inclination can be perceptually relevant.

The variable size of pitch movements

In F_0 registrations (see Figure 1) one can easily observe that the various pitch movements are of unequal size: they cover variable intervals.

Perceptually, however, much of this variability goes unnoticed. One might therefore ask how many intonational distinctions can be established in a language on the basis of different pitch distance covered by rises and falls. Analogously, one might ask how many registers can contrast in a tone language.

't Hart (1981) has found that the differential sensitivity to pitch distance is rather poor. Under circumstances that are typical of running speech, successive pitch movements in the same direction have to differ by at least 3 to 4 semitones in order to be perceived as different in size by the average listener. This insensitivity to interval differences compares unfavorably with the outcome of psychoacoustic experiments in which subjects usually notice a pitch difference of about 55% between stationary sine waves around 1000 Hz, or of about 5% between stationary complex waves in the vicinity of 150 to 200 Hz. As 't Hart points out, the actual speech situation is far more complicated. Listeners have to evaluate the difference covered by successive dynamic rather than stationary tones. Due to the presence of declination, the onset frequencies of the F_0 changes are unequal and, if they span the same interval, their offset frequencies will be different, too.

From 't Hart's findings it follows that in running speech a pitch range of one octave can be quantified into no more than three or four distinguishable intervals: small, medium, large, and possibly extra-large. It may also follow that in tone languages the number of contrasting levels per octave cannot exceed four. Indeed, very few tone languages exploit this maximum number of contrasts among register tones. One example is Feʔ Feʔ, with a low, a raised low, a mid, and a high tone (Hyman 1973). Many more tone languages have tonal systems in which two or three 'level' tones are supplemented with one or more dynamic 'contour' tones, as in Thai, which has a low, a mid, and a high tone in combination with a rising and a falling tone (Abramson 1978).

The timing of pitch movements

In some languages, such as Swedish, Serbo-Croatian, and Japanese, lexical items may contrast in what is called their 'accent'. In Swedish, for example, some words are pronounced with either accent I or accent II, and their meaning varies accordingly. The major phonetic difference that correlates with this accentual contrast resides in the timing of otherwise identical pitch changes (Malmberg 1962; Bruce 1977). A timing difference of not more than 25 ms may be sufficient to cross the perceptual boundary that separates the two accent types.

Even in languages without such an accentual system there are pitch movements that are categorically different in their location with respect to the syllable boundaries. In Dutch, for example, a rise can be located early, late, or very late in the syllable (Collier 1975). The same holds for a fall ('t Hart and Collier 1971). A comparable timing difference can be found in the pitch movements of British English (De Pijper 1980) and American English (Pierrehumbert 1981).

A three-way distinction of rises and falls on the basis of their timing is likely to be a maximum. According to Hill and Reid (1977) the just-noticeable difference for the position of rises is less than 50 ms in the region of the transition of the first formant and 70 ms in the syllable nucleus. Assuming that the average duration of a (stressed or prepausal) syllable is some 200 ms, it follows that no more than three perceptually different positions of pitch rise can be discriminated within its boundaries. For falls the discriminability is not likely to be any better.

Discussion

The physiological and perceptual constraints reviewed in the preceding paragraphs allow one to predict that intonation languages will select their contrastive pitch movements from the inventory that is defined by the attributes and categories listed in Table 1. As can be seen, the number of perceptually distinct pitch movements appears to be 36. They can be superimposed on a baseline that shows either strong, weak, or no declination.

I hope to have shown that the categorization of pitch movement attributes can in principle be explained on external grounds, i.e. by reference to properties of the speech production and perception systems. The present effort is merely illustrative of an approach that may be fruitfully extended in order to solve some of the numerous remaining problems. For instance, on the level of the single pitch movements, it remains to be examined how some of the attributes interact. It may be the case that the timing distinction is relevant for 'fast' pitch movements only (i.e. for those whose duration does not exceed that of a syllable). In such event the number of contrastive pitch movements would be reduced to 24. Also, the distinctions presented here are claimed to hold irrespective of the slope of the declination line, but this still needs to be checked. It may be the case that some distinctions become perceptually irrelevant when declination is weak or absent. Thus, for instance, the contrast between small and medium-sized rises may disappear when both are slow and declination is weak.

I have assumed that the same distinctions are applicable to both rises and falls, but further study of the interaction between attributes may destroy this strict parallelism.

Table 1. *Attributes and categories defining contrastive pitch movements and baseline*

Attributes	Categories
Direction	rise fall
Speed	fast slow
Size	small medium large
Timing	early late very late
Baseline	strong declination weak declination no declination

It is also worth investigating how languages differ in the number of basic pitch movements they use, and examining whether there is a preferred order in which certain types of pitch movements appear when that number increases. Presumably, universal principles regarding ease of production and enhancement of tonal contrast can account for such selection preferences.

In extending the investigation to include the interaction of attributes and the selection of types of contrast, it should be borne in mind that pitch movements do not exist independently of the larger structures of the *pitch contours*, which, in turn, are manifestations of the abstract melodic entities called *intonation patterns*. Which combinations of pitch movements constitute optimum patterns and which patterns form optimum intonation systems cannot be inferred from the survey presented here. Presumably, some of the explanatory principles needed will still be physiological or perceptual in nature, but most of them may reflect constraints of a more cognitive kind.

University of Antwerp
UFSIA
Belgium
Institute for Perception Research
Eindhoven
The Netherlands

Note

1. I would like to thank A. Cohen, J. 't Hart, S. Marcus, and S. Nooteboom for their helpful comments on an earlier draft of this paper.

References

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech* 21, 319–325.
Bruce, G. (1977). Swedish word accents in sentence perspective *Trav. Inst. Ling. Lund* 12, 1–155.
Cohen, A., and 't Hart, J. (1967). On the anatomy of intonation. *Lingua* 19, 177–192.
Collier, R. (1975). Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America* 58, 249–255.
Cooper, W. E., and Sorensen, J. M. (1981). *Fundamental Frequency in Sentence Production.* New York: Springer.
de Pijper, J. R. (1980). A melodical model of British-English intonation. *I.P.O. Annual Progress Report* 15, 54–58.
Elimelech, B. (1974). On the reality of underlying contour tones. *UCLA Working Papers in Phonetics* 27, 74–83.
Hill, R. D., and Reid, N. A. (1977). An experiment on the perception of intonational features. *International Journal of Man–Machine Studies* 9, 337–347.
Hyman, L. (1973). The role of consonant types in natural tone assimilations. In L. Hyman (ed.), *Consonant Types and Tone*, Southern California Occasional Papers in Linguistics, No 1.
Isačenko, A., and Schädlich, H. J. (1970). *A Model of Standard German Intonation.* The Hague: Mouton.
Maeda, S. (1976). A characterization of American English intonation. Unpublished doctoral dissertation, MIT, Cambridge, Mass.
Malmberg, B. (1962). Analyse instrumentale et structurale des faits d'accents. *Proceedings 4th International Congress of Phonetic Sciences*, Helsinki, 1961. The Hague: Mouton.
Meyers, L. (1974). Tone patterns in Hausa: a re-analysis of Hausa downdrift. *UCLA Working Papers in Phonetics* 27, 47–62.
Ohala, J. (1978). Production of tone. In V. Fromkin (ed.), *Tone: A Linguistic Survey.* New York: Academic Press.
Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* 66, 363–369.
—(1981). Synthesizing intonation. *Journal of the Acoustical Society of America* 70, 985–995.
Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7, 71–79.
't Hart, J. (1976). Psychoacoustic backgrounds of pitch contour stylisation. *I. P. O. Annual Progress Report* 11, 11–19.
—(1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America* 69, 811–821.
—, and Collier, R. (1971). A speaker-grammar for Dutch intonation, I.P.O. Report 227 (unpublished).
Thorsen, N. (1980). A study of the perception of sentence intonation — evidence from Danish. *Journal of the Acoustical Society of America* 67, 1014–1030.