JOSHUA STEELE 1775: SPEECH INTONATION AND MUSIC TONALITY

Hunter Hatfield, Linguistics

ABSTRACT

In 1775, Joshua Steele published *An essay towards establishing the melody and measure of speech to be expressed and perpetuated by peculiar symbols*. Steele documented English intonation and proposed a prosodic notation. It is argued that Steele successfully documented that speech varies in pitch, and that the notation system he developed was as accurate a phonetic representation as could be expected at that time. His musical notation, however, was hampered by the lack of a distinction between phonology and phonetics. A musical notation also inherently asserts certain claims about the structure of tonal space, which Steele was not able to justify.

1.0. INTRODUCTION

In 1775, Joshua Steele published *An essay towards establishing the melody and measure of speech to be expressed and perpetuated by peculiar symbols*. This essay was the first treatise in the English language devoted to the study of intonation in speech. The primary motivation for the essay was in fact to rebut comments from an earlier essay, *The origin and progress of language* by James Burnet, in which it was asserted that there is no change in tone in English accents. Steele's essay argues against this claim by providing a method to hear the melodic aspects of speech and, moreover, proposes a notation for speech intonation so that knowledge of "the types of modern elocution may be transmitted to posterity as accurately as we have received the musical compositions of Correlli." (Steele, 1775, p. 14) Naturally, Steele's interests in speech prosody are rather different than contemporary linguistic interests. For instance, for Steele a great outcome from his essay would be the addition of a bass accompaniment to the actor's voice in stage performances—a perhaps worthwhile goal that is yet rarely argued for in contemporary linguistic research.

Despite such different purposes, Steele's essay is quite linguistically nuanced. For instance, he argues for the notation of at least five separate properties of speech intonation – accent, emphasis, quantity, pause, and force, and believes one of the reasons the study of speech intonation had made little progress in his day is because of the conflation of all these items into only a couple terms (Steele, 1775, p. viii). Even 230 years later, working out the exact relationships between pitch, stress, duration, breaks, and intensity, respectively, remains a dominant problem in prosodic research. Additionally, while Steele developed "peculiar symbols" based upon musical notation for recording intonation, he was not under the impression that pitch in speech was strictly speaking musical. Indeed he expanded the musical staff so that pitch targets in speech that did not correspond to a musical semitone could be noted. Indeed one of his main conclusions is "[t]hat the changes of voice from *acute* to *grave* and *vice versa* do not proceed by pointed degrees coinciding with the divisions of the chromatic-diatonic scale; but by gradations that seem infinitely smaller (which we call *slides*)" (Steele, 1775, p. 17; italics in original).

This paper will outline what Steele's goals were with the *Essay* and assess whether or not he met them. It will also walk the line between considering Steele in his historical context to see just what was possible at that time, but also take Steele seriously as a scholar of speech intonation and look at his argument critically. In particular, Steele, despite arguing that speech does not use the chromatic-diatonic scale for its tonal targets, does in fact use a modified music notation for speech. Does this remain a good idea, or does the current world of pitch tracks calculated by software algorithms and ToBI transcriptions render it obsolete? And if measurements of pure acoustic frequency are superior to music notation, why exactly? Answering such questions comes to be the primary point of this paper, because it makes one assess what exactly is being measured with current research tools and its relationship to speech. As a final point before moving on, it is worth noting that Steele's essay concerns the melody and measure of speech, while this paper will look exclusively at melody, or more precisely, tone. This decision is not made out of any perceived lack of interesting comments on the rhythm of speech by Steele, but simply to contain the size of the current paper.

2.0. STEELE'S RESEARCH ON ACCENT

The primary purpose for the publication of Steele's *Essay* was to rebut some comments in the influential *The origin and progress of language* by James Burnet. Indeed the majority of the essay is composed of a published dialogue between Steele and Burnet, in which, by the end, Burnet withdraws many of his major claims about intonation in language. Specifically, Steele wished to refute statements from Burnet such as: "We have accents in English…; but there is no change of the tone in them; the voice is only raised more, so as to be louder upon one syllable than another. …[T]hen is the music of our language, in this respect, nothing better than the music of a drum…." (quoted in Steele, 1775, p. 3).

Other than showing Burnet's error, Steele had several other purposes with his *Essay.* First, he hoped to launch a field of intonation research. He states at one point that he hopes we have "found the land" and future adventurers will follow to improve upon his work. Second, developing a system of notation for speech intonation was critical to Steele because it would allow knowledge transfer between scholars and well as communicating with posterity. Third, such a notation would lead to the improvement of the arts such as stage performance, rhetoric, and general elocution.

As far as pitch, or in Steele's term, accent, is concerned, he argued that accent in speech moves up and down on each syllable in a sliding motion, the voice never resting on a single pitch, except perhaps the end note. This is in opposition to most musical idioms, including Western classical music, which uses discrete pitches in a melody. Moreover, the melodic slides of speech are not part of the musical diatonic scale, but much finer—perhaps quarter tones or smaller (Steele, 1775, pp. 1–12, *inter alia*). In the Western musical scales to which Steele was comparing, there are 12 notes in the chromatic scale, each a half-step or semi-tone apart. In the equal-tempered scales of his day, and most Western music still, these semi-tones are simply 12 equal (logarithmic) steps between a note and its octave equivalent. By asserting that speech operated at the quarter tone level, Steele was asserting that each of the semi-tones could be divided in half or more.[1]

Steele based these claims on the following method for observing pitch: He employed a bass viol, which allows for sliding tones due to its lack of frets along the board. He noted that any instrument which allows for sliding tones would suffice. For the bass viol, paper is attached to the side of the board with each note marked to the quarter tone level. Steele then drew the bow across a string, setting the string vibrating. He would then slide his hand up and down the board as he spoke until he had each turning point marked. Where his finger stopped, say E-sharp, he would note it.

After this, he developed a notation for recording his research. Since speech is composed of slides, and not stable, discrete pitches, he developed note heads which were lines, curves, and circumflex notes that could cover several pitches. To these curved heads, he then added a tail which marked the duration of the syllable in a number of beats within a musical bar. These notes were then placed on a modified musical staff. Figure 1 shows Steele's notation of a line from Alexander Pope.

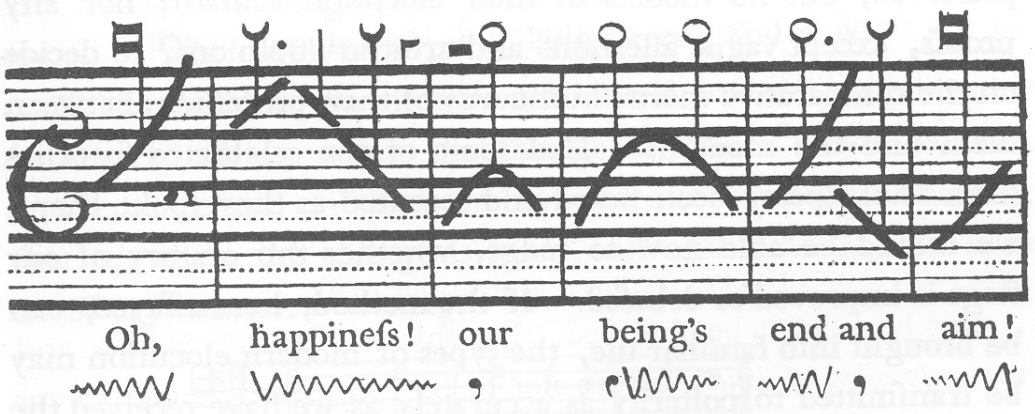

Figure 1. Transcription of Pope's Happiness (Steele, 1775, p. 13)

The broad strokes on each syllable are the decurved note heads, so the word *Oh* is transcribed here as sliding from the thick line right in the middle of the staff to a quarter tone above the staff. Also, below each word, note the wavy lines, which are Steele's invented notation for Force (intensity or loudness). Thus, *happiness* begins loud and ends soft, which indeed it would if one considers the sonority of the first low vowel as compared to the high vowel and fricative at the end. One distinguishing feature of this notation that Steele rarely draws attention to is that the note values are not mentioned here, i.e., there is no record of which notes each line represents. This is not because Steele did not have a fuller notation which record the exact quarter-tone for each syllable, but merely to save time notating when it was not critical to his point. In fact, in most of Steele's notations, even the staff itself is dropped for the most basic hints of the slides' direction and the levels they move through. One can see how cumbersome the notation can be by examining Figure 2, where each quarter-tone is labeled.

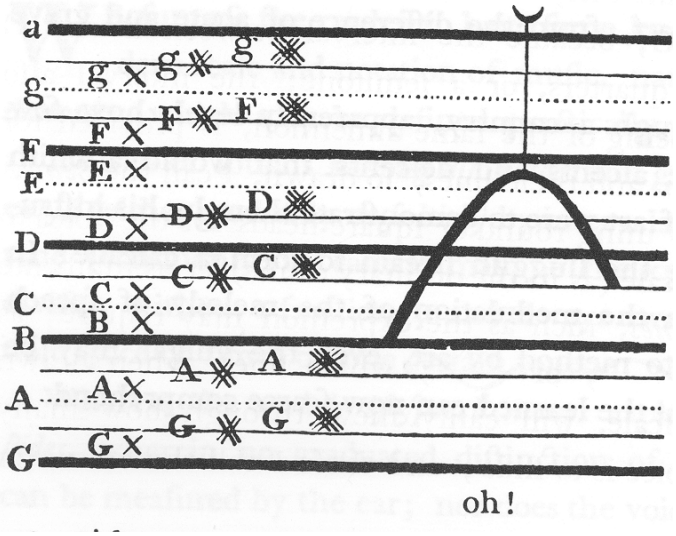

Figure 2. The word *Oh* with all staff lines marked (Steele, 1775, p. 11)

In Figure 2, each x marks an additional quarter-tone. So starting from the bottom, we have G, G+1/4, G+2/4, G+3/4, then A. There is only one quarter-tone between B and C, and between E and F, because those notes are only a semi-tone apart, instead of a whole tone.

Using these methods, Steele made numerous observations about speech accents which have stood the test of time well. For instance, looking through his transcriptions, one notes several micro-prosodic effects. For instance, the /i/ in a diphthong such as /ai/ is frequently noted higher than its companion. Steele had quite nuanced curves for speech slides. One distinction he made was between a straight slide upwards through a syllable versus a curved shape where the first part of the syllable changes pitch relatively slowly, followed by a rapid rise up or down, (Steele, 1775, p. 132). Steele was also quite aware of many of the semantic and pragmatic uses of intonation in English. He noted that sentences with a suspended meaning frequently rise at the end and that "compleated" ones go down. He noted that a question can differ from a statement with only a change in intonation, for example: "That's a bear." versus "That's a bear?" Also, intonation can be used to convey understanding and emotional weight in a conversation. In addition to all this, Steele noted that accent can vary between individual speakers and, more, between dialects of English.

Above are some of the Steelian observations which we still find evidence for today. Naturally, Steele had other interests as well concerning his melodic speech. For instance, he believed his voice had a basic home note, a musical tonic, and that the range of his voice was about a 5th, the 5th being a musical interval where the frequency ratio between the 5th and the tonic (or 1st) is essentially 1.5:1. He practiced the art of accompanying his voice with the bass viol and claimed it seemed quite pleasing when the voice was accompanied by playing the tonic, the 5th, or the 4th musical intervals from his voice's home tonic. He did note, however, that, while his range was approximately a 5th, different people had different ranges, and that was one distinguishing characteristic of English dialects.

3.0. ASSESSMENT—THE PERSPECTIVE OF 1775

We come now to the question of whether any of Steele's methodology was a good idea. Namely, can one truly use a bass viol to notate speech and does this notation system work? First looking at this question from the perspective of 1775, the answer has to be essentially "yes". Current psycho-acoustic and music perception research has shown that people can distinguish hundreds of pitches with an octave. Moreover, several musical idioms in the world, such as Indian and Arabic-Persian, commonly use quarter-tone notes in their musical scales. (Burns, 1999) The relevance of this is simply that humans are fully capable of hearing pitch fluctuations at this level of detail.

Moreover, there was no better way of measuring the frequency of sound waves at the time. For instance, in 1637, the French monk Marsenne constructed 30-meter long ropes that were tense enough to be vibrated. Since the ropes were of this magnitude, it was possible to watch their movement and count the number of waves that vibrated on them. While this is intriguing, it is certainly not very useful for Steele's purposes. It wasn't until 1819, 44 years later after publication of the *Essay*, that the Siren was invented which allowed a person to set it moving at a certain frequency and count the revolutions (Beament, 2001). Even these two inventions still depended on hearing, as they produced a note that a human had to judge in order to know what note you were measuring the frequency of. Thus, in 1775, the human ear was the best method around for measuring frequency.

How about the notation? In one sense it clearly works, as one can still today look at Steele's notations and get the basic idea of how something was said. Some of the transcriptions seem markedly unnatural at times, but that does not disprove the basic validity of the notation. More interestingly, Steele seems to have been largely accurate in his use of the quarter-tone as a sufficiently precise measure of pitch changes in speech. As a test, the author of this paper recorded himself speaking the "Oh Happiness" quote from Pope above, based on Steele's transcription. Next, the recording's absolute frequency was measured using the pitch tracking algorithm of the Praat software package. Each bend in the frequency measurement was noted. Each of the bends was converted from a Hertz unit of measurement to a corresponding quarter-tone note. (Quarter-tone frequency calculations were made by simply dividing the ratio between 2 semi-tones in half.) All of this was then marked in Steele's notation. The result is shown in Figure 3.

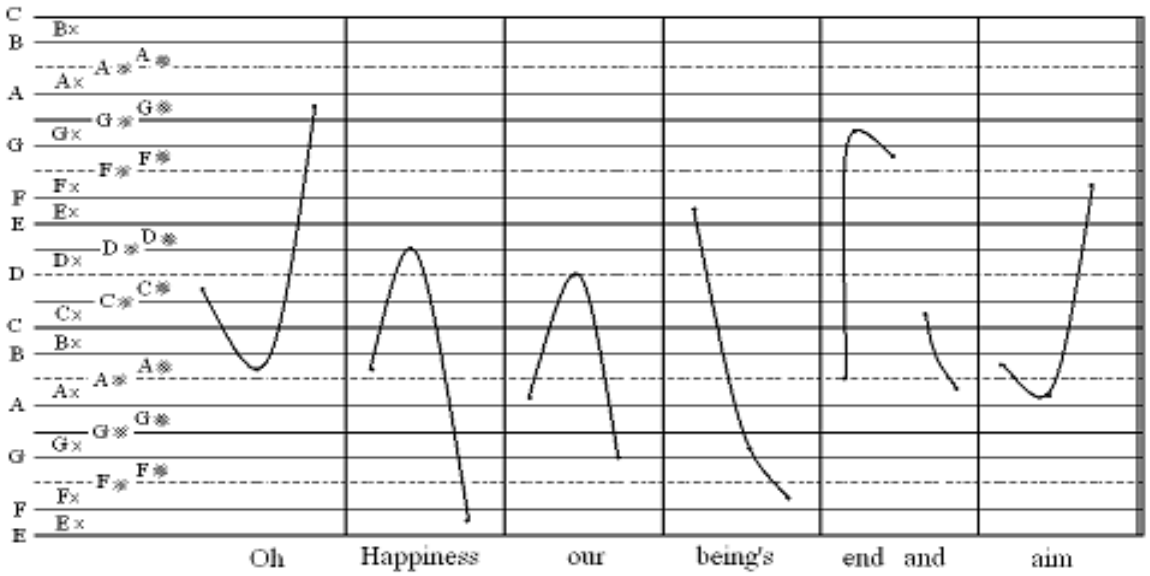

Figure 3: Translation of frequency to quarter-tones

What is significant about Figure 3 is not that it matches Steele's transcription, which it does not. It is simply that (1) the detail available here is more than sufficient to do any type of linguistic research on

intonation, and (2) the difference between absolute frequency in Hertz and a quarter-tone value was never more than 3 Hertz. At present, there is no reason to believe that differences smaller than this are relevant linguistically.

Of course, there are some confusions in Steele's work that would be noticed even in Steele's historical context. On the one hand, Steele makes several comments that Accent is invariable. For instance, he states that accents are determined except for grace note, grace notes being melodic flourishes that are not part of a main melodic line. At another time, he indicates that language has fixed accents on syllables, though he admits some speakers are "indecisive". Steele has clearly distinguished between stress and pitch (Emphasis and Accent in his tones) and he is certainly speaking of tone here. Thus, this is a claim that there are fixed pitch moves that are at a minimum "best" on a syllable, once its context in discourse is specified.

On the other hand, Steele makes comments implying that accents are in fact variable. He states that there are as many accents as tempers and features in men. At another point, he mentions that 2 or 3 quarter-tones this way or that make little difference. Finally, as mentioned earlier, Steele quite frequently drops large pieces of his notation, as if they are not really too important, as long as we know the general movement of a syllable.

In the current, dominant research paradigm in speech intonation, namely the Autosegmental-Metrical approach (e.g., Ladd, 1996), these issues would be resolved by distinguishing between phonology and phonetics. Many of the tones of a language are specified in the language, and knowing those fixed tones is part of what it means to be a competent speaker of the language. However, there is great variation in the phonetic realization of many of these tones. In the end, though, it is hard to fault Steele for this confusion, as Baudouin de Courtenay, generally credited with discovering the phoneme, was not even born until 1845—70 years after Steele's essay.

4.0. MUSIC NOTATION FOR SPEECH

In this section, we will look at whether or not using a modified music notation remains the best notation for speech. Steele's notation works fairly well, but there are many other intonation transcription systems now. Most of these transcription systems start with a measurement of absolute frequency of the utterance as the basis for research, then typically add phonological tones and breaks. The best well-known of such systems is the Tones and Break Indices system (ToBI) (Silverman et al., 1992). Does Steele's notation stack up favorably against ToBI and other such transcription systems? If not, precisely why not? I will argue here that Steele's notation is no longer the best, as a musical notation inherently suggests a certain structure to the tonal space of speech, which just does not seem to be present.

To decide whether or not musical notation is proper for speech, we have to decide what makes a notation particularly "musical". Surveying many of the world's transcription systems, there appear to be two main types—ones based on particular instruments (put your finger here, press this button, etc.) and ones based upon musical intervals.[2] We will set the instrument-based notations aside, as Steele's notation makes no reference to the human instrument—vocal folds, shape of the oral cavity, etc. A musical interval, briefly, is a musical relationship between two pitches. Intervals relate to frequency relationships, but are not equivalent. For instance, intervals can have slightly different frequency ratios and be considered the same interval. Indeed the precise ratios between notes have been modified over time, moving from the Pythagorean to Just Temperament to Equal Temperament. Also, intervals have been discovered all over the world, often without direct knowledge of the ratios between them, but instead based simply on perception. Most importantly, the way musical intervals relate to one another does not follow the physical frequency measurements. An *A* note of 220Hz and an *A* note of 440 Hz are quite far apart physically, but are in fact considered the same note musically; they form an octave. Similarly, a musical note is considered musically closer to its 5th than it is to a note adjacent to it in frequency measurements.

So the question, of course, is whether speech-intonation is based on musical intervals of this type—not intervals necessarily of the diatonic scale, but just musical intervals *simpliciter*. One precondition for intervals in speech would be that speech uses some sort of target, some consistent location, for its high and low tones. If there is no intonational target in speech, then the target cannot be an interval. The good news is that there is considerable evidence that speech indeed has tonal targets. When speech is normalized for a speaker's pitch range, it turns out that they quite reliably hit the same targets in the same intonation patterns (Ladd, 1996, p.

262–265). So, between multiple speakers and between one speaker at different points in time, intonation has consistent tonal points it is aiming for.

The problem is that these targets are only consistent once you normalize for pitch range, Yet, pitch range itself in speech is quite variable. Different speakers, who might have the same basic "tonic," frequently have different ranges or excursion sizes. The same speaker can employ different ranges in different emotional states (boredom, excitement, etc.). In addition, a pitch accent can be much higher on one occasion than it is on another and yet be considered the same phonological tone. In other words, speech has tonal targets, but only within a highly variable speech range. Imagine that English intonation was based on musical intervals. Let's say that the English High tone is a 5th above the utterances initial position. Then, all English speakers aim for a 5th when they wish to use a High tone to, say, lend prominence to a syllable. Moreover, let's say that a 5th for Speaker A is at the top of his range. Remember that speech targets are consistent within pitch range, so a High tone might consistently be at 95% of the range utterance. Now, we come to the problem. If my pitch range is smaller when I am bored, but the High tone is still 95% of that range, I have just modified the interval that we hypothesized my intonation was based on. We have scrunched it up into a compressed space, changing the absolute excursion size, and changing the interval.[3] On one hand, speech tonality is like musical intervals because it is not an absolute frequency that is important. Instead, it is a relationship between tones in the intonational phrase. On the other hand, the tonal space which speech occupies, with its stretching and collapsing range, has no parallel in music.

None of this is to say that speech has no predictable structure; the argument is just that the structure of intonation is not the structure of musical tonality. If, as argued, a musical notation implies that speech does have a musical tonal structure, then it is attributing a structure to speech which is in fact absent. No matter how many reminders one attached to a musical transcription of speech, saying that the speech tones are not true intervals, it remains inherently misleading. It is worth noting that this matches recent neurological studies of people with musical disabilities, where the loss of an ability to perceive general melodic contours in music goes hand-in-hand with the loss of an ability to perceive intonation, but a loss in the perception of musical tonality does not appear to have any effect on the perception of speech (Patel & Peretz, 1997; Patel, Peretz, Tramo, & Labreque, 1998).

5.0. CONCLUSION

This paper has been a basic introduction to Steele's work on melody. I recommend to the reader a look at the Measure sections as well, as many of his observations about metrical structure in speech and verse remain quite relevant today. Besides introducing the reader to Steele's *Essay*, it has been argued that Steele's descriptions of intonation and notations are essentially accurate and functional; however, a system of music notation for speech will inevitably draw researchers into assuming a parallel with music that is unfounded.

NOTES

1. For more on scales, temperament, and the like, see Burns 1999.
2. For a look at musical notations other than the standard Western staff, see Kaufmann (1972), which contains various notation systems from East, Central, and South Asia. A commonality over the centuries and cultures is some form of notating musical intervals, in either absolute or movable terms. See also Upitis (1992) where children re-invent both interval and instrument based notations.
3. After making this argument at a conference, I have since discovered largely the same argument in an unpublished paper by Jackendoff and Lerdahl (2004). Taking the opposite angle, they look at speech intonation to see if it could be a source of musical tonality. Their conclusion is that it cannot for the same issues of pitch range that I raise: "Moreover, in the course of down-drift the frequency ratio between high and low tones get smaller…. In music, the parallel would be a melody in which not only the pitches sagged gradually in the course of a phrase, as if a recording were slowing down, but the intervals also got smaller, octaves gradually degrading to fifths, fifths to thirds, and so on." (Jackendoff & Lerdahl, 2004)

WORKS CITED

Beament, J. (2001). *How we hear music: The relationship between music and the hearing mechanism*. Woodbridge, UK: The Boydell Press.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (2nd ed.) (pp. 215–264). San Diego: Academic Press.

Jackendoff, R., & Lerdahl, F. (2004). The capacity for music: What is it, and what's special about it. Ms., Brandeis University.

Kaufmann, W. (1972). *Musical notations of the Orient: Notational systems of continental East, South, and Central Asia*. Gloucester, MA: Peter Smith.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological approach. In I. Deliege & J. Sloboda (Eds.), *Perception and cognition of music*. East Sussex, UK: Psychology Press.

Patel, A. D., Peretz, I., Tramo, M., Labreque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain and Language, 61*, 123–144.

Silverman K. E. A., Beckman, M, Pitrelli, J. F., Ostendorf, M., Wrightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). TOBI: A standard for labeling English prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing, Vol. 2* (pp. 867–870). Banff, Canada: Institute of Phonetic Sciences.

Steele, J. (1775, 1969). *An essay towards establishing the melody and measure of speech*. Menston, UK: The Scolar Press Limited.

Upitis, R. (1992). *Can I play you my song? The compositions and invented notations of children*. Portsmouth, NH: Heinemann.