# Physical correlates of prosodic structure in American Sign Language

Ronnie B. Wilbur & Aleix M. Martínez
Purdue University

## 1 Introduction

The study of ASL prosody provides insights into the specification of stress, rhythmic structure, and intonation in conjunction with the syntax of a natural signed language. It allows us to test claims about the language -and modality-independence of various prosodic phenomena. Intonation in speech, for example, is dependent on pitch, which is not available in the manual modality. Stress marking in speech depends on pitch, duration, and amplitude. Rhythmic phrasing in speech, by which a beat pattern is established and varied for purposes of linguistic grouping, involves duration and pausing. By comparison of the intonation and rhythm mechanisms in ASL with those in spoken languages, we are able to separate the *functions* of intonation and rhythm in language from their *expression* in different languages and modalities. There is evidence that over time, ASL has accommodated to the production and perception requirements of the manual/visual modality, resulting in a prosodic system that is comparable in function to spoken languages but different in means of expression.

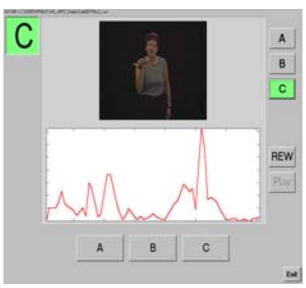
Detailed investigation of perception and production of spoken languages has provided a reasonably clear picture of the relative roles of duration, frequency, amplitude, tempo, and pausing in prosodic structure. In contrast, the prosodic structure of sign languages has not attracted much attention even though it represents a fundamental investigation if our aim is to fully understand how language works. One problem is the difference in physics/kinematics. The variables available for use in signing are displacement (how far the articulator travels), duration, velocity (how much displacement in a unit of time), acceleration (changes in velocity per unit time), and jerk (changes in acceleration per unit time). Existing speech research methods and equipment are not designed to handle these variables. The technological problems are also formidable, comparable to research on speech before the invention of tape recording. However, the linguistic problems are even more daunting, comparable to research on speech before the development of alphabets, not just the IPA. Thus, we begin with old-fashioned 'primitive' measures, namely signers' judgments of input stimuli.

## 2 Goals and assumptions

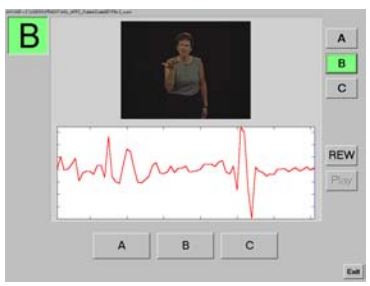
The primarily goal of this paper is to tease apart factors that contribute to the encoding of prosodic information in the signing signal. As a first step, we investigate the role that velocity, acceleration and jerk (which is the derivative of the acceleration over time) of an ASL sentence may play in coding its prosody. This is a difficult task, because prosodic information might be encoded in many distinct forms. Our first and main assumption is that some of the prosodic information is encoded in a kinematic trace of a sentence. We believe this is a reasonable assumption, since previous work has shown that signers are able to tap to the rhythm of input signing in a reliable and coherent manner (Allen, Wilbur & Schick 1991).

Our second assumption is that the most relevant kinematic information of a sign can be extracted from the motion associated to the arms (and hands) of a signer. Note that, even if the arm does not move with respect to the body but the body does, then (in absolute values) the arm is moving. Therefore, our second assumption is not that most of the information is encoded on the arms and hands, but that this can be extracted from them.

When the hands (and equivalently the arms) move on the 3D space of a signer, it creates a displacement pattern that can be captured. This displacement over time is what we know as *velocity*, and can be formally expressed as $v=d/t$ (where $d$ is the displacement and $t$ is the time). We can now visualize the velocity as a function of displacement and time in a 2D plot as shown in (1). We will refer to this plot as the velocity trace.



(1) Velocity



(2) Acceleration

(3) Jerk

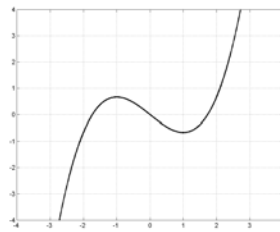
A possible hypothesis is that the prosody of an ASL sentence is encoded in the velocity trace of the sign, since this is the simplest interpretation of the motion of the hands and arms of the signer. This hypothesis has several appealing advantages. The most obvious one is that it is the simplest feature to analyze. Simple implies fast interpretations. It also has the advantage of only representing displacement changes, whereas acceleration and jerk carry other types of information, which makes the plot visually more complex. [1]

---

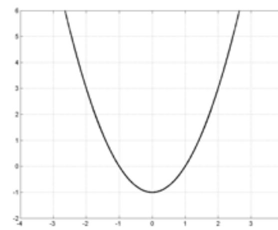
1. In general, the derivative of a function $f$ will correspond to a function $f'$ of lower order. For example, the $f=x^3/3-x$ shown in (n1.1), has a derivative $f'=x^2-1$, shown in (n1.2), of smaller order.

Velocity though has some disadvantages too. Since the plots tend to be very simple (only showing the changes in displacement), these might only carry a limited quantity of information. Obviously, the velocity plot can encode many prosodic cues, but it would be difficult to extract many of them from a simple plot such as the velocity trace. The information might be hidden in areas of small (unclear) displacement changes.

---

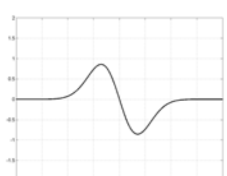


(n1.1)

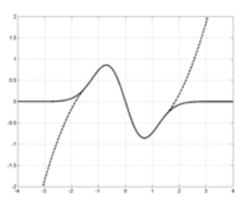


(n1.2)

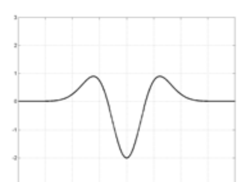
This seems to be in contradiction with our statement that acceleration is generally a more complex function (in the sense of having more peaks) than the velocity. Note however that this is generally the case for pairwise functions (where the value of a function depends on the interval) or functions with continuous changes. For example, the function $f = -2x \cdot e^{-x^2}$, shown in (n1.3) can be seen as composed of two parts; where one of its parts corresponding to the function shown in (n1.1) – see (n1.4) for a comparison of the two functions. We can see now, how for the functions of this type, it is usually the case to have more "peaky" derivatives. The derivatives of the above function is represented in (n1.5)



(n1.3)

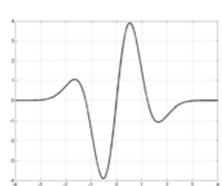


(n1.4)



(n1.5)

Again the derivative of the function shown in (n1.6) will correspond to a function with more peaks, as shown in (n1.6).



(n1.6)

Note that our velocity traces are (in general) of the type shown above and, in consequence, we expect its derivative to be generally more complex (with more peaks) than the original function. The same rational applies to the acceleration and jerk.

In comparison, other plots such as the acceleration can more easily encode all the prosodic features. We argue that this is so, because additional derivatives of the displacement will emphasize more and more changes of the motion pattern (by means of peaks on the plot). For example, acceleration, which is the derivative of the velocity over time (i.e. a=v/t), will have information about the displacement changes (i.e. velocity) as well as information about the velocity changes. This is made clear in the plot (2) above. This graphical representation of the acceleration corresponds to the same sentences as the plot shown in (1). It is thus reasonable to hypothesize that some of the prosodic information in ASL is encoded in the acceleration trace. This would allow the encoding of additional information.

However, we would expect sign naïve (hearing) subjects to prefer velocity even if the prosodic information is better represented by the acceleration or jerk traces. This is so because the velocity is the one that best describes the pattern of motion of the arms and hands. Similar to the way that we can interpret (and enjoy) the pattern of motion of a ballet dancer, we should be able to visually extract the velocity trace of a signer (regardless of whether this is related to prosodic information or not). This is an important point with implications for the results we expect from the experiments reported below. If velocity is so easy and *natural* to obtain, why should sign languages have needed to develop greater level of abstraction such as coding prosodic information in acceleration and jerk? If the analysis of velocity is universally available to human infants, why should sign languages require more complex processing? One possible reason is that: *velocity is not flexible enough to encode all the prosodic parameters of an ASL sentence*. Thus, further derivatives are needed to add more flexibility to the encoding and analysis of the prosodic features.

The rational of the previous paragraph would also apply (probably even more so) to the derivative of the acceleration over time, i.e. the jerk (3). For this reason, we experiment with both the acceleration and the jerk as alternatives to the velocity trace.

A final hypothesis is that prosody is encoded in two or even all three traces described above. This is not the most obvious hypothesis and indeed the most difficult to check, but it is a possible one. We will return to this hypothesis.

## 3 Background preliminary studies

In *speech*, both duration and fundamental frequency are involved in the marking of phrasally prominent (stressed) syllables. There is strong evidence that such syllables are longer and have increased fundamental frequency compared to syllables that are not phrasally prominent (Fry 1955, 1958; Bolinger 1985; Ladefoged 1982). Syllables in phrase final position are also marked by increased duration (Phrase Final Lengthening) and with changes in fundamental frequency that have been interpreted as intonational boundary tones (Beckman & Pierrehumbert 1986). The available physical characteristics of sign movement are duration, displacement (cm), and velocity (cm/s). The question of interest is how *signed* languages mark linguistic prominence and phrase position, given that frequency is not an available marker (and that phrasal prominence is generally preferred in final position; Wilbur & Schick 1987). The answer to this question affects our reasoning regarding the kinematic traces that might be used for carrying larger prosodic structure.

### 3.1 Functions of duration

Liddell (1978) investigated factors affecting sign duration and reported that utterance final signs are marked by increased duration, as are initial signs, although to a lesser degree. In addition to phrase position, the function of a sign, namely whether or not it represents a topic, also affects the duration, with topic signs (which are initial) longer than non-topics. Thus, duration is used as a marker for phrase position. Similar results are reported by Wilbur and Nolen (1986) for syllable measurements in ASL.

Wilbur and Zelaznik (1997; Wilbur 1999) conducted a study of phrasal prominence and position marking in ASL using a 3-d motion analyzer system (WATSMART). Infrared diodes on the thumb and index finger of the dominant hand of 13 ASL users transmit their location to cameras as signers produce target signs in carrier phrases in the 4 relevant contexts. The duration results (Table 1) support previous sign research, in that duration is found to be primarily affected by phrase position, with final signs (205 ms) significantly longer than nonfinal signs (175 ms).

Table 1: Duration (ms) as a function of phrase position and stress

| | Non-final | Final | Mean |
|---|---|---|---|
| Unstressed | 173.8 (53.1) | 201.9 (57.4) | 187.8 ms (57.1) |
| Stressed | 177.6 (44.4) | 207.9 (59.98) | 192.8 ms (54.9) |
| Mean | 175.2 ms (48.96) | 204.9 ms (58.7) | |

### 3.2 Functions of velocity

Wilbur and Zelaznik (1997) reported that increased peak velocity marked phrasal prominence, with prominent signs reaching higher peak velocity than unstressed signs (Table 2). Peak velocity is also slightly affected by phrasal position, with final signs achieving a higher peak velocity than nonfinal signs. However, signers do not have to make a 4-way distinction in peak velocity, as duration is significantly affected only by position and therefore serves as a clear and unambiguous marker of phrase position.

Table 2: Peak velocity (cm/s) as a function of phrasal position and stress

| | Non-final | Final | Mean |
|---|---|---|---|
| Unstressed | 293.8 (108.5) | 323.7 (105.7) | 308.5 (107.95) |
| Stressed | 408.6 (118.1) | 424.7 (117.4) | 416.7 (117.8) |
| Mean | 351.0 (126.9) | 375.2 (122.6) | |

### 3.3 Functions of displacement

The third kinematic characteristic available to signed languages for manipulation is displacement, how far the hand travels during the production of a given sign. Many signs have displacements that are restricted by the anatomy of the hand: when the sign UNDERSTAND is made, the index finger opens from a closed fist, and the displacement is thus directly determined by the length of the signer's index finger and hence not linguistically manipulable. Other signs involve movement from the elbow joint, where greater number of degrees of freedom in movement provides for more flexibility in displacement manipulation. In Wilbur and Zelaznik (1997), the target signs all had downward vertical movement primarily from the elbow joint (CAN'T, MUST, CAN), thus simplifying the measurement of displacement while still allowing the possibility that changes in displacement could be used to signal changes in phrase position or prominence. In fact, displacement is affected by both position and prominence (Table 3). Final signs are larger, that is, travel more distance than nonfinal signs, and stressed signs are larger than unstressed signs. The position by stress interaction is also significant, revealing a complex picture. Nonfinal unstressed signs travel the smallest distance; nonfinal stressed and final unstressed signs travel about the same distance; and final stressed signs travel the largest distance.

Table 3: Displacement (cm) as a function of phrase position and stress

| | Non-final | Final | Mean |
|---|---|---|---|
| Unstressed | 20.6 (9.9) | 29.3 (11.4) | 24.9 (11.5) |
| Stressed | 31.3 (10.1) | 35.7 (11.3) | 33.5 (10.9) |
| Mean | 25.9 (11.3) | 32.6 (11.8) | |

### 3.4 Summary of prior studies

What do these results tell us? These results support the conclusion that Phrase Final Lengthening increases duration, prominence increases peak velocity, and displacement varies as needed. Signers never have to make or perceive more than 2 distinctions in either duration alone or peak velocity alone, and do not have to make distinctions in displacement for either purpose. Like speech, ASL uses duration for Phrase Final Lengthening, but lacking fundamental frequency modification for stress, substitutes peak velocity instead. At the kinematic level, we see a modality effect in marking of linguistic prominence but no modality effect with respect to phrase position: in both speech and sign, phrase final gestures are longer than nonfinal ones. In sum, (1) Phrase Final Lengthening increases duration; (2) stress increases velocity; and (3) when they are in ‘conflict’, displacement is increased.

From these findings, we see that velocity is used for prosodic purposes, making it a viable kinematic variable for carrying information about larger prosodic expanses, such as phrases and clauses. But it is also possible that velocity is already carrying as much information as it can, and that for larger domains, changes in velocity (acceleration) or changes in acceleration (jerk) may be enlisted to avoid overburdening the signer's perceptual capacity with respect to making distinctions within the velocity trace. Thus, in this study, we compare velocity with acceleration and jerk to see which, if any, of these signers rely on for prosodic information.

## 4 Methods

### 4.1 Judges

There were two groups of judges. The Deaf group included 9 native ASL signers, who had attended a state Deaf School. The Hearing group consisted of 9 college-aged sign naïve native speakers of English.

### 4.2 Stimuli

This study used a specially designed signing database collected from 14 signers with stimuli of varying length. This database includes ten ASL multi-sentence sequences with different types of prosodic information. All videos are in digital form and are stored in uncompressed AVI files to guarantee maximum image quality. Note that compressing the image could include artifacts that might distract the judges and might ultimately affect the results reported below.

All videos were captured with the same equipment and under strictly controlled illumination conditions; for this study, the videos used were recorded with diffused illumination. This type of illumination is specially designed to minimize the shadows in the image. Shadows should not be present in the videos because they could visually distract the judges from tracking the arms and hands. The entire database is described in Martinez et al. (2002) and can be obtained for research purposes from the authors. For additional information see http://rvl1.ecn.purdue.edu/~aleix/ASLdatabase.htm

Five multiple-sentence 'stories' were selected from this database as stimuli and 3 different signers performing each of the 5 were used, yielding 15 stimuli. These stimuli were then cut into smaller, more manageable videos of approximately 1-2 sentences length. Graphic plots of each kinematic feature (velocity, acceleration, and jerk) over time were then generated from these stimuli. Each stimulus was shown with each of its 3 kinematic displays, which were randomly assigned to 'A', 'B', or 'C' on each trial; the screen appeared as in (1-3) above. As the videoclip played, a vertical line cursor moved along the graphic display in coordination with the signing itself. Judges had control buttons to change kinematic displays (upper right), and to rewind and replay the clip as many times as they wished (center right). At the bottom of the screen are the three response buttons, from which they selected 'A', 'B', or 'C'.

### 4.3 Task

Video clips were randomized and divided into 4 blocks of stimuli, with breaks between them. The total time per session was under 40 minutes. Each judge started with a different block. Judges were asked to view each videoclip and its corresponding kinematic traces and to pick which graph they felt best exemplified the 'feel' of the signed material.

### 4.4 Data analysis

Responses were automatically recorded by the experiment program as 'v', 'a', or 'j'. These were transferred to an Excel spreadsheets and fed into SPSS for statistical analysis. Each judge's responses were calculated as sums of 'v' response, 'a' response, and 'j' response, which were then converted to percents. Bar graphs were generated within Excel. Within SPSS, chi-square testing was conducted to see if the pattern of 'v', 'a', 'j' responses for each judge varied from chance (33.3%).

## 5 Results

We have divided the analysis of the results into two steps. First, we discuss the results obtained with the native ASL signers. Our goal is to see if any of the hypotheses made above (see Section 2) is adequate. Second, we analyze the results obtained with the hearing judges with no previous experience in ASL. This should help us conclude if the results obtained with the native ASL signers are significant or not. Note that it would be possible for both groups (native ASL signers and hearing) to output the same results. This could indicate that the assumption that prosody is encoded in the kinematic traces of the arms and hands is not well founded. As we will see below though, this is not the case. Hearing responses are almost always at random, whereas native ASL responses can be balanced toward velocity and acceleration. However, we will also show that native ASL signers do not always prefer the same kinematic plot. We have found that this depends on: who is signing the stimuli (each of the signers shown in the video move slightly differently to indicate the same meaning), and what is being signed (different sentences might need different kinematic traces).

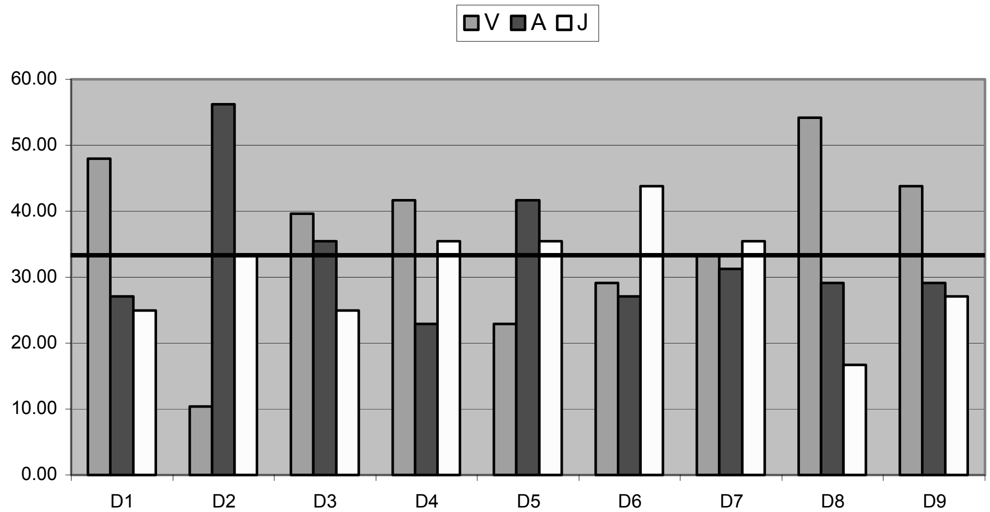
### 5.1. Native ASL signers (Deaf)

In (4) we show the statistical analysis corresponding to the native ASL signers (Deaf). Note that even with a confidence of 10% (0.1), only three of the judges are significantly different from chance.

| | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 |
|---|---|---|---|---|---|---|---|---|---|
| Chi-Square | 4.625 | 15.125 | 1.625 | 2.625 | 2.625 | 2.375 | .125 | 10.500 | 2.375 |
| Df | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Asymp. Sig. | .099 | .001 | .444 | .269 | .269 | .305 | .939 | .005 | .305 |

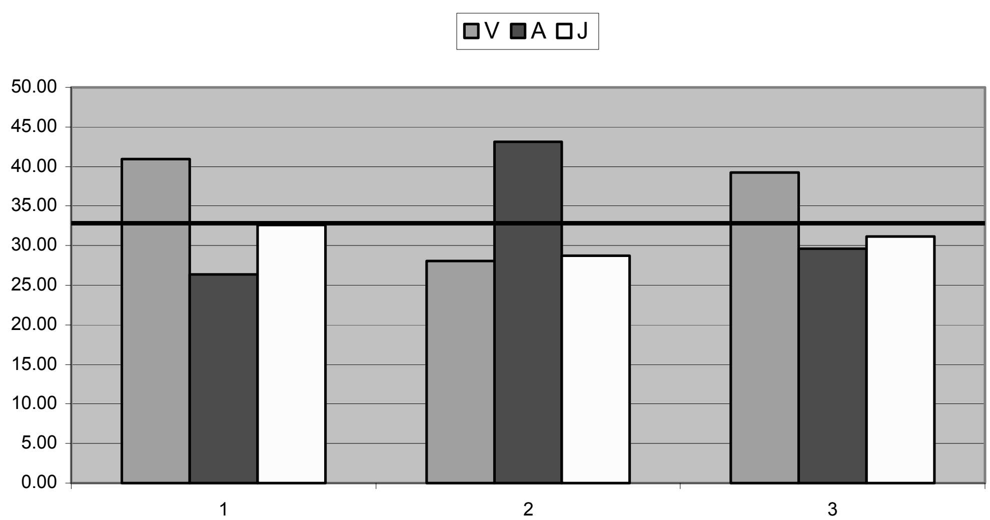
(4) Results of chi-square analysis for each Deaf judge

Note that even with a confidence of 10% (0.1), only three of the judges are significantly different from chance. As a result, we might find it useful to decompose the results by judges and choices. In (5) we show the percentage for each of the options (velocity, acceleration, and jerk) for each of the Deaf judges. The black line is the random selection threshold.




(5) Percent choices of each kinematic variable by each Deaf judge

Judges D1, D2 and D8 are clearly not at random, while the others are at random or close to random. Why is that so? Recall that the judges saw video clips corresponding to different ASL signers, and that each signer will have her/his particular way to sign. Therefore, it would be interesting to see if judges consistently selected one of the three choices (velocity, acceleration, jerk) for each of the signers used in the experiment. This is shown in (6), where we can see judges' preference for velocity for stimulus signers 1 and 3, and a preference for acceleration for stimulus signer 2. Since the graph in (6) shows preferences which depend on the signer, it seems to be the case that the results shown in (5) were obscured by variations in signing style among the stimulus shown to the judges.




(6) Percent judges' choices of kinematic variable by signer of the stimuli

As advanced in Section 2, it is reasonable to assume that the prosodic cues are distributed among the different kinematic traces. Another source of variation we have observed in our results is the dependency on the video clip. Some prosodic cues tend to be encoded in velocity, while others in acceleration depending on the prosodic cue used in that sentence.

Therefore, it will be necessary to further study when the prosodic information is encoded in velocity and when is encoded in acceleration. This can be signer dependent, stimuli dependent or the two.

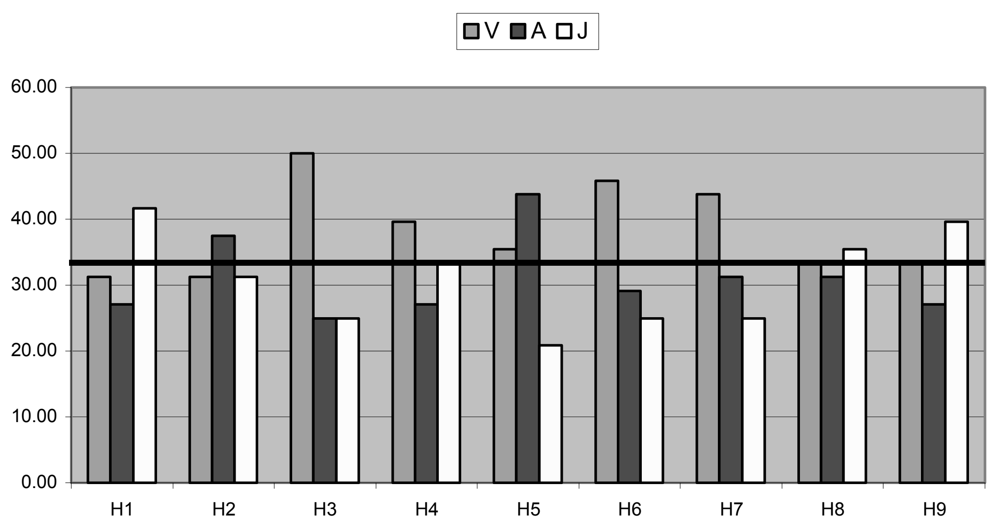
### 5.2. Sign naïve (Hearing)

We show in (7) the statistical analysis of the data of the hearing judges. Note that even with a confidence of 10% (.100), only one subject (i.e. H3) is below random.

| | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 | H9 |
|---|---|---|---|---|---|---|---|---|---|
| Chi-Square | 1.625 | .375 | 6.000 | 1.125 | 3.875 | 3.500 | 2.625 | .125 | 1.125 |
| Df | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Asymp. Sig. | .444 | .829 | .050 | .570 | .144 | .174 | .269 | .939 | .570 |

(7) Results of chi-square test for each hearing judge

Although the statistical analysis shown in (7) are convincing with regard to the claim that most of the responses of the hearing judges are at random, we might find it useful to decompose the results by judges and choices (as we did before with the native ASL). In (8) we show the percentage for each of the options (velocity, acceleration, and jerk) for each of the hearing judges. We note that the results are basically random except for judge Hear3. This judge preferred velocity over the others. As advanced in Section 2, since this is a sign-naïve individual, this performance might be due to the human capacity for extracting and analyzing displacement over time.




(8) Percent choices of kinematic variable for each hearing judge

## 6 Discussion and Conclusions

As described in Section 3, it seems counterintuitive to expect velocity to be the most relevant trace to represent prosody. This is because velocity is already used for several tasks in ASL, such as: aspectual modifications (Klima, Bellugi et al. 1979), intensity (Frishberg & Gough 1973), stress (Wilbur 2000), and rate of signing (Wilbur 2000).

It could also be the case though, that since velocity is already used to achieve several tasks, it will actually be more easy to encode (and extract) the prosodic information from it as opposed to using additional cues. The results reported above seem to indicate that indeed velocity is used to encode some of the prosodic cues of an ASL sentence, but that it is not the unique one. Our findings indicate that depending on the type of prosodic information we want to analyze, different kinematic traces will be used. We do not argue here that this is consistent over judges and stimuli signers. Instead, we argue that our results seem to indicate that prosody is not only (or not always) encoded in the velocity trace, but can be encoded (or extracted) from other sources such as the acceleration and the jerk traces (although we have seen a tendency in preferring the acceleration and the velocity over the jerk). The open question is which prosodic information is (generally) encoded in the acceleration and which is encoded in the velocity. Further experiments will be necessary to answer this general question. It appears that jerk might be an overall distractor, and at least one future study should omit it and pair velocity and acceleration directly against each other in a forced-choice paradigm.

A more important finding was to see that judges tend to prefer one of the three kinematic traces (velocity, acceleration or jerk), depending on who signed the sentences (i.e. depending on the signer shown in the video clips). Several explanations can be given to justify this. Among them it is worth noting that since each signer might have her/his particular way to sign and express things (idiolects), her/his kinematic traces might be biased toward these preferences. Note that all the kinematic traces used above are very dependent on each other, and that the same information can be encoded in each of them (see Section 2). I.e., it is possible to extract the prosodic information from each of the kinematic traces, and only subject dependencies can bias our preference towards one or the other.

These findings are relevant because they can help us to design future experiments. Since we now know that the prosodic cues are encoded differently depending on the sentence and on the signer, we will need to decouple these for separate analysis in future studies. As a consequence, we might conclude that there does not seem to be a unique kinematic trace which encodes all the prosodic cues of ASL.

**Acknowledgements**

The authors would like to thank the people of the Sign Language Linguistics Research lab at Purdue University for their help in preparing and running the experiments shown above. This research was partially supported by NSF grant No. 99-05848-BCS.

**References**

Allen, G. D., Wilbur, R. B. & Schick, B. S. (1991) Aspects of rhythm in American Sign Language. *Sign Language Studies* 72: 297-320.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology* 3: 255-309.

Bolinger, D. (1985). Two views of accent. *Journal of Linguistics* 21: 79-123.

Frishberg, N. & Gough, B. (1973). Morphology in American Sign Language. Manuscript, The Salk Institute for Biological Studies. [Reprinted in *Sign Language & Linguistics* 3:103-131.]

Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27: 765-768.

Fry, D. (1958). Experiments in the perception of stress. *Language and Speech* 1: 125-152.

Ladefoged, P. (1982). *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Klima, E., Bellugi, U., Battison, R., Boyes-Braem, P., Fischer, S., Frishberg, N., Lane, H., Lentz, E. M., Newkirk, D., Newport, E., Pedersen, C. C., and Siple, P. (1979) *The signs of language.* Cambridge, MA: Harvard University.

Liddell, S. K. (1978). Nonmanual signals and relative clauses in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research* (pp. 59-90). New York: Academic Press.

Martinez, A.M., Wilbur, R.B., Shay, R. & Kak, A.C. (2002) The Purdue RVL-SLLRL ASL Database. Tech. Rep. Purdue University.

Wilbur, R. B. (1999) Stress in ASL: Empirical evidence and linguistic issues. *Language and Speech*, 42, 229-250.

Wilbur, R. B. (2000) Phonological and prosodic layering of non-manuals in American Sign Language. In H. Lane and K. Emmorey (eds.), *The signs of language revisited: Festschrift for Ursula Bellugi and Edward Klima*, (pp. 213-241). Hillsdale, NJ: Lawrence Erlbaum.

Wilbur, R. B. & Nolen, S. B. (1986) Duration of syllables in ASL. *Language & Speech* 29/3:263-280.

Wilbur, R. B. & Schick, B. S. (1987) The effects of linguistic stress on ASL signs. *Language & Speech* 30: 301-323.

Wilbur, R. B. & Zelaznik, H. N. (1997, January) *Kinematic correlates of stress and position in ASL.* Linguistic Society of America, Chicago, IL.