23

Iconic Pitch Expresses Vertical Space

NATHANIEL CLARK, MARCUS PERLMAN, AND MARLENE JOHANSSON FALCK

1 Introduction

As bipedal primates, *Homo sapiens* experience the world from a distinctly vertical perspective as they move through and manipulate their environment. Within the first few years of development, human infants grow into mobile, upright toddlers. They learn to balance and walk on two legs, climb up and down, reach for objects high and low, and unfortunately, fall from time to time. Motivated by its prominent role in human experience and development, the conceptual schema of vertical space serves as the basis for making sense of a variety of more abstract concepts (Lakoff and Johnson 1980). Work in cognitive linguistics, for example, has identified a number of conceptual metaphors, so-called orientational metaphors, related to vertical space, including MORE IS UP, LESS IS DOWN; HAPPY IS UP, SAD IS DOWN; and GOOD IS UP, BAD IS DOWN. These metaphors manifest in much of the language we use (Lakoff and Johnson 1980; 1999), in our gestures (Casasanto 2008; Cienki 1998; Cienki and Müller 2008), as well as in noncommunicative tasks (Meier and Robinson 2004).

For speakers of many languages, verticality also serves to structure how people talk about, conceptualize, and experience auditory pitch (Pratt 1930; Rusconi et al. 2005; Zbikowski 1998). In English, the use of vertical spatial terms is the standard (and practically sole) convention to talk about the pitch of sounds. Indeed, it is difficult to avoid tautology in describing this relationship: relatively 'high' fundamental frequencies are described with terms referring to high vertical space, and sounds with relatively 'low' fun-

Language and the Creative Mind.
Michael Borkent, Barbara Dancygier, and Jennifer Hinnell (eds).

damental frequencies are described with low terms. English speakers talk about *high* and *low* musical tones and voices, moving *up* or *down* the scale in music, *rising* and *falling* pitches, and so on. While this mapping is not universal (other languages use *thin*/*thick* and *small*/*large*, for example), it is prevalent in most Western European languages, plus many other non-Indo-European languages (Ashley 2004; Shayan, Ozturk, and Sicoli 2011).

The basis for using vertical spatial terminology for pitch is argued to be motivated by embodied experience, specifically the different parts of the body that resonate during the production of high and low pitches (Pratt 1930; Shayan et al. 2011; Zbikowski 1998). As Zbikowski explains, 'When we make *low* sounds, our chest resonates; when we make *high* sounds, our chest no longer resonates in the same way, and the source of the sound seems located near our head.' Additionally, the position of the larynx further motivates this metaphorical association. The larynx is lowered in the production of lower frequencies and raised for higher frequencies, an action that is experienced proprioceptively by speakers and is sometimes visible in movements of the Adam's apple.

Experiments show that the use of spatial terms to talk about pitch is not just a verbal phenomenon, but reflects a deeper metaphorical conceptualization. A number of studies — couched variously in terms of cross-modal associations, synesthesia, and stimulus-response compatibility — demonstrate an interaction between the processing of pitch and spatial position (Eitan and Timmers 2010; Marks 1996). One pioneering study found that English speakers tend to localize high- and low-pitched sounds as emanating from high and low spatial positions, respectively (Pratt 1930). Participants listened to one of five tones varying by octave intervals that were played in random order by hidden speakers at five randomized vertical positions. Their task was to indicate the spatial position of the sound source on a numbered scale running from the floor to the ceiling. The study found that participants were highly consistent in locating higher pitched tones higher in the vertical array than lower pitched tones. As Pratt expresses, 'The results are clear-cut and unequivocal. *High tones are phenomenologically higher in space than low ones*' (p. 283, italics in original). These results were later confirmed in a more rigorous replication by Trimble (1934), which found the same pattern with nine tones played from a fixed sound source. Roffler and Butler (1968) later expanded this paradigm by manipulating the listener's orientation and distance to the sound source. Additionally, the study found similar results with congenitally blind participants and 4- to 5-year-old children.

Despite the robust findings of these studies, they are subject to the methodological criticism that spatial processing is an explicit part of the task. Thus, people might not normally conflate the perception of pitch and

vertical space, but are primed to do so because of the spatial nature of the task. More recently, Rusconi, Kwan, Giordano, Umiltà, and Butterworth (2005) addressed this confound, demonstrating what they called the Spatial-Musical Association of Response Codes effect (the 'SMARC' effect). For example, in one experiment, musically naïve participants performed a pitch comparison task, indicating whether a probe tone was higher or lower in pitch than a reference tone. Compatible responses — higher frequency tones indicated with a button press at the 'top' of the keyboard ('6') and lower tones with a button press at the 'bottom' of the keyboard ('spacebar') — were made significantly faster than incompatible responses with the buttons reversed. Further experiments found that the effect of stimulus-response compatibility was somewhat attenuated in an instrument identification task (wind or percussion) that did not attend explicitly to pitch, but the compatibility effect was generally strengthened by musical expertise.

Further experimental evidence comes from studies investigating cross-modal, synesthetic correspondences between the auditory dimension of pitch and the visual dimension of vertical position (Ben-Artzi and Marks 1995; Melara and O'Brien 1987; also see similar interactions in infants: Wagner, Winner, Cicchetti, and Gardner 1981). In one set of studies, for example, Melara and O'Brien presented participants with compound stimuli consisting of high- or low-pitched tones matched with high or low positioned dots. Participants made speeded classifications of the stimuli along one or the other of these dimensions, with the irrelevant dimension varied orthogonally in one version of the experiment, or held constant in another. The categorization of either dimension was slower when the irrelevant dimension was varied orthogonally, characteristic of Garner interference. Additionally, classification times exhibited Stroop interference, with faster classifications for congruent combinations of stimuli and slower times for incongruent combinations.

Vocal Iconic Gesture

These various studies provide strong converging evidence that speakers of English and numerous other languages may use vertical space as a source domain to conceptualize auditory pitch. Furthermore, phenomena like the SMARC effect, Garner interference, and Stroop interference suggest that people might also incorporate pitch into their conceptualization of vertical space. These phenomena raise the intuitive possibility that people might sometimes modulate the pitch of their voice when talking about meanings related to high or low vertical space. Indeed, evidence for such spatially motivated modulations of vocal pitch can be readily found in popular music, such as in Garth Brook's hit country song *Friends in Low Places* (Perl-

man 2010). As Brooks sings the line from the chorus, 'I've got friends in low places,' he uses the pitch of his voice to convey a sense of low vertical space, singing the word 'low' at the bottom of his vocal range, a major sixth below the prior note. This vocal effect places an iconic emphasis on the sense of lowness he is expressing, which metaphorically extends to the target domain of social class. Perlman suggests that these kinds of iconic vocal effects are not restricted to performative contexts like popular music, but instead reflect the common process of conceptualizing different kinds of semantic domains in terms of movements of the vocal tract.

Gesture scholars tout the value of studying manual iconic gestures as a window into the conceptualizations that underlie spoken communication and thought (Cienki and Müller 2008; Kendon 2004; McNeill 1992; 2005), and recent research suggests that conceptualizations are also revealed in iconic modulations of the speech signal (Nygaard, Herold, and Namy 2009; Okrent 2002; Perlman 2010; Perlman, Clark, and Johansson Falck, [under review]; Shintel, Nusbaum, and Okrent 2006). These modulations bear important functional similarity to iconic gestures in the visual-kinesthetic modality. Like manual gestures, vocal iconic gestures manifest as spontaneously formed, conceptually motivated movements that tend to be co-expressive with co-occurring speech (Okrent 2002; Perlman 2010). For example, Perlman demonstrated that people spontaneously modulated their speech rate in iconic correspondence with spontaneous, open-ended descriptions of various fast or slow events presented in video clips. Participants tended to speak faster or slower as they described fast or slow events, and especially as they articulated adverbial phrases related to their speed.

Similar results were found in a more controlled study in which participants used prescribed sentences to describe the direction of motion of an animated dot on a computer screen (Shintel et al. 2006). In one experiment, participants described a dot as moving to the left or right, while it also traveled either quickly or slowly. The results showed that participants spoke the prescribed sentences with a longer duration for the slow moving dot compared to the fast one, even though the dot's speed was incidental to the explicit communicative task. A second experiment in this paradigm provided evidence that people also produce iconic gestures of pitch related to the expression of vertical space. Participants increased or decreased their fundamental frequency as they described the dot moving upward or downward, respectively.

Finally, in another study, participants were recorded as they read short vignettes with physical and metaphorical instantiations of meanings related to speed and size, such as someone racing down a highway or signing an enormous contract (Perlman et al. [under review]). Analysis revealed that participants spoke faster when describing fast physical motion, and used

lower pitches when describing both physically and metaphorically large entities. These results not only provide further documentation of the iconic mapping between speech rate and pitch, but also indicate the metaphorical potential of these acoustic features.

2 The Current Study

The current study aimed to explore the possibility that people produce vocal gestures of pitch related to spatial vertical meanings. To assess this possibility, we used a similar methodology as in our earlier exploration of the speech rate-speed and pitch-size mappings (Perlman et al. [under review]). Participants read short stories that related to different meanings communicated by terms of verticality: actual vertical space, emotion, and auditory pitch. For each story, they read either a *high* version in one condition or a *low* version in the other condition. We predicted that participants would use a higher pitch when reading stories related to positive emotion, high vertical space, and high auditory pitch, and a lower pitch when reading stories related to negative emotion, low vertical space, and low auditory pitch.

These three target domains represent an array of different mapping types from conceptualizations of verticality to its realization in acoustic pitch. The mapping from emotion to pitch is a link that is generally well-established. Further, prosody is widely acknowledged to convey affective information (Bolinger 1986), and under some views this function of prosody is universal and possibly evolutionarily deep (Cosmides 1983). In reporting a sound, the specific mapping from the pitch of a reported sound to the pitch of the report of the sound is mimetic rather than metaphorical. In contrast, though, the mapping from vertical space to pitch is cross-modal, exploiting grounded elements of the source domain of pitch to express information about the target domain of vertical space.

Additionally, this study introduces a powerful new methodology for the analysis of vocal gestures. This method of Acoustic Contour Analysis, based on the analysis of mousetracking trajectories, addresses a weakness of previous studies that have examined iconicity in speech prosody. Previous studies have generally focused on the averages of prosodic variables over intervals of speech such as a word, phrase, or entire sentence (e.g., Nygaard et al. 2009; Perlman 2010; Shintel et al. 2006). Consequently, these analyses lose information of the temporal dimension of speech and thus the contour of the acoustic variable. For example, two phrases may have very different pitch contours (starting low and rising vs starting high and falling), but very similar averages across the contour (See Figure 1).

3 Methods

3.1 Participants

Thirty-four undergraduate participants from the University of California, Santa Cruz received course credit in exchange for their time. All participants were self-reported native speakers of English. Two participants' data were excluded due to excessive disfluencies, leaving a total of 32 participants (20 women) in the analysis.

3.2 Materials

Four pairs of short target stories were created, which included two stories related to vertical motion, one story related to emotion, and one story related to a reported sound. (Full texts of the stories appear in the Appendix.) Each story pair contained a high version with terms referring to high spatial position, and a low version with terms referring to low spatial position. Other than this semantic contrast in high or low terminology, the paired stories were identical. The stories followed a general format of containing a short introduction, a plot, and a conclusion, with each section presented slide-by-slide in a PowerPoint presentation. In addition, five filler stories were created, which were similar in structure, but did not focus on vertical space in any way.

The stories were combined together into four counterbalanced lists of four target stories interspersed among the five fillers. The valences of verticality were counterbalanced across the pitch and emotion stories. Participants who read a story about a dog with a high-pitched bark also read a story about a person feeling emotionally low, while those who read about a dog with a low-pitched bark also read about a person feeling emotionally high. The two physical motion stories were also counterbalanced, so that participants who read a story about riding an elevator to the top floor of a building also read a story about bending down for an object on the bottom shelf, and those who read about riding an elevator down to the basement also read about reaching up for an object on the top shelf. The order of story presentation was also counterbalanced across participants.

3.3 Procedure

Participants read one of the four lists of stories from a PowerPoint presentations displayed on a laptop computer. In order to make the task more communicative and meaningful, participants were assigned to read the stories to an undergraduate confederate, who played the role of a listener who would afterwards have to answer questions about the stories. Participants were

instructed to read clearly and 'be an interesting storyteller' so that their partner would be able to understand and remember what was read to them. The participant first read through each story once silently to ensure familiarity with it, and then read it aloud to the confederate. The readings were recorded at 44.2 kHz using a digital audio recorder with a flat boundary microphone sitting unobtrusively next to the computer. Finally, after reading all nine stories, participants filled out an exit questionnaire. In total, a typical session lasted 15-20 minutes.

3.4.1 Analysis of Contrasting and Shared Phrases

The digital recordings were analyzed in Praat (Boersma 2001). For each story, the boundaries of contrasting (the high or low terms contrasted within the story pairs) and shared phrases (the identical portions within the story pairs) were marked in a Praat TextGrid. Disfluencies, such as false starts and repetitions, were marked and later excised from the analysis. The recorded readings and associated TextGrid annotations were fed into a Praat script that computed pitches of each marked interval. Due to the difficulties of accurately measuring F0, each participant's recording was analyzed twice using the autocorrelation pitch tracking algorithm in Praat. The first pass was run with default settings on the raw recordings from each participant. This provided a baseline pitch for each participant, based on their average pitch across the whole task, which was used in the second pass of the pitch tracker to refine the settings for more accurate tracking.

For the second pass, several adjustments were made to the recording and to the default settings of Praat's pitch tracking algorithm. First, each participant's recording was low-pass filtered, with the upper bound of the pass band set at eight semitones (2/3 octave) above the speaker's average pitch across the whole recording. This filter guarantees an absence of doubling for the upper octave of the speakers' constrained vocal range, starting at four semitones below their baseline pitch. The filter also removes formants, reducing the potential influence of supralaryngeal articulation on tracked F0. Further, to ignore amodal (breathy, creaky) voicing and reduce pitch halving, the Silence Threshold of Praat's pitch tracking algorithm was set to 0.075 (from a default of 0.03) and the floor of the pitch tracker window was set to eight semitones below the speaker's previously determined baseline (Boersma 1993).

Pitches for the contrasting intervals of each story were computed as the duration-weighted average pitch of all contrasting phrases; pitch for the shared elements was a duration-weighted average of the pitches of the remaining, shared intervals. In addition, the two stories about physical space both contained the phrase 'all the way,' which allowed for a within-speaker

Acoustic Contour Analysis of phonologically identical material across opposite semantic contexts.

3.4.2 Acoustic Contour Analysis of 'all the way'

The phrase 'all the way,' which occurred in both physical space stories, is ideal for a within-speaker Acoustic Contour Analysis because it is composed entirely of voiced continuants, which means that there is an F0 signal throughout the phrase. This phrase thus afforded a second, finer-grained analysis. Here, rather than simply comparing the average pitch across participants' two productions of this identical phrase, we compared the trajectories of speakers' pitches over time. This approach has the advantage of being able to detect differences in the shape of a pitch contour that do not necessarily change its average pitch, as illustrated in Figure 1.

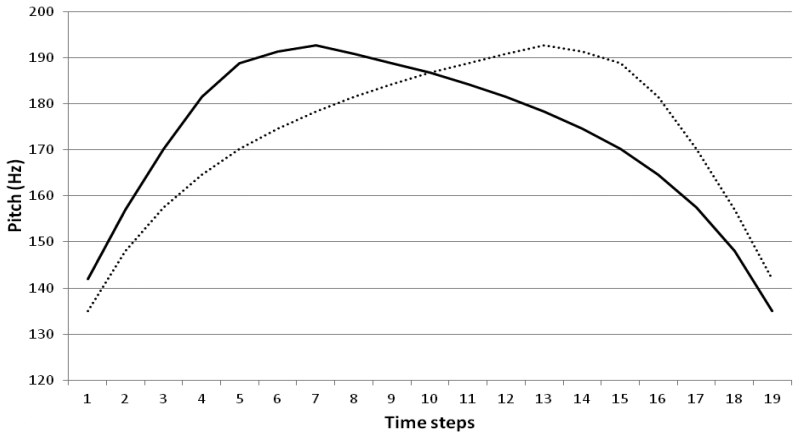

FIGURE 1. Contrasting pitch contours with same mean

The analysis used a Praat script to slice each phrase into twenty intervals and computed the average pitch in each ventile. Then, within-speaker comparisons were conducted for each of the twenty time steps. Following the mousetracking paradigm established by Dale, Kehoe, and Spivey (2007), we then conducted a bootstrap simulation to assess the significance of the largest region of consecutive significant differences. If our contours show a region of consecutive significant differences larger than could be expected by chance, we can conclude that the difference is due to the semantic context in which the identical phonological material was uttered.

4 Results

An overview of our results is presented in Table 1. To reduce the influence of individual differences in the emotion and reported sound stories (for which the 'up' vs 'down' contrast was between speakers), participants' pitch measurements were normalized using their average pitch in the filler stories as a baseline.

		Shared Phrases		Contrasting Phrases	
		Up	Down	Up	Down
Emotion	*M*	178.3	178.2	160.1	161.0
	s	7.1	5.4	8.2	8.0
Reported Sound	*M*	173.0	171.0	160.3	156.9
	s	5.7	6.8	6.6	7.9
Physical Space	*M*	176.4	171.6	166.7	160.7
	s	5.5	5.3	6.5	5.4

Note: For emotion and pitch stories, the 'up' vs 'down' contrast is between-speakers; for physical space stories, the contrast is within-speakers. All statistics are corrected for intra-speaker differences in baseline pitch.

TABLE 1. Descriptive statistics of pitch (Hz) produced in stories

4.1 Contrasting and Shared Phrases

For the emotion stories, a mixed two-way ANOVA was conducted on participants' pitch measurements, with contrastiveness ('contrasting' vs 'shared' phrases) as a within-speaker variable, and semantic valence ('up' vs 'down') as a between-speaker variable. This analysis revealed only a main effect for contrastiveness, and no main or interaction effect involving semantic valence.

For the reported sound stories, the same two-way ANOVA was conducted on participants' pitch measurements, with contrastiveness ('contrasting' vs 'shared' phrases) as a within-speaker variable, and semantic valence ('up' vs 'down') as a between-speaker variable. This analysis re-

vealed also only a main effect for contrastiveness, and no main or interaction effect involving semantic valence.

For the physical space stories, a two-way within-speaker ANOVA targeted participants' pitch measurements with contrastiveness ('contrasting' vs 'shared' phrases) and semantic valence ('up' vs 'down' stories). This analysis revealed that the overall model captured a significant portion of the variance in average pitch scores. There were significant main effects for both contrastiveness and semantic valence ($F(1, 30) = 11.43$, $p = .002$), but the interaction between the two did not reach significance. For the main effect of semantic valence, phrases in stories in the 'up' condition were spoken at a higher pitch (M = 171.5 Hz, s = 4.6 Hz) than those in the 'down' condition (M = 166.2 Hz, s = 5.3 Hz).

4. 2 Pitch Contour of 'all the way' in Physical Space Stories

In the analysis of the phrase 'all the way' as produced in the context of 'up' or 'down' stories, a paired t-test of the two average pitch across the two conditions revealed a marginal difference, $t(31) = 1.97$, $p = .058$, with productions in the 'up' condition (M = 186.1 Hz) averaging about 6.6 Hz higher pitch than those in the 'down' condition (M = 179.5 Hz). However, as previously noted, comparing averages across the whole phrase does not take into account the shape of the contours resulting in these averages (recall Figure 1). We also examined the pitch contour of the phrase, and found that, from the 11th through the 15th ventiles of the duration of this phrase (a region corresponding to the third quartile of the phrase), the average pitch of the phrase when produced in the 'up' condition was significantly higher than the 'down' condition (All $t > 2.25$, all $p < .033$). See Figure 2 for an illustration of this region.

Next, in order to assess the statistical significance of a region of difference totaling 25% of the duration of the phrase, we conducted a bootstrap simulation based on the observed contour data. In each iteration of the simulation, we created a simulated dataset by randomly sampling (with replacement) 32 pairs of productions of the phrase. Then, for each phrase, we randomly reassigned one of the productions to the 'up' condition and the other to the 'down' condition. We then counted the number of consecutive significantly different timesteps in the resampled dataset. For each iteration, this count of consecutive significantly different steps was entered into our Monte Carlo distribution. We ran a simulation of 10,000 iterations, and found that the 95th percentile (corresponding to $\alpha = .05$) of the Monte Carlo distribution was 2 consecutive significant differences, the 99th percentile (corresponding to $\alpha = .01$) of the distribution was 4 consecutive significant differences, and the 99.9th percentile (corresponding to $\alpha = .001$) was 7 consecutive significant differences. Thus, we can conclude that our five-ventile

region of difference is significant at $\alpha = .01$. As shown in Figure 2, in both conditions, the shape of the contour was generally a downward-facing parabola, with its highest point in the 3rd quartile of the phrase. This analysis revealed that region around the peak pitch, from the 11th through the 15th ventiles of the phrase's duration, was spoken at a significantly higher pitch in the 'up' condition.

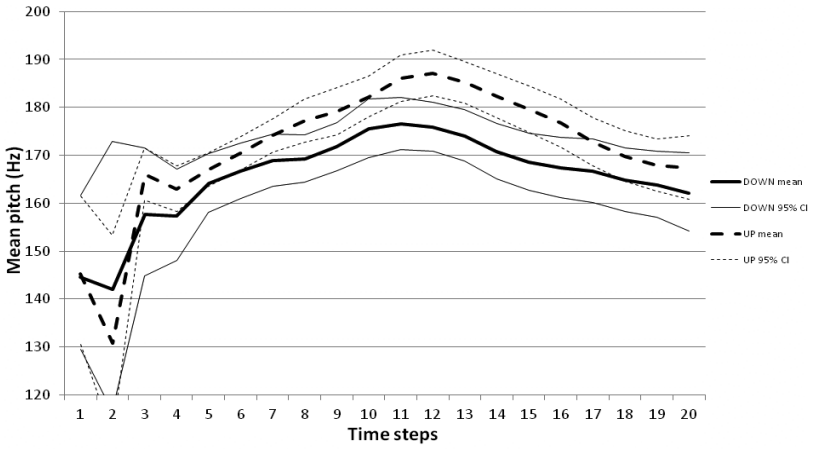

FIGURE 2. Pitch contours of 'all the way' in UP vs. DOWN context

5 Discussion

Previous research has documented the close conceptual relationship between vertical space and auditory pitch, particularly in Western, English-speaking cultures. Yet little is known about the intuitive possibility that people use pitch, embodied through movements and tensions of the vocal tract, to understand vertical space. Based on previous work on vocal iconic gesture and related phenomena, we examined this idea by testing whether speakers produced pitch-related vocal gestures as they read aloud stories related to high and low vertical space, high and low emotion, and high and low pitch. Our prediction was that readers would produce higher pitches when reading 'high' events and lower pitches when reading 'low' events.

The main analysis investigated whether there was a difference in average pitch between the up and down versions of stories, and also whether differences in pitch might be more localized to directly contrastive portions

of the story compared to the identical, noncontrastive portions. Interestingly, we did not find evidence that people made iconic modulations of pitch when reading the stories related to high and low emotion or to the high or low pitch of a dog's bark, despite the less abstract pitch-meaning mapping of these stories. In these cases, the iconic use of pitch to express positive and negative emotions is acknowledged to be a basic function of prosody (Bolinger 1986), and the use of pitch to convey the high or low pitch of a sound involves a directly imitative mapping. We caution, however, that these exploratory null results are based the average pitches across just a couple idiosyncratic stories, and do not license generalization. Future research must use a broader range of stimuli to further examine the role of pitch in conceptualizing these and other polysemous senses of vertical terms, and how these meanings are expressed through vocal gesture.

In contrast, we did find a significant main effect for semantic valence in the two stories about high and low vertical space. In these stories, there was no interaction between semantic valence and contrastiveness, suggesting that the pitch difference was not just localized to contrastive sections, but more spread out across the story. These results support previous research on the close conceptual connection between pitch and vertical space, and specifically demonstrate how pitch modulations manifest as conceptually motivated vocal gestures, produced as people talk about vertical space.

While the main analysis compared average pitches across relatively large segments of speech, it is elucidating to examine pitch with a more fine-grained temporal resolution. Towards this goal, we examined the pitch contours of the phrase 'all the way', which appeared in both semantic conditions of the two vertical motion stories. Our novel Acoustic Contour Analysis, derived from mousetracking studies, compared the pitch contours of speakers' productions of the identical phonological content across the high and low conditions. Over both conditions, the shapes of the pitch contours were generally similar. However, as can be seen in Figure 2, around the peak pitch of the phrase, the pitch contour in the high condition separates upward from the pitch of the low condition. Statistical analysis confirms that the region around the peak pitch, from the 11^{th} through the 15^{th} ventiles of the phrase's duration, was spoken at a significantly higher pitch in the 'up' condition.

In addition to the theoretical value of demonstrating a vocal iconic gesture in which conceptualized vertical space is expressed through acoustic pitch, we stress the value of this Acoustic Contour Analysis as a method for comparing the shapes of pitch contours, and thereby examining co-speech iconic gestures in the vocal modality. Through this analysis, we were able to describe a pattern of speech production that would not otherwise have been detected by pitch averages measured over whole phrases. Since lan-

guage production unfolds through time, it is crucial to investigate pitch as a dynamic variable, rising and falling over the time course of an utterance. This treatment of pitch as a dynamic analog signal underscores the similarity between iconic gestures produced by the vocal tract, and their more well-studied counterparts produced by other, more visible bodily effectors. Acoustic Contour Analysis affords empirical investigation of vocal gestures on the same time scale as the preparations, strokes, and holds of manual gestures. This possibility presents a marked contrast to Hockett's (1978: 274) claim that human language is characterized by an arbitrary relationship between forms and meanings, because iconicity is squeezed out when the four dimensions of reality are reduced into the single dimension language. As our figures make clear, pitch (as well as other acoustic variables like intensity) can be rendered visible in two dimensions, serving as proxies for the obscured movements of the vocal tract over time. Thus, spoken language is not in fact unidimensional, but rather affords substantial iconicity in the dynamics of its many acoustic variables changing over time.

There are, of course, differences between vocal iconic gestures and manual iconic gestures. Whereas manual iconic gestures communicate through the visual modality, independent of the speech channel, vocal iconic gestures are typically superimposed on, and thus constrained by, the articulatory business of producing speech. As a consequence, they can be challenging to identify and study, similar to the challenge posed by the co-sign gestures that are produced by users of signed languages (Duncan 2003; Okrent 2002). Our Acoustic Contour Analysis overcomes this challenge by focusing on phonologically identical phrases in different semantic contexts. When the conventional, linguistic content is identical, but analog, nonconventionalized aspects of the signal differ in ways that are predictable from the semantic context, one can conclude that such differences result from vocal iconic gestures.

Previously, researchers interested in vocal iconic gestures have stressed a modality independent notion of iconic gesture to account for their production across the manual/visual and vocal/auditory modalities (Okrent 2002). As McNeill (1992) observes, vocal and bodily iconic gestures and language (both spoken and signed) seem to behave as deeply interlinked parts of the human conceptualization and communication system. However, the affordances of each modality, and the constraints imposed by the need to coordinate linguistic communication, mold characteristic styles of gesture in each modality. For example, the pitch gestures studied here are based on highly schematic correspondences between pitch and verticality, rather than depicting a more detailed image of the scene as with some manual gestures. On the other hand, vocal iconicity is highly suitable to more detailed depic-

tions of auditory images, such as during the quotation of sounds (Blackwell, Perlman, and Fox Tree [under review]), which is difficult to accomplish through manual gesture. But while the pitch-verticality mapping may abstract and schematic, it is also robust enough to generate the vocal gestures we document, underscoring the prominent role both domains play in everyday human experience.

6 Conclusion

Historically speaking, vocal iconicity has typically been viewed as a marginal aspect of language. The received view of Saussurian arbitrariness between form and meaning, though, has been eroding for some time. For example, Perniss, Thompson, and Vigliocco (2010) have recently argued that motivated sound-meaning relationships are in fact a common feature of language, both spoken and signed. In spoken languages, evidence for this claim includes phonological phenomena like sound symbolism and phonaesthemes, as well as lexical phenomena like onomatopoeia. Additionally, the present work contributes to a growing body of research that focuses on iconic uses of prosody, demonstrating various ways that prosody can express semantic meaning in an utterance (Nygaard et al. 2009; Perlman 2010; Perlman et al. [under review]; Shintel et al. 2006; Shintel and Nusbaum 2007; 2008). This avenue of research is well-suited to the language of gesture studies, and specifically addresses the question of how vocal iconic gestures may be superposed on the articulation of speech.

The present study advances research on vocal iconic gesture in two ways. First, our results expand upon previous findings (Nygaard et al. 2009; Shintel et al. 2006), demonstrating that contrasts of ‘up’ vs ‘down’ are encoded in speakers’ prosody. These results contribute more broadly to a large body of work showing the tight conceptual relationship between vertical space and pitch. Second, our Acoustic Contour Analysis methodology offers researchers a powerful new analytical tool to study the temporal unfolding of vocal gestures. This method affords two-dimensional visualization and comparison of any prosodic variable, making vocal gestures substantially more amenable to the fine-grained description and temporal analysis that has been so productive for studies of manual gesture.

7 Appendix

7.1 Physical Space Stories

Climbing Up High

Melinda is at the store. She is staring up at a box on the very highest shelf. It is high up, and she must climb to get it. Melinda climbs all the way up to the top. She gets the box.

Bending Down Low

Melinda is at the store. She is staring down at a box on the very lowest shelf. It is low down, and she must squat to get it. Melinda bends all the way down to the floor. She gets the box.

Going Up a Skyscraper

Chris is in an elevator with a view. It is going up a skyscraper. He watches out the window as he rises up and up, all the way to the top of the building. He reaches the end of the ride.

Going Down a Skyscraper

Chris is in an elevator with a view. It is going down a skyscraper. He watches out the window as he descends down and down, all the way to the bottom of the building. He reaches the end of the ride.

7.2 Emotion Stories

High on Life

Janet is walking to class. She is in high spirits. She looks at the trees surrounding her. They are uplifting. She feels like she is rising into the sky. Janet is high on life. She reaches school.

Low on Life

Janet is walking to class. She is in low spirits. She looks at the trees surrounding her. They are depressing. She feels like she is sinking into the ground. Janet is low on life. She reaches school.

7.3 Reported Sound Stories

A Deep Bark

Max is a dog with a low-pitched bark. He just spotted the mailman out the window. He runs to the door and barks his deep bark. Ruff! The mailman is not frightened.

A High Bark

Max is a dog with a high-pitched bark. He just spotted the mailman out the window. He runs to the door and barks his high bark. Ruff! The mailman is not frightened.

Acknowledgement

The study was initiated while the third author, Marlene Johansson Falck, was a postdoctoral fellow in the Department of Psychology, University of California, Santa Cruz, and funded by the Swedish Research Council. Her contribution to writing this manuscript was funded by the Royal Swedish Academy of Letters, History and Antiquities, supported by a grant from the Knut and Alice Wallenberg Foundation (KAW 2009.0295).

References

Ashley, R. 2004. Musical Pitch Space across Modalities: Spatial and Other Mappings through Language and Culture. In *Proceedings of the 8th International Conference on Music Perception and Cognition*. Eds. S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster, 64-71. Adelaide: Causal Productions.

Blackwell, N., M. Perlman, & J. Fox Tree. [Under Review]. Quotation as a Multimodal Construction. (Manuscript under review.)

Bolinger, D. 1986. *Intonation and its Parts*. Stanford, CA: Stanford University Press.

Ben-Artzi, E. & L. Marks. 1995. Visual-Auditory Interaction in Speeded Classification: Role of Stimulus Difference. *Perception and Psychophysics* 57: 1151-1162.

Boersma, P. 1993. Accurate analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Institute for Phonetic Science, University of Amsterdam, Proceedings* 17: 97-110.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5: 341-345.

Casasanto, D. 2008. Conceptual Affiliates of Metaphorical Gestures. In *Proceedings of the International Conference on Language, Communication, & Cognition*. Brighton, UK.

Cienki, A. 1998. Metaphoric gestures and some of their relations to verbal metaphoric expressions. *Discourse and Cognition: Bridging the Gap,* ed. Jean-Pierre Koenig, 189-204. Stanford, CA: Center for the Study of Language & Information.

Cienki, A. & C. Müller. 2008. Metaphor, Gesture, and Thought. *Cambridge Handbook of Metaphor & Thought*, ed. R. Gibbs, 483-501. Cambridge: Cambridge University Press.

Cosmides, L. 1983. Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance* 9: 864-881.

Dale, R., C. Kehoe, & M. Spivey. 2007. Graded Motor Responses in the Time Course of Categorizing Atypical Exemplars. *Memory and Cognition* 35: 15-28.

Duncan, S. 2003. Gesture in Language: Issues for Sign Language Research. *Perspectives on Classifier Constructions in Signed Languages*, ed. K. Emmory, 259-268. Mahwah, NJ: Lawrence Erlbaum Associates.

Eitan, Z. & R. Timmers. 2010. Beethoven's Last Piano Sonata and Those Who Follow Crocodiles: Cross-domain Mappings of Auditory Pitch in a Musical Context. *Cognition* 114: 405-422.

Grady, J. 1997. *Foundations of Meaning: Primary Metaphors and Primary Scenes*. Ph.D. dissertation, University of California, Berkeley.

Hockett, C. 1978. In Search of Jove's Brow. *American Speech: A Quarterly of Linguistic Usage* 53(4): 243-313.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Lakoff, G. & M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G. & M. Johnson. 1999. Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought. New York: Basic Books

Marks, L. 1996. On Perceptual Metaphors. *Metaphor and Symbolic Activity* 11: 39-66.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: Chicago University Press.

McNeill, D. 2005. *Gesture and Thought*. Chicago: Chicago University Press.

Meier, B. & M. Robinson. 2004. Why the Sunny Side is Up: Associations between Affect and Vertical Position. *Psychological Science* 15: 243-247.

Melara, R. & T. O'Brien. 1987. Interaction between Synesthetically Corresponding Dimensions. *Journal of Experimental Psychology: General* 116: 323-336.

Nygaard, L., D. Herold, & L. Namy. 2009. The Semantics of Prosody: Acoustic and Perceptual Correlates to Word Meaning. *Cognitive Science* 33: 127-146.

Okrent, A. 2002. A modality-free notion of gesture and how it can help us with the morpheme vs gesture question in sign linguistics, or at least give us some criteria to work with. *Modality & Structure in Signed & Spoken Languages*, ed. R. Meier, D. Quinto, & K. Cormier, 175-198. Cambridge: Cambridge University Press.

Perlman, M. 2010. Talking Fast: The Use of Speech Rate as Iconic Gesture. *Meaning, Form, and Body,* eds. F. Perrill, V. Tobin, & M. Turner, 245-262. Stanford, CA: CSLI Publications.

Perlman, M., N. Clark, & M. Johansson Falck. [under review]. Conceptually motivated prosody in story reading. *Cognitive Science*. [manuscript under review].

Perniss, P., R. Thompson, & G. Vigliocco. 2010. Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology* 1: 1-14.

Pratt, C. 1930. The Spatial Character of High and Low Tones. Journal of Experimental Psychology, 13, 278–285.

Roffler, S. & R. Butler. 1968. Localization of Tonal Stimuli in the Vertical Plane. *The Journal of the Acoustical Society of America* 43: 1260–1266.

Rusconi, E., B. Kwan., B. Giordano, C. Umiltá, & B. Butterworth. Spatial Representation of Pitch Height: The SMARC Effect. *Cognition* 1: 1-17.

Shayan, S., O. Ozturk, & M. Sicoli. 2011. The Thickness of Pitch: Crossmodal Metaphors in Farsi, Turkish, and Zapotec. *Senses and Society* 10: 96-105.

Shintel, H., H. Nusbaum, & A. Okrent. 2006. Analog Acoustic Expression in Speech Communication. *Journal of Memory and Language, 5,* 167-177.

Shintel, H. & H. Nusbaum. 2007. The Sound of Motion in Spoken Language: Visual Information Conveyed by Acoustic Properties of Speech. *Cognition* 105: 681-690.

Shinttel, H. & H. Nusbaum. 2008. Moving to the Speed of Sound: Context Modulation of the Effect of Acoustic Properties of Speech. *Cognitive Science* 32: 1063-1074.

Trimble, O. 1934. Localization of Sound in the Anterior, Posterior, and Vertical Dimensions of Auditory Space. *British Journal of Psychology* 24: 320–334.

Wagner, Y., E. Winner, D. Cicchetti, & H. Gardner. 1981. 'Metaphorical' Mapping in Human Infants. *Child Development* 52: 728-731.

Zbikowski, L. 1998. Metaphor and Music Theory: Reflections from Cognitive Science. *Music Theory Online* 4.