How Pragmatics Shape Articulation: A Computational Case Study in STEM ASL Discourse

Saki Imai[1], **Lee Kezar**[2], **Laurel Aichler**[2], **Mert İnan**[1], **Erin Walker**[3], **Alicia Wooten**[2], **Lorna Quandt**[2], **Malihe Alikhani**[1]
[1]Northeastern University [2]Gallaudet University [3]University of Pittsburgh

Abstract

Most state-of-the-art sign language models are trained on interpreter or isolated vocabulary data, which overlooks the variability that characterizes natural dialogue. However, human communication dynamically adapts to contexts and interlocutors through spatiotemporal changes and articulation style. This specifically manifests itself in educational settings, where novel vocabularies are used by teachers, and students. To address this gap, we collect a motion capture dataset of American Sign Language (ASL) STEM (Science, Technology, Engineering, and Mathematics) dialogue that enables quantitative comparison between dyadic interactive signing, solo signed lecture, and interpreted articles. Using continuous kinematic features, we disentangle dialogue-specific entrainment from individual effort reduction and show spatiotemporal changes across repeated mentions of STEM terms. On average, dialogue signs are 24.6%-44.6% shorter in duration than the isolated signs, and show significant reductions absent in monologue contexts. Finally, we evaluate sign embedding models on their ability to recognize STEM signs and approximate how entrained the participants become over time. Our study bridges linguistic analysis and computational modeling to understand how pragmatics shape sign articulation and its representation in sign language technologies.

Keywords: American Sign Language, phonetic reduction, entrainment, motion capture, STEM, education, pragmatics, dialogue

1. Introduction

Human communication is inherently adaptive (Clark and Brennan, 1991), causing context-sensitive phenomena such as **entrainment**, where speakers adjust their linguistic and articulatory behaviors to one another (Brennan and Clark, 1996), and **effort reduction**, where individuals reduce articulatory effort through repetition (Zipf, 2016). While work in dialogue modeling in spoken languages has progressed in recent years, there have been few efforts to model these processes in signed languages, which have important features related to adaptivity, such as sign lowering (Tyrone and Mauk, 2010), weak drop (Padden and Perlmutter, 1987), and location undershooting (Mauk, 2003), distinguishing them from spoken languages. Models aiming to understand or generate naturalistic signing should accommodate these idiosyncratic, context-dependent articulatory shifts.

Yet, most existing sign language models are trained on interpreter data or isolated sign videos (Desai et al., 2024). While these datasets capture clear, standardized productions, they omit the adaptive and fluid behaviors that characterize dialogue, and often reflect the influence of non-native interpreters (Tanzer et al., 2024). To explore these, we test *whether language models generalize beyond isolated sign data to capture interaction-driven articulatory variation.* Addressing this question requires quantifying how signers adapt their articulation in dialogue, and determining whether existing models are sensitive to these pragmatic differences.

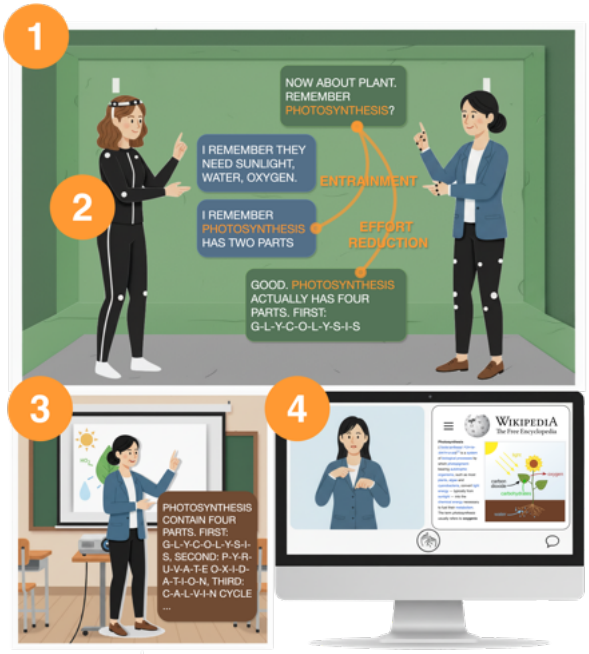

Figure 1: We explore the effects of pragmatics in four different signed STEM contexts: 1) instructor-student dyadic conversations, 2) isolated vocabulary, 3) a signed lecture, and 4) interpreted Wikipedia articles. We computationally analyze how signers establish conceptual pacts across these contexts.

To study this, we collect an American Sign Language (ASL) motion capture dataset that consists of two contexts: 1) an instructor-student dialogue featuring repeated biology signs used in an interactive setting, and 2) isolated productions of biology

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 8476–8490
11–16 May 2026.

vocabulary signs, each produced individually by the student.[1] We supplement these with 3) publicly available recordings featuring a lecture given by the same instructor (to control for individual effort reduction), and 4) interpreter signing (representative of model training data). This approach (Fig. 1) enables us to test whether dialogue exhibits additional convergence beyond per-signer efficiency and to compare natural communicative articulation with the controlled productions that dominate existing signed corpora.

We focus on biology signs because STEM signs in general represent a unique domain for studying communicative adaptation, especially in ASL. Many of these STEM concept signs lack one standardized sign form, with multiple variants circulating among scientists, interpreters, and educators (Lualdi et al., 2023). For instance, ASL users in some schools may sign CELL one way, while students in another region or school may typically fingerspell CELL letter-by-letter, or use a different sign variation, for example CELL (version 2). This fluidity introduces challenges for technologies with specialized vocabulary: systems trained on particular sign forms may fail to recognize the variants used by groups of learners and educators.

Our goal is not to generalize across all ASL use, but to introduce computational tools for quantifying sign articulation, situate articulatory findings from ASL STEM dialogue relative to prior linguistic work, and evaluate whether current sign embedding models capture these context dependent differences that arise in interactive signing. Our contributions are as follows:

1. We introduce a framework that isolates dialogue-specific entrainment from individual effort reduction in ASL, using continuous spatial, temporal, and vertical motion metrics (§ 3.3).
2. We find that dialogue signing exhibits reduced spatial and temporal articulation compared to isolated vocabulary productions, reflecting adaptive changes characteristic of interactive communication (§ 4.1).
3. We show statistically significant spatial and temporal reduction in STEM signs across repeated mentions in dialogue, absent in monologue, indicating that such adaptation may reflect entrainment rather than individual effort reduction (§ 4.2).
4. We demonstrate that existing sign language models fail to generalize to these articulatory differences (§ 4.4).

[1]Data can be made available to researchers upon proper agreements.

2. Related Work

Articulatory Reduction and Interactional Adaptation A line of research argues that spoken languages are shaped by a functional pressure toward ease of articulation and communicative efficiency (Zipf, 2016; Kanwal et al., 2017; Piantadosi et al., 2011; Sigurd et al., 2004; Gibson et al., 2019). In sign languages, similar efficiency pressures shape articulatory form (Napoli et al., 2011; Yin et al., 2024; Caselli et al., 2022). Increased signing rate, for instance, causes signs to be articulated in a lower, more compact space (a phenomenon known as sign lowering) (Tyrone and Mauk, 2010, 2012; Mauk and Tyrone, 2012). Moreover, casual signing tend to involve less joint usage and a shift toward distal articulators (Napoli et al., 2014; Stamp et al., 2022). Beyond individual efficiency, interactive contexts introduce an additional adaptation. In dialogue, signers not only economize their own movements but may also modulate them to align more closely with their interlocutor, a process known as entrainment (Hoetjes et al., 2014), which recent sign language corpora are beginning to capture (Bono et al., 2020, 2024). Despite these insights, characterizations of articulatory adaptation in sign languages remain limited. Prior work has primarily relied on categorical annotation of sign which does not capture the continuous dynamics of articulatory change. In contrast, our work employs motion capture kinematic features to disentangle individual effort reduction from dialogue specific entrainment, building on prior work with motion capture (Lu and Huenerfauth, 2012; Huenerfauth and Lu, 2010).

Sign Language Modeling and Dataset Biases Despite recent progress in sign language recognition, most computational models remain limited in their representation of articulatory variation. Existing large-scale datasets such as PHOENIX-2014T (Camgoz et al., 2018), How2Sign (Duarte et al., 2021), CSL-Daily (Zhou et al., 2021), and WMT-SLT (Mathias et al., 2022), designed primarily for sentence-level translation tasks, often feature interpreter-produced signing performed under controlled conditions. Consequently, this results in a strong bias toward isolated sign articulation, which diverges from the fluid and adaptive behaviors observed in dialogue. Prior work has documented instances where deaf viewers struggle to fully understand such interpretations (Alexander and Rijckaert, 2022). Moreover, the dominant paradigm of sign-to-text translation implicitly treats sign language as a visual analog of spoken language rather than representations grounded in sign-specific discourse structure (Tanzer et al., 2024). Addressing these limitations requires datasets and modeling frameworks that account for the continuous and

discourse-dependent variability of natural signing. Our work contributes to this direction by introducing motion capture analyses of naturalistic ASL dialogue and by testing whether existing sign embedding models capture articulatory differences that emerge through interaction.

3. Methods

In this section, we describe our method of collecting and annotating the STEM dialogue data (§3.1) and finding comparable monologic data (§3.2). Then, we describe our analysis of the data from the perspective of motion capture kinematics (§3.3) and pretrained machine learning models (§3.4).

3.1. Data Collection

Participants Two fluent deaf signers participated in a structured dialogue with one another. One signer is a faculty member in Biology who regularly teaches biology classes in ASL. The other signer is a student familiar with biology concepts from having taken classes with the faculty member previously. Both signers are right-handed.

Signed Content The motion capture data were collected in two conditions:

1. *Isolated STEM vocabulary articulation.* One signer (the student described above) produced a set of 77 STEM signs drawn from introductory biology content. Each sign was produced individually, without a conversational partner. This condition serves as a articulatory baseline, which allows us to estimate each sign's spatial and temporal properties before interactional adaptation occurs. The full list of STEM signs used is provided in the Appendix A.
2. *Biology dialogue.* The two participants engaged in an 8.52-minute spontaneous instructor-student dialogue focusing on key biology topics such as cell structure and genetics, cell cycle, and photosynthesis. While the instructor initiated inquiries to guide the interaction, the dialogue also included extended explanations, clarifications, confirmations, and belief exchanges. Across the dialogue, 17 STEM signs overlapped with the vocabulary condition, allowing for a direct comparison of articulatory variation between isolated and interactive contexts.

Recording Set Up Both motion capture data were recorded at 120 frames per second using a Vicon system. The setup employed 18 high-resolution cameras (8 T160 and 10 Vero) and 73 markers carefully placed on the signer's fingers, hands, and body. Data were captured using Vicon Shogun 1.7, a tool that significantly enhanced the quality of motion capture, particularly in capturing body, hand, and finger movements with high fidelity. In addition to motion capture, we also recorded a 2D RGB video which was used for annotation.

Annotation We annotated our 2D video data using ELAN (Max Planck Institute for Psycholinguistics, The Language Archive, 2025). ELAN uses tiers to annotate different aspects of the source language (e.g. phonology, lexical items). Each video was annotated on two tiers: one for sentence-level translation into English and one for the STEM signs themselves. We followed the SLAASH ID glossing principles (Hochgesang, 2022), specifically focusing on accurate capture of the STEM signs to allow us to perform accurate measures of sign timing differences across the data. A gloss is a conventional written label used by researchers to reference a specific ASL sign. Two hearing, ASL-proficient researchers performed most of the annotation, consulting with the Deaf faculty member signer and a hearing fluent signer with expertise in STEM signs (Figure 2).

3.2. Supplementary Monologic Data

To contextualize the types of signing observed in STEM dialogue, we compared the collected data against *monologic* ASL signing covering the same Biology topics as the dialogue: ASL STEM Wiki (Yin et al., 2024) and Atomic Hands (Wooten and Spieker, 2022).

ASL STEM Wiki provides translations of STEM articles on Wikipedia, signed by certified ASL interpreters. From the 254 articles, we selected samples that match the topics found in the dialogue, namely, *Reproduction* and *Photosynthesis*. From these two articles, we selected excerpts 8 and 12 sentences respectively[2] from the opening section to maximize lexical overlap with the dialogue. This dataset represents interpreter signing that closely resembles the dominant training data for sign language understanding models. It therefore serves as a reference point for evaluating how dialogue signing diverges from data that shape existing technologies.

With permission from the creators, we additionally analyze videos from Atomic Hands website (Wooten and Spieker, 2022). Atomic Hands provides educational resources for STEM topics in ASL, signed by deaf experts in their disciplines. We found one video ("ASL Signs – cell division, mitosis, meiosis") that discusses the relevant topics in the dialogue. The signer in this video is the same participant as the instructor in the dialogue. This

[2]These are contiguous in the article, starting with §5 in Reproduction and §1 in Photosynthesis.

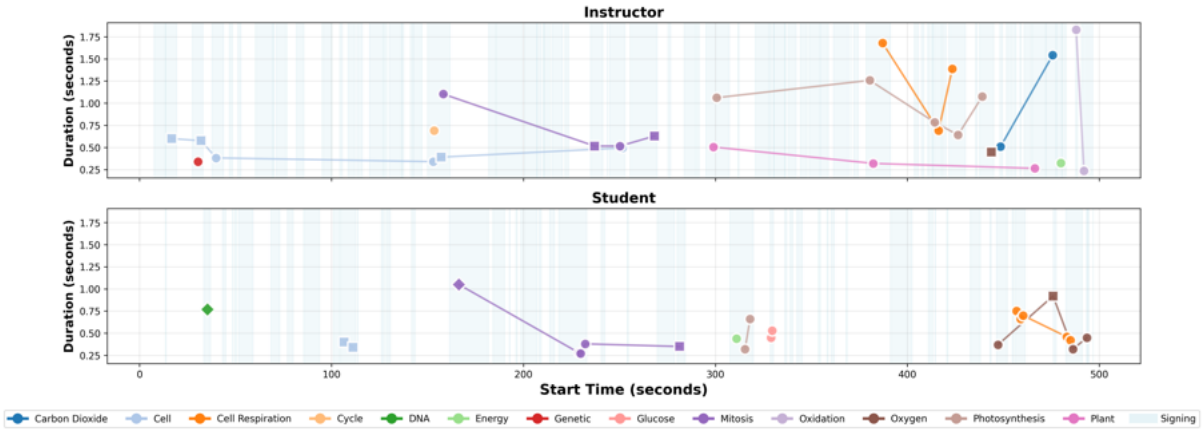

Figure 2: Sign duration patterns over time for instructor-student ASL dialogue. Stacked panels show start time (x) vs. duration (y) of individual signs for both participants, with connected lines indicating temporal progression of repeated signs within semantic groups. Each point represents an individual STEM sign instance, colored by semantic base term and shaped by articulation variation (circles, squares, diamonds for fingerspelling). Light blue background shading indicates active signing periods for each participant.

dataset allows to compare dialogue signing against solo instructional signing by the same individual. This comparison helps distinguish individual effort reduction tendencies from entrainment effects that arise uniquely in dialogue.

We repeat the annotation process on these videos and include them in our experiments and analyses.

Extracting Keypoints While the motion capture recordings provided 3D joint coordinates (as described in § 3.1), the additional video data consisted of 2D RGB recordings in MP4 format. To obtain a comparable set of kinematic features from these videos, we processed them using MediaPipe Holistic (Zhang et al., 2020), which extracts 543 landmarks spanning face, hands, and body. We used the default detection and tracking confidence thresholds of 0.5. We then mapped MediaPipe's landmark indices to the corresponding motion capture joint positions used in our analyses. The full list of joint correspondences between MediaPipe and the motion capture joints is provided in Appendix B.

3.3. Articulatory Metrics

To quantify articulatory variability in ASL production, we adapted and extended kinematic measures commonly used in gesture studies. These metrics capture changes in spatial, temporal, and vertical movements across repeated sign articulations. We use these metrics to separate (i) *individual effort reduction*, where signers economize movement through repetition regardless of interaction, from (ii) *interaction-driven entrainment*, where articulatory patterns adapt in response to an interlocutor. Across all metrics, *reduction* is operationalized as decreases in movement magnitude, duration, or spatial extent across repeated mentions. When such reductions occur in dialogue but are absent or attenuated in monologue by the same signer, we interpret them as evidence of entrainment rather than general articulatory efficiency. Conversely, reductions observed in both dialogue and monologue are interpreted as reflecting individual effort reduction.

3.3.1. Spatial Reduction

We operationalize the spatial properties of sign articulation using two measures:

Spatial Extent captures the overall 3D volume occupied by a sign. For a trajectory $\mathbf{p}_i = (x_i, y_i, z_i)$ over frames $i = 1, \ldots, n$, we compute the per-dimension ranges $\Delta x = \max_i x_i - \min_i x_i$, $\Delta y = \max_i y_i - \min_i y_i$, and $\Delta z = \max_i z_i - \min_i z_i$. The spatial extent is defined as the diagonal length of the corresponding bounding box:

$$\text{SpatialExtent} = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}.$$

This metric captures how much space an articulator occupies during signing (Bevacqua et al., 2006).

Path Length measures the cumulative distance traveled by the articulator across frames (Shaw and Anthony, 2016; Trujillo et al., 2020). For consecutive positions $\mathbf{p}_i, \mathbf{p}_{i+1}$:

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2},$$

and the total path length is $\sum_{i=1}^{n-1} d_i$.

3.3.2. Temporal Reduction

Temporal reduction reflects efficiency in the timing of sign articulation:

Average Velocity captures the mean speed of the articulator over the course of a sign.

Articulation Duration measures the total time span of a sign. Reductions in duration across repeated mentions are commonly interpreted as efficiency gain (Hoetjes et al., 2014). This metric does not utilize kinematic features.

3.3.3. Sign Lowering

Sign lowering captures the vertical displacement of the hands, often observed as signs being produced progressively lower in signing space over time or across repetitions (Tyrone and Mauk, 2010, 2012; Mauk and Tyrone, 2012). For each sign instance, we calculate the mean vertical position of the dominant hand during articulation.

3.4. Pre-Trained Sign Encoders and Evaluation Setup

We next assess whether current state-of-the-art encoder models for ASL can generalize beyond isolated sign data to the articulatory variation found in dialogue. Such models are commonly trained on data that is not representative of signing among deaf and hard-of-hearing people (Desai et al., 2024), such as isolated sign productions and hearing signers of varying skill levels. In contrast, our eight-minute biology dialogue presents a realistic conversation between two deaf signers, using out-of-vocabulary STEM signs while also adapting those signs through reduction, coarticulation, and entrainment. By testing pre-trained encoders in this domain, we ask whether their embedding spaces remain stable across contextual variation and whether they can still recover lexical identity when signs deviate from canonical form.

Models We examine two pre-trained encoders with complementary input modalities and training regimes: SignCLIP and I3D. **(1) SignCLIP** (Jiang et al., 2024) uses MediaPipe skeletal keypoints as input and encodes sign clips via a frozen 3D CNN video encoder paired with a BERT-based text encoder (Devlin et al., 2019). It is trained with a CLIP-style contrastive loss (Radford et al., 2021) to align signs and glosses using 500k sign videos from 44 sign languages in the SpreadTheSign dictionary (Hilzensauer and Krammer, 2015). We use the ASL finetuned checkpoint[3] which is further finetuned on Popsign ASL (Starner et al., 2023), ASL Citizen (Desai et al., 2023), and Sem-Lex (Kezar et al., 2023). **(2) I3D** (Varol et al., 2021) is a 25M-parameter inflated 3D ConvNet ("I3D") pretrained on Kinetics action recognition dataset (Carreira and Zisserman, 2017) with additional transformer layers fine-tuned on 1,000 hours of BSL broadcast footage (BSL-1K; Albanie et al. 2020). It operates on RGB video input and is trained to classify gloss labels in continuous signing[4].

Creating Avatars The video recordings of the dialogue differ substantially from standard datasets for training sign language models. Specifically, the participants wore black motion capture suits, lowering visual contrast, and the signers were positioned at an angle rather than facing the camera directly. To bring these data closer to in-distribution signing, we used Unity (Unity Technologies, 2023) to render 3D avatars from the motion capture recordings. These videos preserved articulatory motion and spatial trajectories while providing consistent signer contrast and viewing angle. Example avatar renderings are shown in Appendix C.

3.4.1. Embedding-based Entrainment Metrics

To explore whether pretrained encoder embeddings are sensitive to conversational adaptation, we analyze repeated STEM sign productions across time. For each gloss g with multiple occurrences per signer, we extract embedding vectors $x_1, x_2, \ldots, x_T \in \mathbb{R}^d$ for each production, using mean-pooled and L2-normalized outputs from the same encoders (I3D, SignCLIP).

We define:

- $\Delta\cos = \cos(x_T^A, x_T^B) - \cos(x_1^A, x_1^B)$: change in cross-signer similarity between first and last tokens.
- $s_{\mathsf{A}\to\mathsf{B}}$: the slope of $\cos(x_i^{\mathsf{A}}, x_1^{\mathsf{B}})$ over i, indicating whether signer A becomes more similar to signer B's initial token over time.
- $\text{selfsim}_{\mathsf{A}} = \cos(x_1^{\mathsf{A}}, x_T^{\mathsf{A}})$: within-signer temporal stability.

We compute these metrics for all glosses with ≥ 2 tokens per signer under both *raw video* and *avatar-rendered* conditions. Positive values of $\Delta\cos$ or $s_{\mathsf{A}\to\mathsf{B}}$ are interpreted as evidence of entrainment, while negative values suggest signer-specific differentiation or self-entrainment.

3.4.2. Sign Spotting Task

To test these models' generalizability, we isolate STEM productions from the dialogue (manually annotated in §3.1) and evaluate them as search queries in a *continuous sign spotting* task, where the model searches a sentence-level sequence

[3] https://drive.google.com/drive/folders/10q7FxPlicrfwZn7_FgtNqKFDiAJi6CTc

[4] https://www.robots.ox.ac.uk/~vgg/research/bslattend/data/bsl5k.pth.tar

for matching occurrences of the query. Success on this task could lead to downstream educational technologies, such as searching a corpus of STEM content for examples of a specific sign. We use the isolated productions (described in §3.1) as queries and search the dialogic and monologic sentences using a sliding window with width = stride = 0.5 seconds. We additionally consider reformatting the dialogic inputs as an avatar to reduce the noise attributed to the motion capture suits.

We implement a baseline model that ranks candidate windows by cosine similarity to the query embedding. From this ranking, we report recall at $k = \{10, 50\}$, defined as the proportion of top-k windows that overlap with the ground-truth sign with intersection over union (IoU) ≥ 0.3. We also report mean reciprocal rank (MRR), computed as the average of $1/\text{rank}(w)$ over all windows w with IoU ≥ 0.3.

4. Results

We first compare dialogue articulations with *Vocabulary* baselines to characterize how signing style differs across communicative contexts (§ 4.1). This analysis primarily uses left and right hand motion capture data, as the hands are the primary articulators that account for the majority of articulatory effort and spatial contrast in signing. The subsequent section expands the analysis to include more joints to examine how these articulatory properties evolve across repeated mentions within the dialogue, and how that differs from a monologue context, such as solo lecture and interpreted articles (§ 4.2). Finally, we evaluate whether current sign language models can capture these lexical variations through *continuous sign spotting* (§ 4.4) experiment, and whether those embeddings are sensitive to articulatory adaptation in dialogue (§ 4.3).

4.1. Comparison with Vocabulary Articulation

To quantify how dialogue signing diverges from isolated vocabulary form and how it changes with repetition, we computed per-sign differences in kinematic properties defined in Section 3.3 between *Dialogue* data and *Vocabulary* conditions. In our setup, the vocabulary signs produced by the student serve as the baseline reference for both participants. Each plotted trajectory (Figures 3) shows the difference in a given metric relative to its vocabulary baseline on the y-axis, as a function of mention order on the x-axis. Thus, points below zero indicate more compact articulation than the vocabulary form, and the slope across mentions indicates how this deviation evolves with repetition.

Path Length Difference As shown in Figure 3, most signs show negative and decreasing Δ Path length across mention order. Although there are clear inter-participant differences, these patterns appear modulated by the communicative goals of each role. The instructor shows relatively stable articulation across repetitions to ensure clarity, whereas the student who often reuses STEM terms already introduced by the instructor shows greater reduction from vocabulary form. The average trajectory line for each plot suggests a general downward trend, but large standard deviations show that the extent of reduction varies substantially by lexical item and interactional context.

Duration Difference Across all STEM signs with matching vocabulary baselines, signs were on average **24.6%** shorter for the instructor, and **44.6%** shorter for the student, relative to their corresponding vocabulary articulation, both statistically significant reductions ($p < .01$). These results suggest that interactive signing involves not only spatial contraction but also temporal compression, which is consistent with efficiency-driven adaptation (Tyrone and Mauk, 2010; Hoetjes et al., 2014).

Sign Lowering No significant differences were found in vertical hand position for either hand ($p > .2$), suggesting that while dialogue articulation exhibits clear spatial and temporal reduction, it does not consistently involve a downward shift relative to vocabulary baselines. This aligns with prior findings that sign lowering is highly context dependent and influenced by various factors such as sentence position, coarticulatory effects and prosody (Tyrone and Mauk, 2012).

4.2. Repeated Mention Analysis

To quantify articulatory reduction across repeated mentions of a sign, we computed relative percentage change using the first occurrence of each sign as a baseline. For every sign that appeared at least twice in the corpus, we extracted joint-level motion features for all mentions and calculated the proportional reduction of each subsequent token relative to its first mention. This approach controls for signer-specific baseline differences in articulation while comparing within-sign trajectories across repetitions.

For each joint, we then tested whether the degree of reduction increased systematically with mention order using Spearman rank correlations (Spearman, 1961) between mention index and percent reduction. Positive correlations indicate reduction (i.e., smaller spatial extent or path length, or faster velocity) as a function of repetition. Table 1 reports the resulting correlations for the instructor's produc-

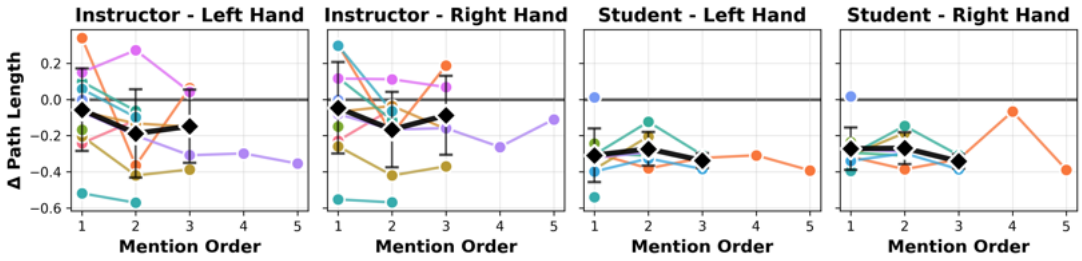

Figure 3: Path length differences between dialogue and vocabulary articulation for left and right hands as a function of mention order. Colors correspond to individual signs and the bolded black line is the mean with standard error. A value of 0 indicates no difference between dialogue and vocabulary path lengths. Negative values indicate shorter trajectories in dialogue. Overall, dialogue signing shows consistently shorter and progressively reduced movement paths relative to vocabulary signs.

Joint	Spatial Reduction			Path Reduction			Velocity Increase		
	Dialogue	Monologue		Dialogue	Monologue		Dialogue	Monologue	
Fingers (L)	$+0.66^{***}$	-0.41^{*}	$(-)$	$+0.63^{***}$	-0.24	$(-)$	-0.52^{**}	$+0.23$	$(-)$
Fingers (R)	-0.08	-0.17	$(+0.54)$	$+0.09$	-0.33	$(+0.78)$	$+0.14$	$+0.28$	$(+0.45)$
Hand (L)	$+0.71^{***}$	-0.29	$(+0.34)$	$+0.68^{***}$	-0.35	$(+0.34)$	-0.55^{**}	$+0.50^{**}$	$(+0.19)$
Hand (R)	$+0.30$	-0.34	$(+0.30)$	$+0.62^{***}$	-0.39^{*}	$(+0.64)$	-0.12	$+0.48^{**}$	$(+0.48)$
Forearm (L)	$+0.48^{*}$	-0.25	$(+0.73^{*})$	$+0.64^{***}$	-0.34	$(+0.41)$	-0.36	$+0.41^{*}$	$(+0.44)$
Forearm (R)	-0.04	-0.38^{*}	$(+0.66)$	$+0.24$	-0.41^{*}	$(+0.66)$	-0.05	$+0.51^{**}$	$(+0.57)$
Arm (L)	$+0.44^{*}$	-0.23	$(+0.40)$	$+0.50^{*}$	-0.33	$(+0.35)$	-0.48^{*}	$+0.45^{*}$	$(+0.06)$
Arm (R)	$+0.35$	$+0.08$	$(+0.62)$	$+0.23$	-0.12	$(+0.57)$	$+0.04$	$+0.15$	$(+0.66)$

Table 1: Relative Change Analysis (Spearman correlation). Stars denote significance (* $p < .05$, ** $p < .01$, *** $p < .001$). Values in the parentheses indicate the *interpreter* condition, and non-parenthesized *monologue* values correspond to the solo lecture by the same instructor in the *dialogue* condition. Dialogue signing shows systematic spatial and temporal reduction, particularly in the left hand and arm, while monologue and interpreter signing exhibits weaker, inverse or inconsistent trends. The Fingers (L) values for the interpreter condition could not be computed due to Mediapipe failing to extract hand keypoints.

tions in the dialogue (§ 3.1) and monologue (§ 3.2) contexts. Comparing with the monologue condition (of the same instructor) allows us to study whether articulatory reduction across repetitions arises from individual tendencies toward effort reduction or from adaptation to an interlocutor. Results for the student signer are included in Appendix D, as the limited number of repeated signs in the dialogue (see Fig. 2) reduced the statistical power of within-sign analyses.

4.2.1. Spatial Reduction

Spatial Extent and Path Length Both spatial extent and path length showed systematic reduction across repeated mentions in the dialogue condition. The instructor showed significant positive correlations between mention order and reduction for several left-hand joints. This pattern may suggest a phenomenon known as *weak drop* in which one hand (typically the non-dominant hand) is omitted during production of a two handed sign (Padden and Perlmutter, 1987).

In contrast, the monologue condition showed little evidence of spatial reduction and, in some cases, negative correlations. This likely reflects the communication goals of lecture-style signing, where clarity and precision take precedence over articulatory economy.

4.2.2. Temporal Reduction

Average Velocity Temporal metrics revealed similar patterns. In the dialogue condition, several left-side joints exhibited significant negative correlations between motion order and velocity increase. In contrast, positive correlations were found for some right-side joints in the monologue, which may reflect emphasis through larger and more deliberate motions to maintain clarity.

Articulation Duration In the dialogue condition, the instructor's sign durations decreased significantly with repetition ($r = 0.279, p < .001$), indicating that signs become approximately 27.9% shorter

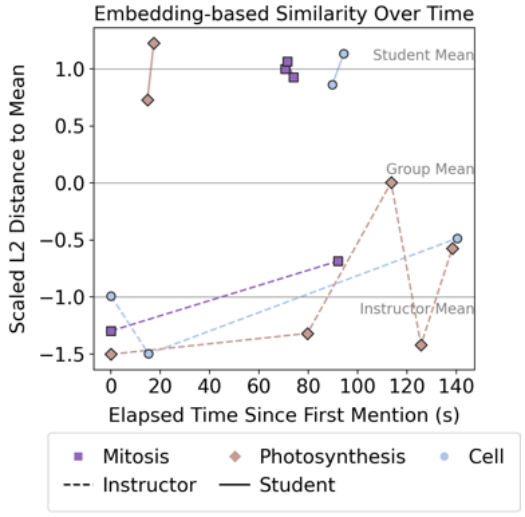

Figure 4: Sign productions plotted with respect to their embeddings' L2 distance to the mean of each signer's average and the group's average.

Input	Model	MRR	R@10	R@50
$D_{\text{inst.}}$	I3D	0.009	0.043	0.087
	SignCLIP	0.004	0.000	0.000
$D^{\Diamond}_{\text{inst.}}$	I3D	0.003	0.000	0.000
	SignCLIP	0.003	0.000	0.000
$D_{\text{stud.}}$	I3D	0.013	0.048	0.095
	SignCLIP	0.003	0.000	0.000
$D^{\Diamond}_{\text{stud.}}$	I3D	0.011	0.048	0.095
	SignCLIP	0.013	0.048	0.095
$M_{\text{inst.}}$	I3D	0.026	0.086	0.229
	SignCLIP	0.021	0.029	0.171
M_{wiki}	I3D	0.005	0.000	0.071
	SignCLIP	0.010	0.000	0.214

Table 2: Retrieval performance (mean reciprocal rank, recall@k) by input context and encoder. Key: D=dialogue, M=monologue, $\Diamond$=avatar rendering.

on average over repeated mentions. The student data followed a similar trend ($r = 0.295, p < .001$). These reductions across interlocutors suggest a temporal entrainment effect, where both signers adapt their articulatory timing patterns over the course of interaction (Brennan, 1996; Garrod and Anderson, 1987). In the monologue, by contrast, no significant duration change was found ($r = 0.026, p = 0.895$). The absence of reduction when signing alone further suggests that the observed compression in dialogue arises from interactional adaptation rather than articulatory efficiency.

4.3. Embedding-Based Entrainment

Having examined the effect of pragmatic context using kinematic measures, we now evaluate the learned representations from pre-trained sign language models.

We focus on the SignCLIP encoder, excluding I3D due to near-zero variation across time ($\Delta \cos < 0.0002$). This is likely a result of the model's training on frontal videos, which mismatches the side-facing camera perspective in our dataset. By contrast, SignCLIP operates on normalized pose inputs. Three signs met the minimum requirement of $n > 1$ productions for each signer: MITOSIS, PHOTOSYNTHESIS, and CELL.

All signs exhibit a clear increase in cross-signer similarity from the first to the last repetition, with $\Delta \cos > 0$ for all three (CELL: $+0.0070$, MITOSIS: $+0.0422$, PHOTOSYNTHESIS: $+0.0785$), providing limited evidence of entrainment.

The student's self-similarity is consistently higher than the instructor's across all signs with average scores of $\text{selfsim}_{\text{stud.}} = 0.963$ and $\text{selfsim}_{\text{inst.}} = 0.942$ across the three signs. This indicates that the student's productions remained more internally stable across time.

Slope-based analyses show that the student tends to diverge slightly from the instructor $s_{\text{stud.}\to\text{inst.}} < 0$, with slopes of -0.0329 for PHOTOSYNTHESIS, -0.0022 for CELL, and $+0.0039$ for MITOSIS. Instructor-to-student slopes are consistently positive, with slopes of $5e^{-4}, 9e^{-4}, 1e^{-4}$ respectively.

Figure 4 visualizes each embedding x_t's alignment relative to the mean embeddings of the student $\mu_{\text{stud.}}$ and instructor $\mu_{\text{inst.}}$. To quantify this alignment along a single axis, we compute a similarity score by projecting x_t onto the line that connects the signer means, and scale to $[-1, 1]$:

$$\text{sim}(x_t) = \frac{2(x_t - \mu_{\text{inst.}})^\top (\mu_{\text{stud.}} - \mu_{\text{inst.}})}{\|\mu_{\text{stud.}} - \mu_{\text{inst.}}\|^2} - 1$$

The student's embeddings generally move *away* from the instructor's mean (by a small amount) while the instructor's move *toward* the student's. Taken together, these patterns suggest that the instructor adapted more to the student's signing than vice versa, consistent with instructor-side entrainment rather than student convergence. However, these findings are limited by the models' performance on recognition tasks, like sign spotting.

4.4. Sign Spotting

We evaluate sign spotting performance using pre-trained encoders across the four contexts. Table 2 reports MRR and recall at $k = \{10, 50\}$.

Performance varies widely across signer and context. The highest scores are observed in the

Instructor-monologue condition, with I3D achieving the top MRR of 0.026 and SignCLIP close behind at 0.021. This aligns with expectations, as the monologue setting most closely resembles the direct-facing data seen during model pretraining. In contrast, the dialogue settings consistently lead to lower performance.

Searching over the student's signing yields slightly better results than the instructor's, with I3D scoring 0.013 MRR in the former compared to 0.009 in the latter. This asymmetry supports the hypothesis that query and search embeddings drawn from the same signer are more similar in latent space than embeddings drawn across signers. SignCLIP, however, performs poorly in both dialogue contexts, failing to retrieve any correct items within the top-50 ranks. When the student's signing is rendered with an avatar, the SignCLIP model gains 0.01 MRR, but in all other cases the avatar reduced performance.

Overall, the findings confirm that signer identity and pragmatic context heavily influence retrieval outcomes. No encoder achieves strong performance across all conditions, suggesting that current sign-receptive models do not generalize well to lexical innovations, such as in STEM contexts. Furthermore, high performance in monologue but low recall in dialogue highlights the need for models that are robust to pragmatic variation and signer-specific features.

5. Discussion

This work provides the first quantitative evidence that pragmatic adaptation in sign language follows measurable articulatory principles comparable to spoken dialogue (Zipf, 2016; Kanwal et al., 2017; Piantadosi et al., 2011). Signs produced in dialogue were both spatially and temporally reduced relative to isolated sign articulations, with reductions intensifying across repeated mentions. These findings mirror the functional pressures toward articulatory economy observed in spoken language (Zipf, 2016; Lindblom, 1990).

Significant reductions in left hand and arm movements indicate selective economization of non-dominant articulators, consistent with the weak drop phenomenon (Padden and Perlmutter, 1987). This suggests that pragmatic adaptation in sign language is not a uniform compression but a targeted modulation that preserves communicative clarity while minimizing redundant effort. Additionally, different patterns in the instructor and the student's signing suggest that communicative goals and power dynamics may exert pressure upon signed productions. We observed that the instructor was more likely to sign slowly and match the student's signing. The instructor may prioritize clarity and match the student's articulations, while the student aims for conciseness and demonstrates a lesser tendency to align with the instructor's signing.

From a computational perspective, our findings highlight the gap between linguistic adaptation and machine representation. Although current sign embedding models such as SignCLIP and I3D achieve high accuracy on interpreter and isolated sign benchmarks, their performance degrades in interactive settings that feature pragmatic variation. This suggests that these models largely capture lexical and visual regularities but fail to encode the gradient articulatory variability characteristic of real-world signing.

Our results highlight a broader limitation in how current sign language technologies conceptualize variation: models are trained to recognize the sign produced, but not how or why it varies. By incorporating motion capture kinematics into dialogue analysis, this study emphasizes the importance of modeling approaches that view signing as an adaptive and interactional system, rather than a sequence of static lexical items.

6. Conclusion

This study presented an empirical analysis of how dialogic ASL signing differs from isolated vocabulary and monologic articulations through quantitative analyses of motion capture recording in STEM discourse. Using continuous spatial, temporal, and vertical motion metrics, we found that dialogic signing is characterized by articulatory reduction, which is absent in solo lecture contexts. Our analyses further revealed that current embedding models struggle to generalize to such pragmatic variation due to the limitations of training paradigms that rely predominantly on interpreter or isolated vocabulary data.

By modeling how STEM signs are adapted in interactive ASL dialogue, our findings highlight the importance of developing educational technologies and sign language models that are robust to pragmatic variation. Future work should extend these analyses to larger datasets, diverse signers, and additional discourse contexts to better understand how communicative pressures drive articulatory adaptation across modalities.

Limitations

Data Limitations While our dataset provides a unique and naturalistic ASL dialogue grounded in STEM education, it remains limited in scope. The number of participants is small, consisting of a single instructor–student pair, and the duration of the recorded dialogue (8.52 minutes) restricts the range of lexical and interactional variation that can

be analyzed. For example, a different pair of interlocutors are likely to use different regional variations for STEM concepts and may show different levels of entrainment and effort reduction. Consequently, our findings should be interpreted as a case study rather than generalization.

The video modality of the dialogue features participants wearing black body suits, which creates visual noise for both human annotators and machine learning models. We attempted to remedy this effect with an avatar-based rendering of the motion capture data; nonetheless, both formats are considerably out-of-distribution for the pretrained models, as evidenced by our findings.

Annotation Challenges The authors who annotated the ASL data are hearing signers. Additionally, the analyses were conducted by hearing signers and non-signers. This lack of annotation and analysis directly from native ASL users may introduce bias into the analyses and interpretations, potentially leaving out certain valid explanations of the observed phenomena or undermining the quality of the annotations. Proofing was conducted across both signers' annotations; however, judgments about translation correctness and sign start/stop frame are inherently subjective.

One challenge of annotating ASL and other sign languages is the three-dimensional nature of the languages and the lack of a widely accepted writing or annotation system. Because of this gap, language researchers have historically relied on glossing ASL with words borrowed from English. This presents several problems: 1) glosses chosen are often not consistent, leading to confusion about which sign is meant, 2) the English meaning of the word chosen often does not line up with the meaning of the ASL sign which it is supposed to represent, which can introduce skews in understanding when communicating about signs, 3) the written gloss does not show the form of the sign, leading to a lack of access for researchers who do not already sign. The SLAASH ID Glossing Principles (Hochgesang, 2022) address some of these longstanding issues. They emphasize the importance of bridging the distance between actual sign production and the data presented to researchers by linking annotations directly to video and always providing a video link for signs that are glossed.

Ethical Considerations

Given that ASL is a low-resource language and the language closely associated with a minority community in the United States (and to some extent, other languages), we note the importance of having cultural and linguistic knowledge that derives from lived experience, and not only academic study. Multiple authors of this paper use ASL in their daily or workday lives and have concrete connections within Deaf communities of the United States. One author/project leader is a Deaf fluent signer, as was the "instructor" in the motion capture dataset. We recognize the inherent difficulties and weaknesses of carrying out research on minority languages with sizable research teams whose relationships to the languages vary widely, and hope that by sharing our positionality, we may be better able to "join the conversation" regarding signed language research and emerging technologies (Desai et al., 2024).

The influence of pragmatics is complex, and neither the data introduced in this work nor the analyses we conducted are fully representative of this complexity. This limitation is especially important in educational contexts, where a diverse range of learning preferences, cultural backgrounds, and ways of languaging influence the surface form of signing. The work presented here represents an early attempt to computationally characterize specific aspects of sign language pragmatics and inform further research on this topic.

Acknowledgement

This research was supported in part by the U.S. National Science Foundation under Awards No. 2418662, 2418663 and 2418664. We thank Deanna Dunlop, Jason Lamberton, Thalia Guettler, and Joseph Palagano for their assistance with data collection and annotation.

Bibliographical References

Samuel Albanie, Gül Varol, Liliane Momeni, Triantafyllos Afouras, Joon Son Chung, Neil Fox, and Andrew Zisserman. 2020. Bsl-1k: Scaling up co-articulated sign language recognition using mouthing cues. In *European conference on computer vision*, pages 35–53. Springer.

Dhoest Alexander and Jorn Rijckaert. 2022. News 'with'or 'in'sign language? case study on the comprehensibility of sign language in news broadcasts. *Perspectives*, 30(4):627–642.

Elisabetta Bevacqua, Amaryllis Raouzaiou, Christopher Peters, George Caridakis, Kostas Karpouzis, Catherine Pelachaud, and Maurizio Mancini. 2006. Multimodal sensing, interpretation and copying of movements by a virtual agent. In *International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*, pages 164–174. Springer.

Mayumi Bono, Tomohiro Okada, Victor Skobov, and Robert Adam. 2024. Data integration, annotation, and transcription methods for sign language dialogue with latency in videoconferencing. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 26–35, Torino, Italia. ELRA and ICCL.

Mayumi Bono, Rui Sakaida, Tomohiro Okada, and Yusuke Miyao. 2020. Utterance-unit annotation for the JSL dialogue corpus: Toward a multimodal approach to corpus linguistics. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 13–20, Marseille, France. European Language Resources Association (ELRA).

Susan Brennan and Herbert Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1482–1493.

Susan E Brennan. 1996. Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, 96:41–44.

Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

João Carreira and Andrew Zisserman. 2017. Quo vadis, action recognition? a new model and the kinetics dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4733.

Naomi Caselli, Corrine Occhino, Bruno Artacho, Andreas Savakis, and Matthew Dye. 2022. Perceptual optimization of language: evidence from american sign language. *Cognition*, 224:105040.

Herbert H Clark and Susan E Brennan. 1991. Grounding in communication.

Aashaka Desai, Lauren Berger, Fyodor Minakov, Nessa Milano, Chinmay Singh, Kriston Pumphrey, Richard Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, et al. 2023. Asl citizen: a community-sourced dataset for advancing isolated sign language recognition. *Advances in Neural Information Processing Systems*, 36:76893–76907.

Aashaka Desai, Maartje De Meulder, Julie A. Hochgesang, Annemarie Kocab, and Alex X. Lu. 2024. Systemic biases in sign language AI research: A deaf-led call to reevaluate research agendas. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 54–65, Torino, Italia. ELRA and ICCL.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i Nieto. 2021. How2sign: a large-scale multimodal dataset for continuous american sign language. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2735–2744.

Simon Garrod and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181–218.

Edward Gibson, Richard Futrell, Steven P Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. How efficiency shapes human language. *Trends in cognitive sciences*, 23(5):389–407.

M. Hilzensauer and K. Krammer. 2015. A multilingual dictionary for sign languages: "spreadthesign". In *ICERI2015 Proceedings*, 8th International Conference of Education, Research and Innovation, pages 7826–7834. IATED.

Marieke Hoetjes, Emiel Krahmer, and Marc Swerts. 2014. Do repeated references result in sign reduction? *Sign Language & Linguistics*, 17(1):56–81.

Matt Huenerfauth and Pengfei Lu. 2010. Eliciting spatial reference for a motion-capture corpus of american sign language discourse. In *signlang@ LREC 2010*, pages 121–124. European Language Resources Association (ELRA).

Robert E Johnson and Scott K Liddell. 2010. Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies*, 11(2):241–274.

Robert E Johnson and Scott K Liddell. 2011a. A segmental framework for representing signs phonetically. *Sign Language Studies*, 11(3):408–463.

Robert E Johnson and Scott K Liddell. 2011b. Toward a phonetic representation of hand configuration: The fingers. *Sign Language Studies*, 12(1):5–45.

Robert E Johnson and Scott K Liddell. 2012. Toward a phonetic representation of hand configuration: The thumb. *Sign Language Studies*, 12(2):316–333.

Robert E Johnson and Scott K Liddell. 2021. Toward a phonetic description of hand placement on bearings. *Sign Language Studies*, 22(1):131–180.

Jasmeen Kanwal, Kenny Smith, Jennifer Culbertson, and Simon Kirby. 2017. Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165:45–52.

Lee Kezar, Jesse Thomason, Naomi Caselli, Zed Sehyr, and Elana Pontecorvo. 2023. The semlex benchmark: Modeling asl signs and their phonemes. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–10.

Robert M Krauss and Sidney Weinheimer. 1964. Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1:113–114.

Scott K Liddell. 1984. Think and believe: sequentiality in american sign language. *Language*, pages 372–399.

Scott K Liddell and Robert E Johnson. 1986. American sign language compound formation processes, lexicalization, and phonological remnants. *Natural Language & Linguistic Theory*, 4(4):445–513.

Scott K Liddell and Robert E Johnson. 1989. American sign language: The phonological base. *Sign language studies*, 64(1):195–277.

Scott K Liddell and Robert E Johnson. 2019. Sign language articulators on phonetic bearings. *Sign Language Studies*, 20(1):132–172.

Björn Lindblom. 1990. Explaining phonetic variation: A sketch of the h&h theory. In *Speech production and speech modelling*, pages 403–439. Springer.

Pengfei Lu and Matt Huenerfauth. 2012. CUNY American Sign Language motion-capture corpus: First release. In *Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, pages 109–116, Istanbul, Turkey. European Language Resources Association (ELRA).

Colin P Lualdi, Barbara Spiecker, Alicia K Wooten, and Kaitlyn Clark. 2023. Advancing scientific discourse in american sign language. *Nature Reviews Materials*, 8(10):645–650.

Müller Mathias, Ebling Sarah, Camgöz Necati Cihan, Jiang Zifan, Battisti Alessia, Moryossef Amit, Rios Annette, Bowden Richard, and Wong Ryan. 2022. Wmt-slt focusnews: Training data for the wmt shared task on sign language translation.

Claude E Mauk and Martha E Tyrone. 2012. Location in asl: Insights from phonetic variation. *Sign Language & Linguistics*, 15(1):128–146.

Claude Edward Mauk. 2003. *Undershoot in two modalities: Evidence from fast speech and fast signing*. The University of Texas at Austin.

Donna Jo Napoli, Nathan Sanders, and Rebecca Wright. 2011. Some aspects of articulatory ease in american sign language. *Handout from Stony Brook University, May*, 6:2011.

Donna Jo Napoli, Nathan Sanders, and Rebecca Wright. 2014. On the linguistic effects of articulatory ease, with a focus on sign languages. *Language*, 90(2):424–456.

Carol A Padden and David M Perlmutter. 1987. American sign language and the architecture of phonological theory. *Natural language & linguistic theory*, 5(3):335–375.

Steven T Piantadosi, Harry Tily, and Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*.

Alex Shaw and Lisa Anthony. 2016. Analyzing the articulation features of children's touchscreen gestures. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16, page 333–340, New York, NY, USA. Association for Computing Machinery.

Bengt Sigurd, Mats Eeg-Olofsson, and Joost Van Weijer. 2004. Word length, sentence length and frequency–zipf revisited. *Studia linguistica*, 58(1):37–52.

Charles Spearman. 1961. The proof and measurement of association between two things.

Rose Stamp, Lilyana Khatib, and Hagit Hel-Or. 2022. Capturing distalization. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 187–191.

Thad Starner, Sean Forbes, Matthew So, David Martin, Rohit Sridhar, Gururaj Deshpande, Sam Sepah, Sahir Shahryar, Khushi Bhardwaj, Tyler Kwok, et al. 2023. Popsign asl v1. 0: An isolated american sign language dataset collected via smartphones. *Advances in Neural Information Processing Systems*, 36:184–196.

C Stokoe William. 1960. Sign language structure: An outline of the visual communication systems of the american deaf (studies in linguistics occasional papers 8) buffalo. *NY: University of Buffalo*.

Ted Supalla. 1978. How many seats in a chair? *Understanding language through sign language research*.

Garrett Tanzer, Maximus Shengelia, Ken Harrenstien, and David Uthus. 2024. Reconsidering sentence-level sign language translation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6262–6287, Miami, Florida, USA. Association for Computational Linguistics.

James P Trujillo, Irina Simanova, Harold Bekkering, and Asli Özyürek. 2020. The communicative advantage: How kinematic signaling supports semantic comprehension. *Psychological research*, 84(7):1897–1911.

Martha E Tyrone and Claude E Mauk. 2010. Sign lowering and phonetic reduction in american sign language. *Journal of Phonetics*, 38(2):317–328.

Martha E Tyrone and Claude E Mauk. 2012. Phonetic reduction and variation in american sign language: A quantitative study of sign lowering. *Laboratory phonology*, 3(2):425.

Unity Technologies. 2023. Unity. Game development platform.

Kayo Yin, Terry Regier, and Dan Klein. 2024. American Sign Language handshapes reflect pressures for communicative efficiency. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15715–15724, Bangkok, Thailand. Association for Computational Linguistics.

Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*.

Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, and Houqiang Li. 2021. Improving sign language translation with monolingual data by sign back-translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1316–1325.

George Kingsley Zipf. 2016. *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio books.

7. Language Resource References

Hochgesang, J. 2022. *SLAASh ID glossing principles, ASL Signbank and annotation conventions*. PID https://doi.org/10.6084/m9.figshare.12003732.v4.

Jiang, Zifan and Sant, Gerard and Moryossef, Amit and Müller, Mathias and Sennrich, Rico and Ebling, Sarah. 2024. *SignCLIP: Connecting Text and Sign Language by Contrastive Learning*. Association for Computational Linguistics. PID https://aclanthology.org/2024.emnlp-main.518/.

Max Planck Institute for Psycholinguistics, The Language Archive. 2025. *ELAN (Version 7.0) [Computer software]*. Max Planck Institute for Psycholinguistics, The Language Archive. PID https://archive.mpi.nl/tla/elan.

Varol, Gül and Momeni, Liliane and Albanie, Samuel and Afouras, Triantafyllos and Zisserman, Andrew. 2021. *Read and Attend: Temporal Localisation in Sign Language Videos*. PID https://arxiv.org/abs/2103.16481.

Wooten, Alicia and Spieker, Barbara. 2022. *Home*. PID https://atomichands.com/.

Yin, Kayo and Singh, Chinmay and Minakov, Fyodor O and Milan, Vanessa and Daumé Iii, Hal and Zhang, Cyril and Lu, Alex Xijie and Bragg, Danielle. 2024. *ASL STEM Wiki: Dataset and Benchmark for Interpreting STEM Articles*. Association for Computational Linguistics. PID https://aclanthology.org/2024.emnlp-main.801/.

A. Full List of STEM Signs

This appendix summarizes the STEM vocabulary used in both the isolated vocabulary condition described in § 3.1. Table 3 reports the unique STEM terms included in the dataset and the number of distinct signed variants observed for each term.

STEM Term	# Distinct Signed Variants
System	3
Energy	2
Species	3
Theory	2
Cell	3
Plant	2
Carbon	2
Gas	3
Iron	2
Mass	2
Chemical	2
Element	2
Oxygen	5
Organism	2
Properties	3
Photosynthesis	4
Chloroplast	1
Chlorophyll	1
Carbon Dioxide	2
Glucose	2
Autotrophs	1
Chemical Equation	4
Oxidation	2
Reduction	2
Cell Respiration	2
Mitosis	5
Meiosis	1
DNA	1
DNA Replication	1
DNA Transcription	1
DNA Translation	2

Table 3: Unique STEM terms and the number of distinct signed variants observed in the dataset. Variants include fingerspelled forms and annotated phonological alternatives.

B. Keypoint Mapping

To ensure comparability between the 3D motion capture data and 2D MediaPipe outputs, we manually defined a correspondence between the motion capture joint names and MediaPipe's landmark indices. Table 4 lists the specific MediaPipe keypoints corresponding to each motion capture joint for both the right and left sides of the body.

Figure 5: Example frames from avatar rendered motion capture recordings.

C. Avatar Examples

Figure 5 shows example frames from the avatar rendered versions of the motion capture recordings used in our experiments. Using Unity, we rendered 3D humanoid avatars driven by the recorded 3D joint trajectories. Avatar rendered videos were used exclusively for evaluating pretrained sign language models (§ 3.4), in order to assess whether reducing out-of-distribution visual noise improves model performance.

D. Relative Change Analysis - Student

This section reports the relative change analysis for the student signer, conducted using the same method described in § 4.2. For each joint, we computed Spearman rank correlations between mention index and the percentage change in spatial extent, path length, and velocity across repeated mentions of the same sign. Positive values indicate articulatory reduction (i.e., smaller movements or increased efficiency) with repetition.

As shown in Table 5, compared to the instructor, the student produced fewer repeated signs, which limited the statistical power of within-sign analyses. None of the correlations reached significance. This suggests no consistent pattern of reduction across repetitions for the student signer in the dialogue condition.

	Joint Name	Mediapipe Keypoint
Right	RightArm	pose_12
	RightForeArm	pose_14
	RightHand	pose_16
	RightHandMiddle1	right_hand_9
	RightHandMiddle2	right_hand_10
	RightHandMiddle3	right_hand_11
	RightHandMiddle4	right_hand_12
	RightHandRing	right_hand_13
	RightHandRing1	right_hand_14
	RightHandRing2	right_hand_15
	RightHandRing4	right_hand_16
	RightHandPinky	right_hand_17
	RightHandPinky1	right_hand_18
	RightHandPinky2	right_hand_19
	RightHandPinky4	right_hand_20
	RightHandIndex	right_hand_5
	RightHandIndex1	right_hand_6
	RightHandIndex2	right_hand_7
	RightHandIndex4	right_hand_8
	RightHandThumb1	right_hand_1
	RightHandThumb2	right_hand_2
	RightHandThumb3	right_hand_3
	RightHandThumb4	right_hand_4
Left	LeftArm	pose_11
	LeftForeArm	pose_13
	LeftHand	pose_15
	LeftHandMiddle1	left_hand_9
	LeftHandMiddle2	left_hand_10
	LeftHandMiddle3	left_hand_11
	LeftHandMiddle4	left_hand_12
	LeftHandRing	left_hand_13
	LeftHandRing1	left_hand_14
	LeftHandRing2	left_hand_15
	LeftHandRing4	left_hand_16
	LeftHandPinky	left_hand_17
	LeftHandPinky1	left_hand_18
	LeftHandPinky2	left_hand_19
	LeftHandPinky4	left_hand_20
	LeftHandIndex	left_hand_5
	LeftHandIndex1	left_hand_6
	LeftHandIndex2	left_hand_7
	LeftHandIndex4	left_hand_8
	LeftHandThumb1	left_hand_1
	LeftHandThumb2	left_hand_2
	LeftHandThumb3	left_hand_3
	LeftHandThumb4	left_hand_4

Table 4: Mapping between motion capture joints and Mediapipe keypoints.

Joint	Spatial	Path	Velocity
Fingers (L)	$+0.231$	$+0.137$	$+0.171$
Fingers (R)	$+0.334$	$+0.352$	-0.297
Hand (L)	$+0.309$	$+0.309$	-0.160
Hand (R)	$+0.302$	$+0.322$	-0.290
Forearm (L)	$+0.137$	-0.005	$+0.265$
Forearm (R)	-0.071	-0.154	-0.070
Arm (L)	$+0.188$	$+0.147$	$+0.114$
Arm (R)	-0.030	-0.233	$+0.114$

Table 5: Relative Change Analysis (Spearman correlation). Stars denote significance (* $p < .05$, ** $p < .01$, *** $p < .001$). None of the correlations reached significance.