# The Use of Multi-pitch Patterns for Evaluating the Positive and Negative Valence of Emotional Speech

*Norman D. Cook & Takashi X. Fujisawa*

Department of Informatics
Kansai University, Takatsuki, Osaka, Japan
cook@res.kutc.kansai-u.ac.jp

## Abstract

We report the application of a psychophysical model of harmony perception to the analysis of speech intonation. The model was designed to reproduce the empirical findings on the perception of musical chords, but it does not depend on specific musical scales or tuning systems. Application to speech intonation produces values corresponding to the total dissonance, tension and affective valence among the dominant pitches used in the speech utterance.

## 1. Introduction

It is well known in both music and speech that a slower tempo, a lower pitch register and a smaller amplitude are commonly used to convey negative affect, whereas a faster tempo, higher pitch and greater amplitude are used to convey positive affect. From empirical studies of speech intonation, however, it is also known that statistics on pitch, tempo and volume alone do not allow for unambiguous distinctions between the negative affect of sadness and the positive affect of contentedness, nor between the negative affect of anger and the positive affect of happiness [1, 2]. From a musical perspective, what is missing from such analysis is the contribution of multi-pitch patterns. In a word, current techniques in speech analysis lack the concepts of melody and harmony, which are crucial for determining the affective valence of music, notably, the emotional content inherent to the major and minor modes (keys, scales and chords).

Because of the fixed pitches used in most music and the apparent cultural specificity of musical styles, most linguists have not pursued the analogy between music and speech (Fonagy [3] being a notable exception). As a consequence, the harmonic factors that play such an important role in music have not been studied in relation to speech. Despite of the known differences in the pitch phenomena of music and speech, we have studied the psychophysics of harmony perception with the purpose of developing a general model of pitch perception that does not rely on fixed pitches or on concepts from traditional Western harmony theory (but, importantly, can be applied to the harmonies of diatonic music). We have found that the positive/negative affect of the pitch combinations in music can be explained quantitatively on the basis of the acoustic signal and without recourse to the complex and culture-specific concepts of traditional music theory [4-7].

The psychophysical model (Section 2) is not complex, but the application to speech phenomena encounters two significant difficulties – one empirical and one "political". The empirical problem concerns how to extract the "dominant" pitches used in emotional speech from the continuously changing acoustic signal. We have chosen a technique in which the pitch profile is reconstructed using Gaussian base functions. The main pitches and their relative amplitudes can then be obtained in a fully-automated, quantitative manner that does not require editing of the raw pitch data. Alternative methods and linguistically more-sophisticated techniques for selecting important pitches are of course possible, but the above technique has the merit of using all of the pitch information in the utterance. In any case, the novel aspect of our technique is the quasi-musical analysis that is possible once extraction of the dominant pitch structure of the utterance has been achieved.

The political problem is how to encourage both linguists and music theorists to reconsider the acrimonious divide between the pitch phenomena of speech and music. As shown below, a general model of pitch perception requires that linguists familiarize themselves with the core ideas of musical harmony – and to suppress the tendency to declare proudly that they "know nothing about music!" By the same token, traditional music theorists must be willing to consider pitch phenomena that are fundamentally non-diatonic – whether music from other cultures or the pitch phenomena of speech. Theoretical questions about the *Origins of Music* [8] and its relation with language remain unanswered, but a basic understanding of both music theory and linguistics is required for a sensible dialog between these two realms.

## 2. Recent Experimental Work

Detailed discussion of our experimental results and of the harmony model can be found in Cook et al. [7] and are available over the internet [9]. The main conclusions of that work, summarized below, have prompted us to undertake a related analysis of the same data. Final results will be reported at SP2006, but the basic techniques and the rationale for the re-analysis will be explained here.

### 1.0 Cluster analysis

At the heart of our analysis is a method for producing a histogram of the pitches in a spoken sentence. This inevitably results in a complex, jagged curve with several dominant peaks that we then reconstruct as a summation of Gaussian functions using Bouman's (2002) algorithm [10]. The reconstruction provides us with a small number of pitches of known amplitude, which can then be used for musical analysis.

As shown in Figure 1, in the case of instrumentally played or sung melodies, the technique gives an accurate reconstruction of the pitch structure of the music. In the case of speech utterances, occasionally distinct pitch structure can be observed in the raw data (Figure 2), but more usually reconstruction using the Gaussian curves is necessary to simplify the raw pitch histogram (Figure 3), and allow for the quasi-musical analysis.

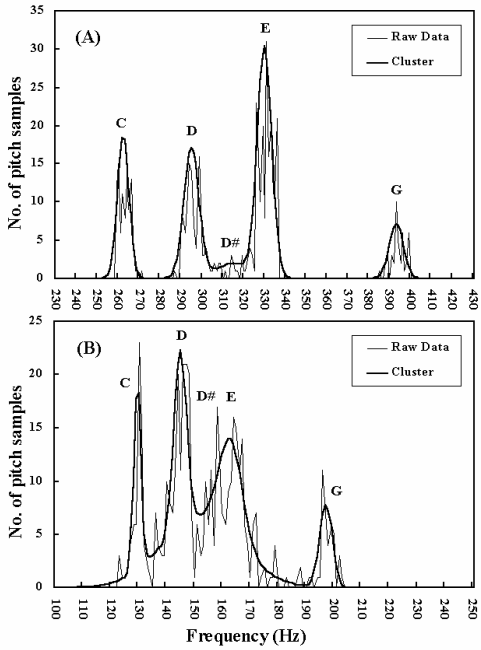


Fig. 1. The raw pitch data and the fitted clusters for the first 26 notes of "Mary Had a Little Lamb". (A) shows the narrow clusters obtained precisely at the four diatonic tones (CDEG) played on an electronic piano, and (B) shows the somewhat broader clusters obtained from a male voice sung one octave lower. Note that both versions show a fifth cluster at D#. The fifth cluster in the piano version is negligible, whereas that of the singer is a rather large deviation from the C-major melody, indicative of an amateur singer's difficulty in maintaining the major key modality.

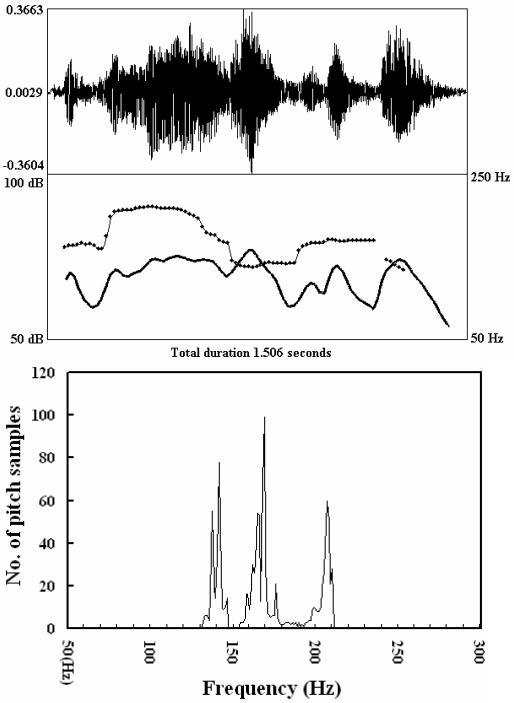


Fig. 2: Above, the wave-form, pitch and intensity curves for a Korean utterance. Below, Korean is not a tone-language, but the frequency histogram shows three distinct pitches.

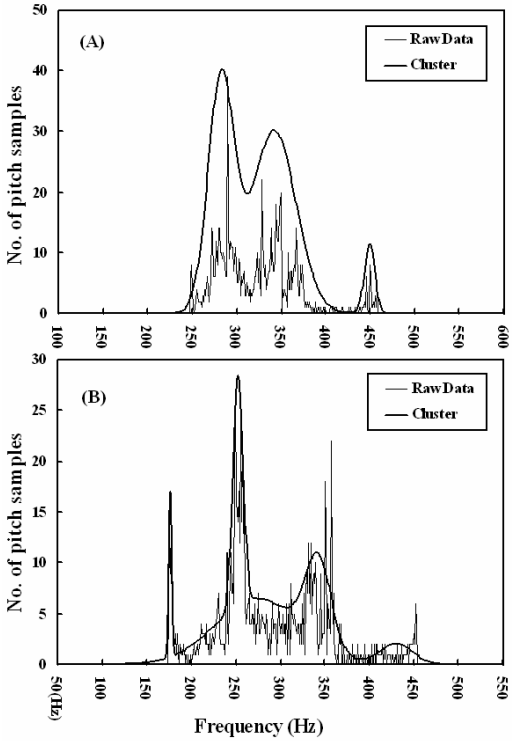


Fig. 3: Typical cases of the pitch distribution of Japanese speakers uttering emotional sentences. Several overlapping pitch regions are commonly found. In (A) clusters are at 283, 342 and 450 Hz. In (B) clusters are at 177, 253, 276, 343 and 431 Hz.

The pitch structure of most music is straightforward insofar as only the tones of musical scales are used. Normal speech, however, entails a continuous range of pitches not tied to any given scale. Empirically, certain pitch regions are often emphasized, others not emphasized, and the entire speech utterance usually displays a multimodal distribution around the speaker's tonic pitch. Figure 2 shows a particularly clear example of harmonic structure in a Korean speaker proclaiming that: "The salesman cheated me out of 200 dollars!" Figure 3 shows more typical examples of pitch substructure with fitted Gaussians.

Once the dominant pitches have been identified (using Bouman's cluster algorithm [10] or more traditional linguistic methods), three musical measures can be calculated using our model of pitch effects: (i) the total dissonance among pitch pairs, (ii) the total tension among pitch triplets, and (iii) the total major/minor modality among pitch triplets.

## 2.0 The harmony model

The first calculation is the total dissonance of pitch intervals, following the model of Plomp & Levelt [11] (Figure 4A). Variations on this model have previously been used to model the perception of the consonance/dissonance of musical intervals and are considered to be an important psychophysical basis for explaining music perception (e.g., Parncutt [12]).

As important as the dissonance curve in Figure 4A is for music psychology, it is noteworthy that the summation of the dissonance of multi-tone combinations ***fails*** to explain the relative stability/instability of the triads of diatonic music and does not lead to an understanding of the different affect of

major and minor chords. This problem has been discussed in detail elsewhere [5, 12]. Suffice it to say that the relative spacing of the tones in 3-tone chords (not solely the dissonance of 2-tone intervals) must be brought into consideration to explain the phenomena of harmony.

By measuring the difference in interval sizes among any 3 tones, the inherent tension of the chord can be calculated. Using the model curve in Figure 4B, the unresolved "tension chords" (i.e., diminished and augmented chords) show high tension values, whereas any of the resolved major or minor chords show low tension values – in good agreement with behavioral results (augmented>diminished>minor>major)[13].

The difference in magnitude of the intervals in a 3-tone chord can also be used to calculate the major or minor modality of the chord. As seen in Figure 4C, when the lower of the 2 intervals is larger than the upper interval (giving a positive difference score), the chord is major (and vice versa for minor). When the effects of upper partials are included in the calculation, the modality scores for all known triads are in agreement with their major or minor modality, as known from traditional harmony theory [5, 9].

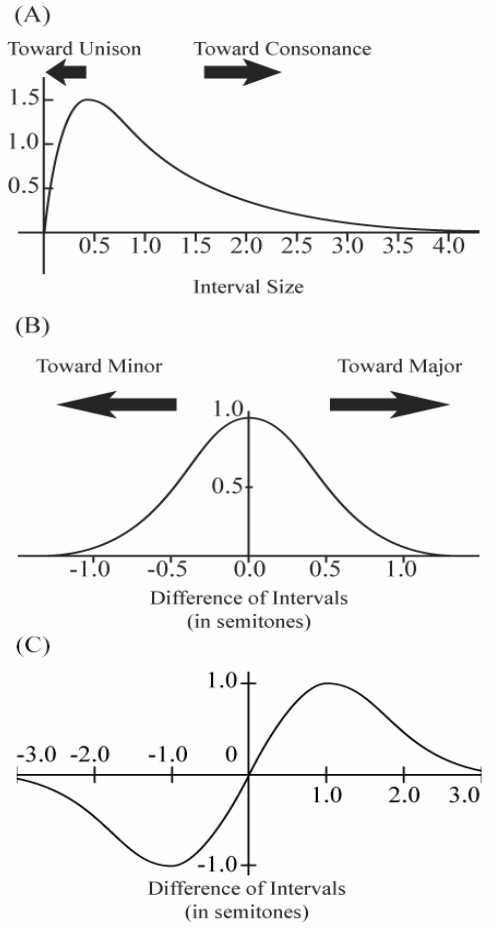


Fig. 4: The three factors underlying a psychophysical model of harmony: (A) defines the relative dissonance of any pair of tones; (B) defines the tension inherent to any three-tone combination, based on the relative size of the two intervals in any triad; and (C) defines the major-like (positive) or minor-like (negative) modality of any triad (for details, see Cook et al. [4-7, 9]).

### 3.0 Experimental results

The main advantage of the model outlined in Figure 4 is that it allows for a quasi-musical analysis of any combination of pitches, unrelated to musical scales. When pitch combinations approximate major or minor chords, a modality score near to 1.0 or –1.0 is obtained, but, for most speech samples, smaller (but generally non-zero) values are found. Our main result using this technique in the study of emotional prosody (144 utterances of positive and negative affect) was a significant difference in the modality of happy and sad sentences [7]. The statistical results were not overwhelming, but significant in the predicted direction. This result was interpreted as indicating a hidden "harmony" in emotional speech

### 4.0 The need for re-analysis

Application of the harmony model to speech phenomena gave quantitative indications of major-like and minor-like pitch combinations in, respectively, happy and sad utterances. Similar attempts by others have provided further evidence of a link between musical melody and speech prosody [14], but the musicality of speech is notable and indeed perceptible only in the special circumstances of motherese and school-teacherese. In other words, only when the intonation of speech is exaggerated in a fashion that might be described as operatic, melodramatic, "affected" or "motherly" does it reveal its musical acoustic properties.

The mystery therefore remains why music-like harmonies are so ***infrequent*** in normal speech – despite the fact that speakers typically use multiple (3-7) pitches over a range of 7-12 semitones in 1-3 second sentences [15] (e.g., Fig. 3). In other words, normal spoken prosody could theoretically be fully melodious in a musical sense, but clearly it is not (except under the unusual circumstances noted above). Although speech is not music, it is nonetheless true that normal people are quite capable of inferring the emotional state of others from pitch information without linguistic cues [1, 2] – in much the same way that non-musicians from cultures of the East and West, and children as young as 4 years-old can infer the affect of major and minor chords [5]. Are we therefore obliged to draw the non-parsimonious conclusion that there are completely different codes involved in deciphering the meaning of the pitch information in language and music? Or is there a deeper code [16]?

Indications of an answer come from the longstanding mystery of why patients with left hemisphere (LH) damage are still able to sing and why patients with right hemisphere (RH) damage are still able to produce and understand intonation for syntactic purposes. Apparently both hemispheres are involved in the pitch control of the voice – and this suggests the possibility that the tonotopic representation of musical harmony in the RH is modified during interhemispheric communications. It is this possibility that we have tested in a re-analysis of the same data as published previously [7].

Results of the new analysis will be reported at SP2006, but the basic idea is as follows: There is abundant evidence that the RH has a tonotopic representation of musical pitch in the form of cortical maps. Such maps could be the basis for the RH's superior performance in recognizing melodies, harmonies, musical affect and the major and minor modes. That tonotopic map can be utilized for the expression of affect directly from the RH in the form of song (or motherese, etc.), but is normally an indirect influence on the expressive speech

of the LH. Our hypothesis is that if the callosal transfer of the cortical affect of the RH is inhibitory, then precisely those tones that would serve to express the musical affect would be suppressed in the LH, leaving it to express emotional prosody through the tonotopically-organized tones that are not under the active suppression of the RH. The pattern of pitches used by the LH is thereby determined by the affective coding of the RH, but the tones available to the LH lie in-between the tones used in a musical sense.

This theoretical "inversion" of the tones activated in the LH and RH is analogous to the sense/anti-sense coding of genes by the two strands of DNA. Both versions have the same informational content, but only one is "meaningful" in the sense of coding for proteins (diatonic harmonies).

What this hypothesis implies is that speech utterances will be harmonic in a musical sense only in so far as the RH is in control (or RH affect is directly "leaked" in affected speech). At other times, emotional prosody will be heard as emotional precisely because the expected musical harmonies are suppressed. In effect, it is the non-barking of the dog in a Sherlock Holmes' mystery that provides the essential affective information.

## 3. Discussion

Similar to many others with an interest in prosody, our original motivation was to develop a quantitative technique that would allow distinctions between emotional states based solely on the non-linguistic acoustic signal of spoken utterances. Because the bulk of intonation theory has been developed for the narrow purposes of syntactic analysis, we have studied musical acoustics in search of techniques to deal quantitatively with pitch phenomena. To our surprise, we found that music psychology also does not have methods even for distinguishing between the major and minor modes – the classic and empirically strongest indicator of mood in all of diatonic music! In a word, there is no established theory of the "psycho-acoustics" of musical affect. As a consequence, when discussing the emotional effects of musical pieces, theorists invariably refer to the "established major key" or the role of the "minor third" – using concepts from music theory to explain auditory perception, rather than vice versa. This is comparable to the linguist distinguishing between a joyful utterance and an angry utterance by referring to the speaker's use of words meaning "hatred" in the sentence.

Belatedly, we discovered Meyer's (1956) explanation of musical phenomena from the perspective of Gestalt psychology [17]. His ideas are easily translated into a psycho-physical model of harmony [4-7, 9] and provide the basis for a coherent view of diatonic harmony that is consistent with both the musician's and the non-musician's common sense understanding of harmony. With the stalemate over the psychoacoustics of musical mode broken, the model has allowed us to return to the problems of speech prosody with a new tool. Its application to speech raises new questions about the relationship between music and language, but questions that can be answered experimentally.

## 4. Conclusions

Our conclusions are yet tentative. On the one hand, it appears likely that the affective valence of speech prosody depends critically on multiple pitch patterns, i.e., the "melody" of speech. On the other hand, speech is not melodious in an unambiguous, recognizable, musical sense. The pitch contour in most spoken utterances, even in spontaneous, unmistakably emotional circumstances, is distinctly prosodic, not musical.

The hypothesis that we are testing in the present study is based on the idea that the cortical representation of positive and negative affect takes the form of the excitation of tonotopic maps. Strictly within a musical context, that is a plausible view of the RH's cortical (as distinct from limbic) representation of musical affect. If, however, such tonotopic mapping of musical affect is essentially a RH function, then its transfer to the LH across the corpus callosum can take two main forms, corresponding to excitatory and inhibitory callosal effects. The implications of these hypotheses are straight-forward and will be examined in a re-analysis of previously published data [7].

## 5. References


[1] Scherer, K. R., 1986. Vocal affect expression. *Psychological Bulletin*, 99, 143-165.

[2] Scherer, K. R., Johnstone, T., & Klasmeyer, G., 2003. Vocal expression of emotion. In, R. D. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*, Oxford University Press, Oxford, 433-456.

[3] Fonagy, I., 2001. *Languages within Language*. Benjamins, Amsterdam.

[4] Cook, N.D., 2001. Explaining harmony. *Annals of the New York Academy of Sciences* 930, 382-385.

[5] Cook, N.D., 2002. *Tone of Voice and Mind*. Benjamins, Amsterdam.

[6] Cook, N.D., Fujisawa, T.X., & Takami, K., 2003. Evaluation of the affect of speech intonation using a model of the perception of interval dissonance and harmonic tension. *Eurospeech2003*, Geneva, Switzerland.

[7] Cook, N.D., Fujisawa, T.X., & Takami, K., 2006. Evaluation of the affective valence of speech using pitch substructure. *IEEE Transactions on Audio, Speech & Language Processing* 14, 142-151.

[8] Wallin, N.L., Merker, B., & Brown, S., 2000. *The Origins of Music*. MIT Press, Cambridge, Mass.

[9] http://www.res.kutc.kansai-u.ac.jp/~cook.

[10] Bouman, C.A., 2002. *Cluster: an unsupervised algorithm for modeling Gaussian mixtures.* www.ece.purdue.edu/ ~bouman.

[11] Plomp, R., & Levelt, W.J.M., 1965. Total consonance and critical bandwidth. *Journal of the American Society for Acoustics* 38, 548-560.

[12] Parncutt, R., 1989. *Harmony: A psychoacoustical approach*. Springer, New York.

[13] Roberts, L.A., 1986. Consonant judgments of musical chords by musicians and untrained listeners. *Acustica* 62, 163-171.

[14] Schreuder, M., van Eerten L., & Gilbers, D., 2005. Speaking in minor and major keys. http://odur.let.rug.nl/ ~schreudr

[15] Fitzsimons, M., Sheahan, N., & Staunton, H., 2001. Gender and the integration of acoustic dimensions of prosody. *Brain & Language* 78, 94-108.

[16] Juslin, P.N., & Laukka, P., 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* 129, 770-814.

[17] Meyer, L., 1956. *Emotion and Meaning in Music*. Chicago, Chicago University Press.