

A PARALLEL BETWEEN MUSIC AND SPEECH: TONALITY AND TONE

John T. Hogan
University of Alberta

ABSTRACT

The goal of this paper is to discuss a substantive rather than a formal parallel between language and music and to suggest that musical notation may have value in representing some speech material. The domain of interest is limited to speech perception and phonetics only. The phenomenon of interest is tone, especially the tonal sequences that occur in some African languages which have downstep.

1. INTRODUCTION

Parallelism and analogies between language and music have been pursued by Winograd (1968) and, more notably, by Lerdahl & Jackendoff (1977, 1983). Both have used a formal approach for the perception of musical structure with Winograd's major influence from artificial intelligence and Lerdahl & Jackendoff's ideas from generative grammar in linguistics.

The most extensive work to date has been accomplished by the latter authors. Lerdahl & Jackendoff's model is based on the premise that music has a hierarchical organization similar to that of phrase structure which can be produced by rules of well-formedness and preference. For Lerdahl & Jackendoff, the structural analysis of music is divided into four types: two kinds deal with rhythmic structure, namely, grouping and metrical analysis, and two with pitch organization, namely, time-span analysis and prolongation reduction.

Grouping rules are based on repetition, symmetry of sequences and the manner in which their boundaries are determined. Metrical analysis deals with the pattern of strong and weak beats. The formal devices to indicate beats and groups are carried out by a pattern of subscripted dots for beats and by nested bracketing for groups.

Time-span analysis follows from the grouping and metric analysis. Within groups a dominant event (such as a tonic note) is located and elaboration locally around that event is specified. A hierarchy of stability around a tonic is based on factors of consonance, triads in root position as opposed to inverted positions, circle of fifths from the tonic, etc. In terms of tree structure, the dominant event is associated with a major node and branching from that node would represent elaborations. Finally, prolon-

gation analysis indicates the elaboration through repetition and contrast by introduction of different pitch events for the development of tension and relaxation in a piece of music. This form of analysis also used rules of formation similar to phrase structure rules.

Lerdahl & Jackendoff's model was an attempt to elucidate the musical structure of western tonal music as perceived by the listener. Jackendoff (1992) further suggests that, like language perception, music perception involves unconscious construction of abstract structures from the musical surface and that the musical 'processor' is a module in Fodor's sense (1983). Their works largely exploited the formal system of linguistics to analyze for types of structural relationships in music via rules of well-formedness which specify all possible sequences and rules of preference which indicate the perceptual tendencies of a listener.

From a psychological perspective, Handel (1989) notes that processing of tonal sequences and melody is facilitated when the tonal context creates a perceptual structure. The parallel to this in language is that comprehension and understanding are faster when the stimuli are presented in a context. The structure among the elements is perceived more readily than the actual physical sequence. More in the same line, Handel (1989) states that scales and chords bring about a sense of context as lexical items bring about a priming effect in a discourse. Also, the garden-path effect which allows context to override ambiguous interpretation is paralleled in music where a chord may be a member of several keys but resolved in the context of all the other elements in the piece.

2. PITCH AND TONE

Pitch as measured by the fundamental frequency of a periodic speech signal conveys a variety of information in languages. It is usually associated with prosodic (suprasegmental) elements such as intonation, cues for stress, and tone. Pitch variation also indicates the emotional and attitudinal state of a speaker and is indicative of individual speaker characteristics. That is to say, different pitch ranges indicate the gender and age of a speaker.

3. LINGUISTIC FUNCTION OF PITCH

The linguistic function of pitch ranges from the syntactic to the lexical. For example, syntactic information signaled by rising or falling pitch indicates the declarative-interrogative distinction. Pitch accent systems have

a distinctive function in differentiating bisyllabic words as in Swedish. For monosyllabic words, pitch differences may be used for lexical differences and are designed as syllable-based tonal systems. Languages employing tone are typologically divided into two major types: (1) contour tone languages, and (2) register tone languages. The relevant feature of contour tone languages is the change in pitch, i.e., the trajectory or shape of the pitch change. The characteristic of a level tone system is the height of the tone. Tone changes in this system are a result of a mixture of level tones. The most common number of tones is usually two, namely high and low tones (Maddieson 1978). Less frequent are three tone systems with high, mid and low tones. Languages with four and five tone levels are much rarer.

In many African languages of the Niger Kordafanian group, some two and three tone languages have a lowering of the high tone: this phenomenon, which differs from declination and downdrift, has been designated as downstep, which differs from the phenomena of declination and downdrift. Declination refers to a slight lowering in pitch found in most languages. Downdrift is a marked lowering of a high tone only after it follows a low tone. No succeeding high tone can rise to a higher level from the downdrifted high tone. The downdrift effect can result in high tones at the end of an utterance being lower in pitch than a low tone near the beginning of an utterance. For this to occur there has to be some lowering of the low tone throughout the utterance.

Downstep is a marked lowering of a high tone similar to that of downdrift without the presence of a preceding low tone. Downstep in these Niger Kordafanian languages has a phonemic function. The explanation for the downstepped high tone is based on either the historical fact that the vowel bearing a preceding low tone was lost due to apocope but the lowering effect of the low tone has survived or that the low tone becomes a high tone due to tone spreading from a preceding high tone. This is a perseveratory assimilation process. However, the effect of downdrift is as if the assimilated low tone were present and were still operative.

The above description of downstep is very traditional to African linguistics (Welmers 1973). However, Snider & van der Hulst (1993) report linguistic attempts to represent multiple tone heights with the use of a high and low tone feature alone plus downstep as elements in various formal phonological theories such as autosegmental phonology, metrical theory, feature geometry, etc. Van den Berg et al. (1992) have attempted to carry out a formal analysis of Dutch intonation with the categories of High tone,

Low tone and downstep along with reset, the interruption of downstepping which shifts the tonal register upwards.

4. KONO: A TWO TONE SYSTEM WITH DOWNSTEP

4.1. Measurement

The data used for explication of language and music parallels will come from Kono, a Northern Mande language spoken in Sierra Leone. Utterances made by M.M. (a male speaker) were recorded on a TEAC A-7030 tape recorder and the fundamental frequency values were measured both from a Frøkjaer-Jensen transpitch meter and on the Kay Digital Sona-Graph 7800.

For the spectrographic analysis, a 45-Hz filter was used to produce narrow-band sonagrams. With an overlay of a spectrum of a 100 Hz sawtooth wave calibration tone which is the finest gradation possible for this machine (2mm apart), the highest resolution in measurement is in 5 Hz steps. Fundamental frequency values were taken from higher harmonics and then divided by the harmonic number since changes in the f_0 region are hardly visually perceptible.

The duration of the pertinent syllables varied from 200 to 250 ms in duration. The f_0 measurement was made approximately 125 ms from the syllable onset in order to obtain values from a steady state portion and avoid any perturbations of f_0 . Apart from the syllable onset, f_0 values were fairly level throughout the syllable nucleus duration.

For the examples IPA transcription is used with ´ indicating high tone, with ` indicating low tone and ↓ after the relevant syllable indicating downstep.

4.2. Results

An example with subscripted fundamental frequency values of a minimal distinction by downstep is as follows:

1. è táwá cé tíná
 100 135 135 130 130 90
 He/she/it is in the cooking place
2. è táwá cé↓ tíná
 105 135 135 120 120 90
 He/she/it is cooking

Longer examples are as follows:

3. ná fétú sàŋ nǐ í má tó↓ ní dùmbí' à
 140 140 90 90 120 120 110 110 80 100 80
 'my pineapple buy if you neg. like neg. orange the'
 Buy my pineapple if you do not like the orange.

The syllables with the single underline indicate an example of downdrift after a low tone and the double underlined syllable is a case of downstep.

An example of two downsteps is as follows:

4. ní dǒ kèmá má tó↓ né ké ì má dón↓ ní
 130 130 100 120 120 110 110 110 100 115 100 80
 I say Kema neg. stay have but you neg. agree neg.
 'I say that Kema should not stay here but you do not agree.'

In all of the above examples there is a final lowering of the fundamental frequency. This most likely indicates a sentence final boundary. Also, in example 4, the second downstepped syllable dón↓ has a fundamental frequency as low as the earliest low tone in the sequence.

5. CONVERSION OF FUNDAMENTAL FREQUENCIES TO MUSICAL NOTATION

When one listens to the above utterances, one is struck by the fact that there is a pattern of pitches that sounds like a melody in music. When the sentences are imitated, the easiest way to remember the pattern is to think of it as a melodic sequence. With this in view, what kind of notes from a scale does the melody consist of? In many of the utterances the lowest frequency on average was 90 Hz. Selection of this value as a tonic note of a musical scale, scale based well-tempered tuning, for example, can be constructed.

Well-tempered tuning was an invention during the Baroque period in order to circumvent two earlier forms of tuning, namely the Just scale tuning and Pythagorean tuning. In Just tuning the seven notes of the diatonic scale have fundamental frequencies that are in small whole number ratios to each other. Roederer (1974) claims that this tuning will minimize the amount of harshness that will result when two upper harmonics fall within the same critical band in hearing. (Frequencies that fall outside a critical band will be processed independently of each other whereas two frequencies that do not have the same value will produce beats or harshness if they fall within a critical band of hearing). Both scales are designed to maximize consonance of notes and minimize dissonance. Although these tunings are natural tunings, they have one great disadvantage in that a keyboard instrument tuned thusly can only be played in one key. The equal tempered tuning circumvents this difficulty by having each of the twelve

half tones in an octave on a keyboard instrument in a ratio of $12\sqrt{2} = 1.059$ of its successor. There is apparently a minor cost of a slight increase of dissonance when any two notes are played on a well-tempered instrument. The frequency values in a well-tempered scale with their diatonic scale notation for the speaker of Kono are as follows:

Frequency Values in a Well-tempered Scale
[for M.M., a speaker of Kono]

90.00 Hz	=	F
95.31 Hz	=	F [#] = G ^b
100.98 Hz	=	G
107.01 Hz	=	G [#] = A ^b
113.40 Hz	=	A
120.15 Hz	=	A [#] = B ^b
127.27 Hz	=	B
134.82 Hz	=	C
142.83 Hz	=	C [#] = D ^b

For the shorter sentences the total pitch range is roughly a perfect fifth and a minor sixth interval for the longer sentences which have to accommodate to a greater number of downsteps and downdrift.

The shorter sentences usually will start with a C[#] for the high tone and G or G[#] for the low tone. In (1) a declination occurs on the fourth and fifth syllable. The difference between the downstepped cé in (2) and the cé in (1) is 10Hz which is equivalent to an interval slightly less than a whole tone from C to A[#]. For the longer sentence (3), the tonal range is expanded between 80 Hz and 140 Hz from 90 Hz to 135 Hz. In musical notation, the range is an interval from E to D, a minor seventh interval in musical terms and the interval for the short sentence ranges from F[#] to C[#], a perfect fifth interval. For the wider range in sentences (3) or (4), six out of eleven possible half-tones were utilized and five out of nine half-tones were used for the short sentences.

Apart from some declination effect in sentence (1) where the fundamental frequency is lowered by 5 Hz, about a quarter-tone in music, all major tone changes of a high tone or low tone consisted of musically a whole tone or more.

6. SOME PARALLELS WITH MUSIC

6.1. Discreteness of intervals

In register tone systems and musical scales there appears to be a tendency for a fixed number of intervals (Dowling & Harwood, 1986). In music, pitch glides such as glissandos, portamentos and other ornamentation occur but are not essential to the musical structure. This is not the case for all tone languages since the contour tone systems of south-east Asia consist of continuous tone change throughout a syllable.

6.2. The number of usable pitch relationships per octave

Most Western instruments contain 12 notes per octave which comprises the chromatic scale. However, much of Western music is written with one of the 12 major and minor diatonic scales which contains seven steps per octave. It has been reported that the musical systems of India and the Arab-Persian musical systems use fifteen to twenty-five intervals per octave. However, Burns & Ward (1982) indicate that the microtones (i.e., tonal intervals less than a half-tone) are not played as discrete intervals but as a variation of intonation around a note. It appears that the microtones used in these systems are variants on a smaller number of intervals (less than 12).

The number of notes per scale and the number of level tones fall within G. Miller's (1956) range of 7 plus or minus 2 elements along one psychological dimension when a person is carrying out a music or speech processing task along several different dimensions (phonetic to semantic for language, and melodic, rhythmic and harmonic for music).

Given Miller's range from five to nine elements, the number of tone levels, 5 or 6, fall within this range and this is compatible with the pentatonic (5 notes) and diatonic scales (7 notes), both not including the octave note. This limitation may be independent of language or music but it constrains both.

Notable exceptions to this limitation are persons with absolute pitch who can identify about 75 categories throughout the auditory range (Ward & Burns 1982). The average number is five.

6.3. Octave generalization

Tones or notes separated by a musical octave are regarded as the most similar pair of notes or as near equivalents in almost all musically developed cultures. Psychophysically, an octave would produce no dissonance. Tone sequences that are an octave apart are decoded as equivalent se-

quences. This obviously is the case when men and women or children are speaking a tone language where their ranges are roughly an octave apart. Based on averages taken from Peterson & Barney (1952), the average male voice was 132 Hz which is near C, the average female voice 223 Hz, nearly A above C and the average child voice 264 Hz which is roughly an octave above the male voice. In terms of singing, children's voices fall in the range of a soprano.

Burns & Ward (1982) report that intervals less than a fourth are perceived as perceptually wider. In production, this may lead to a compression of an interval relative to the equal temperament scale. Usually larger intervals than a fourth are heard as perceptually narrower than the equal tempered interval. In reporting those intervals subjects tend to stretch the scale. Thus, the measurements for downdrift or downstep may be slightly less than a whole tone distance but larger than a half-tone.

6.4. Categorical perception

Burns & Ward (1978) suggest that musical half-tones are categorized in a similar fashion to speech sounds, especially the acoustic phonetic categories for manner or place of articulation for stops.

Burns & Ward (1978) carried out an experiment where subjects had to identify the number of musical notes over a three half-tone range. Psychophysicists divide that whole octave into 1200 cents so that each half-tone is 100 cents. In the identification experiment, stimuli were presented in 10 cent steps. Three identification curves were obtained corresponding to 3 half-tone categories. Discrimination tasks were also carried out and, for musicians, discrimination was significantly better for two stimuli steps that straddled a category boundary than for two stimuli drawn from within a category. In the model of speech perception presented by Studdert-Kennedy et al. (1970), categorical perception holds. However, these results were drawn from musicians who may mediate some music percepts on the dominant or speech side of the brain. Results from the musically untrained showed high discrimination throughout the range of stimuli. Since the differences in frequency were just larger than the just noticeable differences in a psychophysical task, the musically untrained did not seem to be categorizing the stimuli as music. This type of perception is classified as continuous perception. Massaro (1994) claims that the above distinction between categorical and continuous perception is a result of a chosen model of perception and all perception may be on a gradient between the perception of discrete to continuous categories.

7. MUSICAL APPROACH

One major advantage of the use of musical scales and their notations is the invariance of pattern for different voice registers. This is similar to melodies played in different keys. When the average fundamental frequency for a speaker's lowest tone is determined, a chromatic scale can be constructed by multiplying the fundamental frequency value by $12\sqrt[12]{2}$ and then each successive product by the same value until the full 12 note chromatic scale is constructed. Thus, the tonal series for a speaker can be registered with a musical notation if desired. Information for more formal phonological analysis can be recoded from the musical notation.

The physical values obtained from the fundamental frequency measurements form a logarithmic scale. For example, an octave above a speaker's lowest fundamental is double that lowest value. The next octave is double the previous octave value and four times the value of the lowest fundamental frequency. Octave relationships are usually plotted on logarithmic scales for physical values in music. Does this scale have any value in perception of musical and other frequency relationships? The earliest psychophysical scale developed for pitch-frequency relationships was the mel scale based on the method of fractionation whereby subjects were asked to adjust a pure-tone frequency of 1000 Hz to a value where a subject thinks the frequency is twice as high. Similarly, the subject also adjusts the frequency of a 1000 Hz tone to a point where the adjusted frequency sounds half as high. This process is continued so that a mapping between physical frequencies and a psychological scale is established. For pure-tone frequencies the relationship is neither logarithmic nor linear throughout the whole range of hearing but roughly an ogive shape mapping. The relevant interval for the range of fundamental frequencies for the human voice is roughly linear when mapping from a physical logarithmic scale to the linear mel scale. Other psychophysical scales have been developed that capitalize on other aspects of hearing. The Bark scale is based on the critical bands of hearing which determines when frequencies of complex tones are resolved if they fall in two separate critical bands or produce beats or rough tones if they fall within any one of the twenty-four critical bands. The mapping between a logarithmic physical scale and a linear band scale is an arctangent function. The scale that is the most apt for this discussion is the semitone scale developed by Johan 't Hart (1990). This scale focuses on the perception of pitch in terms of distances between frequencies rather than on the absolute frequencies themselves. 't Hart defines the distances in semitone units as follows: $D = 12 \log_2 f_1/f_2 = (12/\log_{10} 2) \log_{10} f_1/f_2$. Thus,

for example, the f_0 difference of 150 to 100 Hz is 7.02 for a male voice and 270 to 180 Hz is 7.02 semitones for a female voice. This captures the equivalencies of downstep and downdrift among speakers with different registers.

Finally, should speakers whose voices are of a different range due to age, sex, or individual differences be regarded as using a pattern from a musical key? If this is assumed, then one is committed to positing a tonal center (tonic note) for each person. This may be difficult for spoken language since the fundamental frequency values may not be values associated with those of a standard tuned musical instrument. This also makes the design of a musical transcription system somewhat less defensible. It would be more judicious to posit a chromatic or whole tone scale which can begin or end anywhere (Hut & Anderson 1974).

8. CONCLUSION

In summary, this paper addresses the similarities between music and language in terms of tonality. The best cases for this comparison are register tone languages with downstep. A musical notation is suggested to capture the relationships of the tones. Some parallels between music and language were drawn in terms of discreteness of intervals, the finiteness of steps, and categorical perception. Finally, in terms of perception, a logarithmic scale is drawn from J. 't Hart to describe the distance relations among the tones. This scale is quite consonant with the chromatic musical scale.

REFERENCES

- BURNS, EDWARD M. & DIXON WARD. 1978. Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America* 63: 456-468.
1982. Intervals, scales and tuning. In Diana Deutsch (ed.), *The Psychology of Music*. New York, NY: Academic Press, 241-269.
- DOWLING, W. JAY & DANE L. HARWOOD. 1986. *Music Cognition*. New York, NY: Academic Press.
- FODOR, J.A. 1983. *The Modularity of Mind*. Cambridge, MA: M.I.T. Press.
- HANDEL, STEPHEN. 1989 *Listening*. Cambridge, MA: M.I.T. Press.

- 't HART, J., COLLIER, R., & COHEN, A. 1970. *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.
- HUT, LUCIEN & JODY R. ANDERSON. 1974. *Handbook of Musical Scales*. Missoula, Montana: Mountain Press Publishing Co.
- JACKENDOFF, RAY. 1992. Musical processing and musical affect. In Mari Riess Jones & Susan Holleran (eds.), *Cognitive Bases of Musical Communication*. Washington, D.C.: The American Psychological Association, p. 51-68.
- LERDAHL, FRED & RAY JACKENDOFF. 1977. Toward a formal theory of tonal music. *Journal of Music Theory* 21: 111-172.
1983. *A Generative Theory of Tonal Music*. Cambridge, MA: M.I.T. Press.
- MADDIESON, IAN. 1978. The frequency of tone. *UCLA Working Papers in Phonetics* 41: 43-52.
- MILLER, GEORGE. 1956. The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81-96.
- MASSARO, D. 1994. Psychological aspects of speech perception: Implications for research and theory. In Morton Ann Gernsbacher (ed.), *Handbook of Psycholinguistics*. San Diego, CA: Academic Press, 219-263.
- PETERSON, GORDON E. & BARNEY, HAROLD L. 1952. Control methods used in the study of vowels. *Journal of the Acoustical Society of America* 24: 175-184.
- ROEDERER, JUAN G. 1974. *Introduction to the Physics and Psychophysics of Music*. London: The English Universities Press.
- SNIDER, KEITH & HARRY VAN DER HULST. 1993. Issues in the representation of tonal register. In Harry van der Hulst & Keith Snider(eds.), *The Phonology of Tone: the Representation of Tonal Register*. Berlin: Mouton de Gruyter, 1-27.
- STUDDERT-KENNEDY, M., A.M. LIBERMAN, K. HARRIS & F.S. COOPER. 1970. The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review* 77: 234-249.
- VAN DEN BERG, ROB, CARLOS GUSSENHOVEN & TONI RIETVELD. 1992. Downstep in Dutch: implications for a model. In Gerard J. Docherty

- & D. Robert Ladd (eds), *Papers in Laboratory Phonology II*. Cambridge: Cambridge University Press, 335-359.
- WARD, W. DIXON & EDWARD M. BURNS. 1982. Absolute pitch. In D. Deutsch (ed.), *The Psychology of Music*. New York NY: Academic Press, 431-451.
- WELMERS, WILLIAM E. 1973. *African Language Structures*. Berkeley: University of California Press.
- WINOGRAD, T. 1968. Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory* 12: 3-49.