Understanding Challenges and Opportunities of Technology-Supported Sign Language Learning

SARAH FALTAOUS, University of Duisburg-Essen, Germany
TORBEN WINKLER, University of Duisburg-Essen, Germany
CHRISTINA SCHNEEGASS, LMU, Germany
UWE GRUENEFELD, University of Duisburg-Essen, Germany
STEFAN SCHNEEGASS, University of Duisburg-Essen, Germany

(a) (b) (c) (d)

Fig. 1. We investigated the learning effect of four different learning conditions to teach sign languages. In the example pairs of conditions and signs shown, the participant receives (a) audio instruction for the sign *strong*, (b) visual instruction for the sign *past*, (c) electrical muscle stimulation for the sign *have*, and (d) a combination of both visual and EMS for the sign *devour*.

Around 466 million people in the world live with hearing loss, with many benefiting from sign language as a mean of communication. Through advancements in technology-supported learning, autodidactic acquisition of sign languages, e.g., American Sign Language (ASL), has become possible. However, little is known about the best practices for teaching signs using technology. This work investigates the use of different conditions for teaching ASL signs: audio, visual, electrical muscle stimulation (EMS), and visual combined with EMS. In a user study, we compare participants' accuracy in executing signs, recall ability after a two-week break, and user experience. Our results show that the conditions involving EMS resulted in the best overall user experience. Moreover, ten ASL experts rated the signs performed with visual and EMS combined highest. We conclude our work with the potentials and drawbacks of each condition and present implications that will benefit the design of future learning systems.

CCS Concepts: • **Human-centered computing** → **Interaction techniques**.

Additional Key Words and Phrases: Sign language learning, electrical muscle stimulation, audio, visual.

ACM Reference Format:
Sarah Faltaous, Torben Winkler, Christina Schneegass, Uwe Gruenefeld, and Stefan Schneegass. 2022. Understanding Challenges and Opportunities of Technology-Supported Sign Language Learning. In *AH '22: Augmented Humans International Conference, March 13–15, 2022, Kashiwanoha, Chiba, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3519391.3519396

1 INTRODUCTION

With almost half a billion people living with hearing loss [32], sign languages have become a widely used alternative to verbal communication, empowering many people with hearing loss to better communicate [30]. Moreover, learning sign language can benefit everyone: it can be a solution for hearing people if verbal communication is not feasible (e.g., it offers an opportunity to communicate in noisy environments), and, more importantly, it makes communication between the hearing and deaf or hard-of-hearing (DHH) members of a community easier. Currently, sign language learning is often done either with the support of a teacher who provides in-person training or with subbed video tutorials [25]. However, autodidactic learning of sign language via video remains challenging because signs are complex. They consist of hand and arm movements, facial expressions, and body language [37], which are difficult to perform correctly using only videos. Hence, learning sign language without in-person training (e.g., over distance) remains challenging.

Previous research has demonstrated that technology has the potential to support the process of teaching sign language (e.g., through augmented reality [4] or with the help of robotics [48]). However, despite the promising results from studies demonstrating that electrical muscle stimulation (EMS) can support the execution of various muscle skills (e.g., [8, 35]), the feasibility of the technology to support sign language learning has not been investigated. Yet it remains interesting, if the EMS technology can effectively support learning muscle skills that have the complexity of signs. Fundamentally, EMS works by inducing an external electrical signal, which is sent to the human muscle; the human muscle then contracts, leading to a movement that corresponds to the actuated muscle. In this paper, we explore the potential of this technology to support sign language learning. To further investigate the suitability of the traditional language learning approaches, namely *audio* and *visual*, we conducted a comparative study in which *EMS* is compared to both *audio* and *visual* modalities. Additionally, we consider the combination of *EMS* and *visual* for our comparison as well. By comparing these technologies in terms of their effectiveness for learning sign language, the findings could contribute to making autodidactic learning more efficient.

Sign language, like any spoken language, is composed of many elements. Although many of them are common to spoken languages (e.g., grammar, vocabulary, structure, etc), other elements are more focused on sign execution (e.g., hand movement, fingerspelling, facial expressions). Our main goal is to explore the role that technology plays, through different modalities, in supporting sign language learning and execution. In this work, we focus on exploring signs that mainly rely on movement from the user's arms and hands, with fewer complex details added (e.g., facial expressions). This is defined as the first out of four strata in Sandler's categorization [37]. To be exact, we investigate the effect of using different interaction modalities for the autodidactic acquisition of sign language, namely *audio*, *visual*, *EMS*, and a *combined* condition consisting of both *visual* and *EMS*. As we do not only focus on deaf or hard of hearing (DHH) users, but rather address users who are interested in learning sign languages (e.g., family and friends of DHH people), we included the *audio* modality in our studies. While the *audio* and *visual* modalities communicate instructions for performing a sign to the user, *EMS* allows direct actuation of the user. Thus, it involves performance based on direct feedback from a trainer (e.g., training device).

In a user study, we asked participants to perform four different signs instructed through the three modalities (*audio*, *visual*, *EMS*) as well as the combination of *visual* and *EMS*. After a two-week break, participants were asked to perform the signs again, this time without instructions. We were primarily interested in the correctness of sign execution, recall ability, and user experience for each modality type. During the evaluation process, we received support from ten ASL experts (with 22 years of experience on average) to judge the quality of the signs' execution. Our results show a significant difference between the *combined* and *audio* conditions, where the execution of the signs in the *combined* condition was better. In general, conditions involving *EMS* received better ratings in terms of user experience.

Contribution. Our work contributes to (1) a better understanding of the challenges and potential for technology-based sign language learning. Moreover, we (2) provide insights from a user study (N=17) in which we compare the learning effects of different instruction modalities on the example of American Sign Language including the rating from 10 ASL experts evaluating the performance of the participants.

2 SIGN LANGUAGE LEARNING

Sign languages are a means of communication benefiting a large percentage of the world's population, including not only those living with hearing or speech impairments, but also many people in their communities, such as family members, friends, and teachers [7]. Sign languages are more than a system of communication for an existing language; they are "a true human language" ([46], p.7). However, there is no universal sign language, but rather a great variety of sign languages across the globe. As digitalization has progressed over the last 20 years, several technological approaches have been brought forward to support communication between the hearing and people with hearing or speech impairments.

2.1 Signs vs. Gestures

Until the end of the 19th century, the words *gesture and signs* were used interchangeably [22, 23]. In the 20th century, however, signs started to be considered linguistic [23, 24, 41, 42]. Previous research plotted the evolution from simple gestures to sign languages [21, 29], which was later also explained as transferring from no convention and speech (gestures) to convention and no speech (sign languages) [19, 26]. It further highlighted that people who are deaf or hard-of-hearing perceived gestures as a method of communication with hearing people who cannot sign [26]. Another perspective suggests that gestures could be seen as a natural feature that accompanies a language regardless of its modality (sign or speech) [12]. In this paper, we will use the term sign instead of gesture to avoid confusion.

Sandler divided an existing sign language into four strata of increasing complexity in terms of gestures and grammar. The first stratum uses only a hand to communicate, while the fourth stratum uses the hands, head, face, and body [37]. In this work, we focus mainly on the first stratum and we use American Sign language (ASL). Thus, as a first step, we explore the role of different technologies in supporting the hands and arms in executing signs. While starting with rather simple signs is not recommended for machine-based sign language recognition, generation, or translation [3], they can provide a good initial set to investigate human-centered sign language learning as they reduce the number of influencing factors and make the results easier to interpret. Nevertheless, this choice certainly influence the generalizability of the results, and thus, future exploration of more complex signs is required.

2.2 Technological Support for Sign Languages

Technology can support communication between people with and without hearing and speech impairments in multiple ways. The main goal of previously-studied assistive technologies was to develop a medium for post-hoc or instant translation, usually from ASL to verbal or written English. The basis for the automated translation is the recognition of the signs. Examples of applying technology to sign recognition include using RGB [2] or depth cameras [27]. Further, sensing tools such as data gloves [11, 20] provide more detail about the signer's hand positions. To make the process of hand sign recognition less cumbersome, researchers applied sign recognition through Electromyography. For example, Abreu et al. [1] and Paudyal [33] used a consumer device for hand movement tracking called the Myo armband[1]. This armband uses Electromyography to sense the electrical activity in the muscles of the arm that is generated by hand and finger movement.

In their recent work, Gugenheimer et al. [14] evaluated existing assistive technology in this domain and requested a change in the design perspective. They highlighted the importance of supporting learning of the subordinate rather than the dominant language in a society, identifying sign language as a subordinate language based on the number of signers compared to the speaking community. Hence, we take the stand that technological support should be designed to foster the learning of sign languages by people without hearing or speaking impairment.

ASL is a complex language and is difficult to acquire without the help of a proficient teacher. However, technology can be used to enable autodidactic learning anytime and anywhere, appealing to a broader audience.

2.3 Teaching of Sign Languages

Besides real-life courses teaching sign language, digital tutorials have become increasingly popular over recent decades. In these mostly video-based lessons, the learner watches a recording of a person or avatar executing a certain sign or sentence and is asked to repeat it. This form of digital video-based learning enables self-paced and remote learning without an actual human teacher. To assess the quality of the learner's sign execution and provide feedback, the system has to recognize the learner's movements.

There are several ASL learning applications available on the market (e.g., ASL Coach[2] or American Sign Language ASL[3]). These apps offer a variety of video tutorials but lack performance recognition; thus, they also lack feedback on the accuracy of a learner's sign execution. In recent studies, Paudyal et al. [34] highlighted the need for corrective feedback and investigated a combination of a camera-based sign recognition system with an intelligent tutoring system to teach ASL. By analyzing a learner's joint locations, hand and arm movements, and hand shape, the system compares the execution to an expandable database of target signs and provides feedback [34]. However, the effect of the feedback on the learning process has not yet been evaluated.

2.4 Research Gap

The most common technical approach to sign language teaching is video-based learning, ie. the learner is asked to repeat a sign performed by a human or avatar in a video. This approach is applied to mobile applications, online tutorials, and intelligent systems such as SCEPTRE [33] or Learn2Sign [34]. Yet, video-based learning has the drawback that it requires the learner to accurately mimic the signs performed by the teacher on their own and be aware of mistakes. Further, it remains unclear how well video-based learning actually performs in terms of recall compared to other modalities.

3 TEACHING MODALITIES FOR SIGN LANGUAGE LEARNING

The Cognitive Load Theory describes the allocation of working memory resources in learning processes. It defines the terms *Intrinsic Load* (load induced by the task itself), *Extraneous Load* (load induced by the instructional design), and *Germane Load* (resources needed for schema construction and memory integration, and thus, learning). Confusing instructions can use up cognitive resources that would otherwise be available, and can thus hinder learning [5]. Thus, one goal is to design the instruction in a way that it does not induce more additional load than necessary.

[1] Myo armband. https://support.bynorth.com/myo, last retrieved March 9, 2022.
[2] ASL Coach. https://play.google.com/store/apps/details?id=com.PLMUN.myASL, last retrieved March 9, 2022.
[3] Sign Language ASL. http://play.google.com/store/apps/details?id=tenmb.asl.americansignlanguagepro, last retrieved March 9, 2022.

In the following, we will outline different modalities for sign language learning instruction, which we will comparatively evaluate in a user study. Our teaching modalities include *audio* as we include people without hearing or speech impairment, representing relatives that use sign language as a mean of communication with their deaf or hard-of-hearing friends and family members. We propose investigating the performance of the frequently-used (1) video-based instructions in comparison to three other instruction modalities: (2) audio instructions, (3) direct muscle actuation using EMS, and (4) a combination of visual and EMS. We decided to include *video* and *audio* as separate conditions to better understand their individual influence on the learning effectiveness. As previously mentioned in the related work, signs could be divided into four strata [37]. In our work, we focus mainly on signs using the hands (i.e. first stratum). We designed the instructions by observing professional online teaching videos (both front and side views) and cutting down each sign to a sequence of movements.

3.1 Audio

In our case, the audio modality includes a description of the movement. Describing a movement using words offers the chance to include specific details regarding its execution, which may not be obvious in a picture or video. Furthermore, the sign can consist of multiple smaller movements of different body parts (e.g., a hand posture combined with a head tilt or facial expression). In the audio format, the individual signs are outlined successively, while their order of execution is described verbally. As a result, the learner can focus on each individual movement required to execute a sign instead of perceiving all parts simultaneously, leading to a more profound way of processing.

3.2 Video

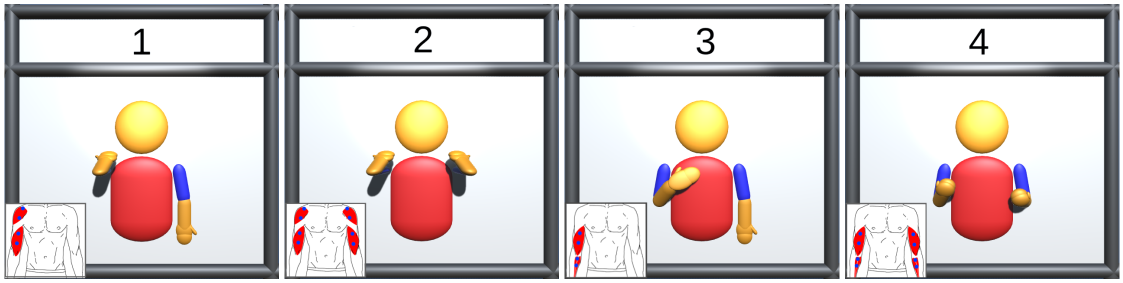
As described in Section 2, many sign language learning applications employ videos of either an avatar or a human demonstrating signs. This technique is founded on the assumption that observing a movement triggers similar processes in the human brain as does the actual movement execution [40]. Therefore, the combination of visual presentation and subsequent practice has the potential to create lasting motor memories.

3.3 EMS

Electrical Muscle Stimulation (EMS) is a method of externally stimulating the human body's muscles using small electrodes attached to the skin that send electrical impulses. Through EMS, a muscle can be actuated and, depending on the impulse's strength, can create a sense of force-feedback or lead to movement execution [38]. In human-computer interaction, EMS has been investigated as a means to provide realistic force-feedback in virtual reality, but also to train certain movements. For example, Hassan et al. [15] applied EMS as a teaching tool to help runners improve their fine-tuned movement execution while running. They were able to show that EMS outperformed slow-motion video-based feedback on their movements. Other researchers have focused more on hand movements, e.g., to improve typing skills [39], bowling skills [44], or the playing of musical instruments [17, 43]. For teaching sign language, we actuate the learners' muscles using EMS to provide them with the correct movement initially, reducing the need for corrective feedback.

4 EVALUATION

This work focuses on teaching American Sign Language (ASL) signs, as it is one of the most widely used sign languages in the world. Although the actual number of ASL signers is hard to specify, the estimates range between 250-500,000 signers in the US alone [31]. Here, we focus on teaching signs that primarily rely on hand and arm movements.



(a) Sign past. (b) Sign devour. (c) Sign have. (d) Sign strong.

Fig. 2. Visual representation and electrode placement for each sign. The audio descriptions for each are (a) *throw your right hand over your right shoulder*, (b) *throw both hands over your shoulders*, (c) *tap your chest with the fingers of your right hand*, and (d) *make a fist with your hands and bend both arms, bend the right arm more than the left one.*

Since our main goal is to compare learning of signs across the different conditions, we avoided any kind of linking between the word meanings and the sign itself (i.e., iconicity [45]). Consequently, we presented the executed signs by number for memorability (e.g., sign 1). There is a huge pool of signs performed using the hands in American Sign Language (ASL). Given the current limitations of EMS (e.g., limited number of simultaneously actuated muscles or precision of actuation) and in favor of a simpler study design with less influencing factors, we chose signs that would ensure a fair comparison across all conditions. Thus, we chose four signs: two that require movement of both arms and two that require only one arm to move. We communicated these signs in four instruction modalities. The first is *audio*, where the user hears a description of what is being communicated. The second modality is a *visual* representation that uses a video of an avatar performing the sign. The third modality is actuating the user via *EMS* to induce performance of one of the signs. The fourth and last modality is a *combined* instruction, where the user can see the visual input and be simultaneously actuated via EMS. Any combination with audio instruction was excluded because the user has to hear the whole description, process it, and then execute it. This is unlike visual and EMS instructions, where the user starts to execute the sign the moment the instructions are communicated.

4.1 Initial Involvement of Signers

We asked a sign language student (20 years, female) and a teacher (60 years, male) for their opinion regarding the design of the different modalities and collected their feedback. For the *EMS* condition, particularly the student was positive by indicating that the *"exact movement could be controlled."* She expected the *audio* condition to be appropriate to communicate the correct arm movement. However, she was afraid that the audio could be easily misinterpreted. The teacher, on the other hand, expected that the *visual* condition would support the learning process best as he has been using it for his 25 year long career. He was afraid that without visuals, learning will not be successful. Both agreed that the combination of visual and EMS can be beneficial.

4.2 Study Design

To avoid that participants see any meaning in the signs that would help to remember them, we assigned each sign a number (cf., Figure 2). In general, we had three modalities to communicate the teaching instructions: *audio*, *visual*, and *EMS*. In addition to these, we had a condition that we refer to as *combined*, which communicated the signs visually and with EMS simultaneously. Each sign was repeated 10 times in each condition. For both the *EMS* and *visual* conditions as well as the *combined* condition, the instructions were communicated for 3*sec*, with 3*sec* intervals between repetitions. However, for the *audio* condition, the communicated description lasted 6*sec* instead with a 3*sec* interval between repetitions. The audio instruction lasted longer to ensure that the participants had enough time to process the meaning and execute the movement in 3*sec*. This asynchronicity concerning the different times for *visual* and *audio* was one of the major reason to not provide it as a combined condition but rather to rely on a combination of *visual* and *EMS*. To be able to record and later evaluate their performance in recalling as well as executing the signs, we recorded the movements via video documentation (mounted from the perspective of a communication partner) and an OptiTrack system[4] (to allow precise replay of the movements from different perspectives). Our approach would help us to identify the best memorized and correctly executed condition after the study.

4.3 Participants and Procedure

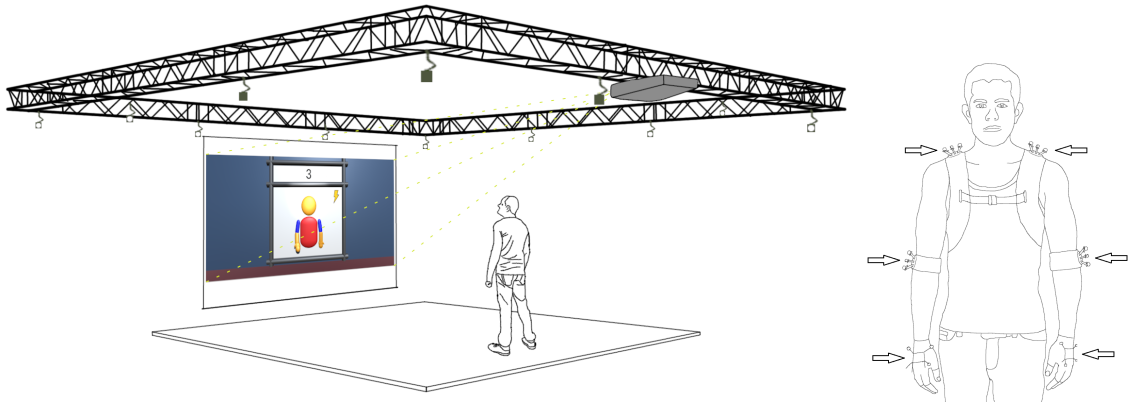
Overall, we had 17 participants (13 male, 4 female) with no prior knowledge of any sign language, aged between 22 and 32 years (mean=27.56, SD=2.98). The quantitative data of one participant was excluded as the actuation in the EMS condition yielded a different behavior after being actuated only once (i.e., the participant's forearms twitched toward the abdominal area instead of raising toward the shoulders). The study was conducted in two sessions, each lasting for about an hour, with a span of two weeks in between, as was done in previous work [13]. Each participant had to perform every sign in a counterbalanced order. Furthermore, the signs communicated in each modality were counterbalanced to eliminate any learning effect.

First Session. At the beginning of the first session, we welcomed the participants and gave them the consent form, which included the safety regulations of the EMS as well as the study description. The study was conducted following the ethical guidelines of our institution. The first session consisted of two parts. The first part was the muscle calibration, in which we actuated the muscles of the participants and visually checked until the target motion was achieved similar to previous work [9, 36]. To be exact, we segmented each sign into a sequence of movements and after placing the electrodes we visually verified if all the segments were applied. In the second part of the first session, we asked the participants to wear OptiTrack markers in the form of arm bands, a vest, and gloves (cf., Figure 3b). We recorded images of the positioning of the arm bands to ensure applying the same positioning in the second session. The participants stood in the lab facing a projection, through which we communicated the sign number (cf., Figure 3a).

Then, we communicated the different instruction modalities and signs. The mix of the modalities and the signs followed a Latin-square design, where each condition appeared four times in our overall data set. Between each condition, we asked the participant to fill in a user experience questionnaire [28]. At the end of the session, we asked the participant to rate each condition on a 7-point Likert-item, indicating to what extent the sign was clearly communicated and to what extent they would remember it, as well as their learning abilities (e.g., ability to memorize and learn).

Second Session. After two weeks, we asked the participants to come into our lab again for the second session. First, we asked them to indicate on a 7-point Likert-item to what extent they remembered the signs they had learned in the first session. Then, we asked them to put on the OptiTrack markers and adjusted the markers' positions and orientations according to the images captured in the first session. We again presented the sign number known from the first session and requested them to execute each sign as soon as they saw the sign number. They were told that they should perform each sign as they remembered it from the first session. We then visually confirmed whether it was the correct sign. If the participant said that they had forgotten the sign or performed a wrong one, we provided a hint. We had a total of

[4]OptiTrack. https://optitrack.com, last retrieved March 9, 2022.



(a) Apparatus used for our evaluation.

(b) Body-mounted trackers.

Fig. 3. The general setup of our study, where the participant stands facing the projection showing the different instructions (cf., 3a). The participants' movements are recorded by a tracking system that tracks the participant's markers (cf., 3b).

three hints that would be communicated in the same order for each condition and across all participants. Each hint would only be communicated if the previous one failed to remind the participant of the sign. In order, the three hints provided were: (1) the sign modality, (2) the number of arms used, and (3) the visual representation of the sign.

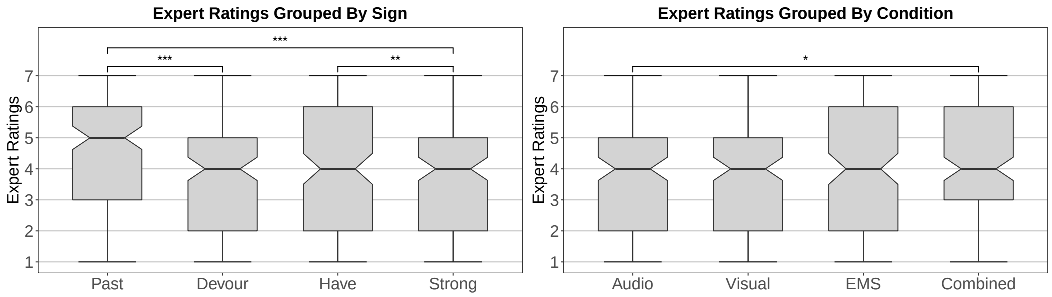
When the participant remembered the sign, they had to rate on a 7-point Likert-item to what extent they remembered the exact execution of the sign. They then had to repeat the sign 10 times, which was guided by signals (i.e., the sign number appearing on the screen). The signals appeared for 1.5*sec* and were separated by 3*sec* breaks. At the end of the session, we conducted an interview with the participant to gain more insight into how each modality affected their ability to remember the signs.

5 RESULTS

Since the *EMS* actuation for one participant caused problems for the last trials of the second session, we excluded that participant for our performance measures but considered their feedback for the semi-structured interviews. Below, we report mean (M), median (Md), and interquartile-range (IQR).

5.1 Ratings from ASL Experts

To understand how interpretable participants performed the learned signs during the second session of our study, we invited ten ASL experts (6 female, 3 male, 1 preferred not to say), aged between 28 and 54 (M=35.7, SD=8.2) with an average experience in ASL of 22.2 years (SD=14.6) to participate in an online questionnaire rating how well participants performed each sign. All Participants are residents in the USA. In particular, the questionnaire showed the video recordings of the performed signs (from the perspective of a conversation partner) and the intended sign. Then, we asked the experts to rate the statement "the participant correctly executed the arm and hand movements of the sign" (1=strongly disagree and 7=strongly agree). As we are inspecting the movements of only the arms, all the faces of the participants were blurred to eliminate any influence of facial expressions. Overall, each one of the ten ASL experts rated the signs of the 16 participants that performed all four signs each, resulting in 640 ratings (i.e., 10 experts * 16 participants * 4 signs).



(a) Boxplots for the ratings of the four different signs. (b) Boxplots for the ratings of the four different conditions.

Fig. 4. Ratings from the ASL experts on the signs performed by the participants at the second session of our user study. The significance levels are given by stars: *(<0.05), **(<0.01), and ***(<0.001).

For the different signs, the mean (median, interquartile-range) ratings of participants' sign execution rated by the experts are (in descending order): *past*=4.53 (Md=5, IQR=3), *have*=4.11 (Md=4, IQR=2), *devour*=3.68 (Md=4, IQR=3), and *strong*=3.51 (Md=4, IQR=3) (see Figure 4a). Since we do not assume normality, we performed a Friedman test that showed a significant effect of the signs on the expert ratings (χ^2(3)=25.61, p<0.001, N=10). A post-hoc test using Wilcoxon Signed-rank with Bonferroni correction revealed significant differences between the conditions. We found a significant difference between *past* and *devour* (W=15950, Z=4.394, p<0.001, r=0.25), *past* and *strong* (W=16675, Z=4.665, p<0.001, r=0.26), and *have* and *strong* (W=14938, Z=3.268, p=0.006, r=0.18). Here, we can conclude that *past* is rated significantly better than *devour* and *strong*, and *have* is rated better than *strong*.

For the different conditions, the mean (median, interquartile-range) ratings of participants' sign execution rated by the experts are (in descending order): *combined*=4.20 (Md=4, IQR=3), *EMS*=4.06 (Md=4, IQR=4), *visual*=3.87 (Md=4, IQR=3), and *audio*=3.7 (Md=4, IQR=3) (see Figure 4b). Since we do not assume normality, we performed a Friedman test that revealed a significant effect of the conditions on the expert ratings (χ^2(3)=9.338, p=0.025, N=10). A post-hoc test using Wilcoxon Signed-rank with Bonferroni correction showed a significant difference between *audio* and *combined* (W=10992, Z=$-$2.794, p=0.030, r=0.16). We conclude that *combined* is rated significantly higher than *audio*.

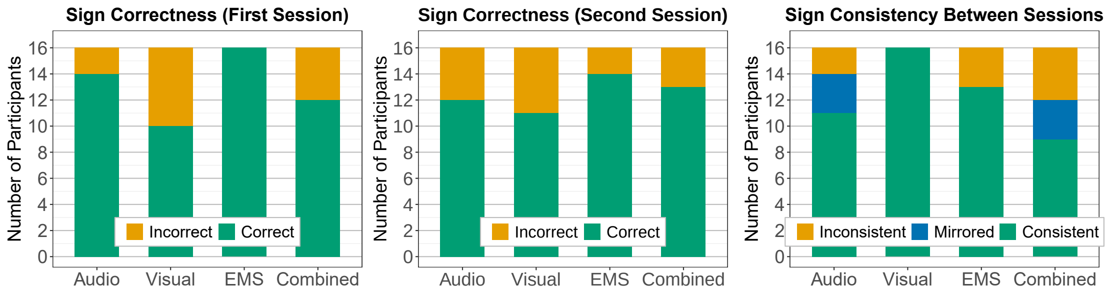
Moreover, after rating the sign execution of our study participants, experts provided general comments about the rated performances. Experts noted that some signs came off a bit aggressive. In particular, because the signs were executed too quickly. Here, the timing for some signs (*devour*, *strong*) is more important than for others (*past*, *have*). Furthermore, experts highlighted that facial expressions and finger spells are important for correct execution of signs.

5.2 Memorability

We report the number of hints needed as well as the correctness of the sign execution within and across sessions.

Number of Hints. At the beginning of the second session, we asked the participant to start executing the sign as soon as we displayed the sign number. If participants had trouble remembering a sign, they could get up to three hints (one at a time). In the following, we report the number of hints needed per condition (in ascending order): *audio*=12 (Md=0, IQR=1.25), *combined*=15 (Md=1, IQR=1.25), *EMS*=19 (Md=1, IQR=2.25), and *visual*=20 (Md=1, IQR=3).

Since the total number of hints is strongly influenced by individual participants taking multiple hints for a particular condition, we looked at the binary decision if hints were provided or not (i.e., we used a binary code: 1, if we gave any



(a) Correctness in first session. (b) Correctness in second session. (c) Consistency between sessions.

Fig. 5. The correctness and consistency of the signs in each condition. In the first two subfigures, we compared the signs executed in each session to the intended ones (i.e., correctly executed). In the *sign correctness* figure, we explored the learning effect by comparing the signs executed in the first and second sessions. For consistency, we assume that signs were correctly executed in the first session.

hints, and 2, if no hints were given). The results indicate that equal numbers of participants needed hints for *audio* and *combined* ($N = 4$), followed by *EMS* ($N = 5$) and *visual* ($N = 6$). A Friedman test that showed no statistical difference across the four different modalities (χ^2(3)=2.18, p=0.534, N=16).

Sign Correctness and Consistency. We further explored how well the signs were executed within and across sessions. We based our evaluation on the video recordings of the two sessions.

In our analysis of the video recordings, we judged the correctness of the sign execution in each of the two sessions. We did this by visually comparing the videos to the intended sign (cf., sign correctness in Figure 5). We focused on the used arm, the direction of the hand motion, and the final sign execution. Furthermore, we then carried out a third comparison to explore the learning effect (cf., sign consistency in Figure 5). In this, we compared the sign execution in the second session with that of the first one. Here, we ignore a wrong sign execution in the first session and simply assume participants practiced the correct sign to deduce to what extent they remembered the signs learned in the first session. To do so, we categorized each compared pair into *consistent, inconsistent,* or *mirrored* based on whether the second session sign was the same, was completely different, or involved the wrong arm(s), respectively.

For the first session, *EMS* was the best in terms of correctness (N=16). *Audio* scored the second best, as 14 participants performed the intended motion correctly. This was followed by the *combined* condition (N=12) and finally the *visual* condition (N=10).

In our second comparison, which evaluated the correctness of the second session performance, *EMS* was again the best (N=14). This was followed by *combined* (N=13), *audio* (N=12), and then *visual* (N=11). There was no discernable pattern among the signs that were wrongly executed.

In the third comparison, the sign consistency was measured by comparing each participant's performance in the second session to their performance in the first one. The sign consistency was best observed in the *visual* condition, in which all the participants, upon remembering the sign, executed it the same way as they did it in the first session (N=16). The second-best was *EMS* (N=13), followed by *audio* (N=11), and finally the *combined* condition (N=9). *Audio* and *combined* resulted in two and three participants performing the sign mirrored, respectively.

5.3 User Experience Questionnaire

Based on the results of the User Experience Questionnaire (UEQ) [28], the overall user experience is equal for both *combined* (Md=1.5, IQR=1) and *EMS* (Md=1.5 , IQR=1.1), followed by *visual* (Md=0.3, IQR=1.06) and *audio* (Md=−0.31, IQR=1). Statistical analysis using a Friedman test revealed significant differences between the four conditions (χ^2(3)=36.38, p<0.0001, N=16). A Wilcoxon-Pratt Signed-rank test showed that there is a significant difference between *audio* and *EMS* (W=0, Z=−3.517, p<0.001, r=0.62), *visual* and *EMS* (W=0, Z=−2.600, p<0.01, r=0.46), *audio* and *combined* (W=3, Z=−3.518, p<0.001, r=0.62), and *visual* and *combined* (W=10, Z=−3.000, p<0.001, r=0.53). For the user experience, we can conclude that *EMS* and *combined* > *audio* and *visual*.

For the pragmatic quality *combined* was rated best ($Md = 1.7, IQR = 0.75$), followed by *visual* ($Md = 1.5, IQR = 0.6$), *EMS* ($Md = 1.25, IQR = 1.3$) and last but not least audio ($MD = 0.87, IQR = 1.5$). Statistical analysis using Friedman test, showed significant difference between the four modalities (χ^2(3)=11.8, p<0.01, N=16). A Wilcoxon-Pratt Signed-rank test shows that there is a significant difference between audio and combined ($Z = -2.8, p < 0.01, V = 7.5$) as well as *EMS* and *visual* ($Z = -2.3, p < 0.05, V = 19$).

For the hedonic quality the same rating as the overall was observed where *combined* was the best rated ($Md = 2.5, IQR = 0.9$) followed by the *EMS* ($Md = 2, IQR = 1.25$) then *visual* ($Md = -0.75, IQR = 1.25$) and at last the *audio* ($Md = -1.25, IQR = 0.88$). Statistical analysis using Friedman test, showed significant difference between the four modalities (χ^2(3)=32.7, p<0.0001, N=16). A Wilcoxon-Pratt Signed-rank test shows that there is a significant difference between audio and combined ($Z = -3.4, p < 0.001, V = 0$), ems and visual ($Z = 3.4, p < 0.0001, V = 120$) as well as combined and visual ($Z = 3.3, p < 0.001, V = 132$).

5.4 Individual Likert-items

We asked the participants to rate several Likert items across the two sessions. The results are presented in the following section. In the first session, we asked them to indicate for each condition the likelihood that they would remember the sign in two weeks. The results show that all the conditions were rated the same (Md=4). A Friedman test showed no significant differences across the four conditions (χ^2(3)=1.06, p=0.7, N=16). We also asked them to rate the extent to which extend the sign was clearly communicated. The participants rated *visual* highest (Md=7, IQR=1), followed by *combined* (Md=6.5, IQR=1), then *EMS* (Md=5.5, IQR=1.25) and then *audio* (Md=5, IQR=4.25). A Friedman test again showed no significant differences across all four conditions (χ^2(3)=7.4, p=0.06, N=16). During the second session, we asked them to rate to which extent they remembered the sign in each condition. The participants rated *combined* as best (Md=6, IQR=2), followed by *EMS* (Md=5, IQR=4) and *audio* (Md=5, IQR=4) in the same rank. The least-ranked signs were those communicated via *visual* instructions (Md=4, IQR=5). A Friedman test showed no significant differences across all four conditions (χ^2(3)=3.88, p=0.2, N=16). In general, the self-rating of memory skills was higher in the first session (MD=5, IQR=3), than in the second session (Md=4, IQR=2.25).

5.5 Semi-structured interviews

Finally, we report the general comments mentioned by the participants during the study along with the comments from the interviews. At the end of the second session, we asked our participants to provide further insights regarding their preferred modality and to share with us any general feedback.

Audio instructions. Out of the seventeen participants, eleven commented on the audio instructions. Nine of these expressed their confusion while receiving the instructions. They described it as *"the most confusing"* [P12] and the *"the* *most unclear"* [P8]. P9 further elaborated: *"I was confused about what the audio needed from me at the end. The thing is it might have meant that I raise my hand super high [...] audio was self-interpreted."* This was further supported by P7 and P13: *"[the] audio condition is not my favorite because of the uncertainty, it is intuitive [...] ground truth unclear, unclear which fingers and how the fingers need to be positioned on the chest and where to put the arm specifically"* and *"as usual you hear something you interpret according to your own understanding."* Two participants mentioned that it was the *"easiest"* to remember as *"the nature of the condition is different"* [P15] and *"it was not like the others [conditions]"* [P13]. P3 further elaborated that he *"had to think about the audio but none of the others."* This was further supported by P2, who mentioned that *"I was not sure if I was doing the right thing, maybe that is why it stayed because it was special".*

Visual instructions. Eight of our participants commented on the visual modality. All comments concerning the visual condition were positive. They described it as *"quite delivering"* [P8] and *"pretty clear"* [P9]. P12 added that *"visual has a better connection to the brain than audio."* P9 further elaborated that for *"everything without visual,"* he was unsure if he executed it correctly. P10 further confirmed that *"in the visual condition, I am pretty sure I did the same as the one displayed."* However, P13 disagreed with that, reporting *"I thought visual should be worst while learning, it just disappeared from my memory."*

EMS condition. Out of our 17 participants, 10 commented on the EMS, with 7 of these mentioning that they remembered it best. P6 said that his *"hands moved unintentionally."* P11 further elaborated that they *"remembered EMS the best; from the EMS one can just notice the feeling of being actuated and that was easier to notice."* This was further supported by P12, who mentioned that *"EMS is just the best because it was a stimulation that the one did not learn or know from daily life."* Six participants said that they might have changed the sign execution during the second session. P3 explained that he *"adjusted it because of cables in the first time that might have influenced the movement."* Others based their reasoning on the lack of noticeable details. P11 explained that *"especially in EMS, the height of arms was not obvious."* P9 further elaborated that *"EMS is an impulse that indicates the direction, but the specific movement (end point) is hard to learn."* Two participants preferred having a more *"natural"* feeling [P9, P10].

Combined Visual-EMS instructions. From our participants, 7 commented on the combined instructions, with 6 of them rating it positively. They described it as a *"good way to learn"* [P10], *"better than the others"* [P5] and *"hard to forget"* [P8]. P15 further explained that *"it was clearly communicated."* P8 elaborated, saying *"combined was the best. However, EMS was dominant in combined, like it was opposing what I was trying to do."* P13 provided further explanation: *"seeing and being actuated leads to not forgetting the sign [...] EMS and visual worked best because of the actuation, as I knew the correct movement because ems helped me to move and then I could double check with the visual".*

6 DISCUSSION

In this work, we explore the feasibility of various instruction modalities or a combination of them for sign language learning. For that purpose, we investigated the use of audio, visual, EMS, and a combination of EMS and visual instructions. Our findings highlight challenges of each modality and potential for future use.

Our findings are based on two main results: (1) the data that we collected from the participants, including the Likert-items, and semi-structured interviews, and (2) the results from the online questionnaire with ten ASL experts along with their additional feedback. The expert evaluation of the signs showed that signs executed by one hand (*past*, *have*) are better learnt than signs executed by two hands (*devour*, *strong*; see Figure 4a). This may indicate that the more body parts a sign involves, the harder it is to perform it correctly. Thus, our findings can probably be seen as an upper limit for technology-based sign language learning as they focus on a selection of rather simple signs. Moreover, for both the *have* and *past* signs, the facial expression and the speed of execution would depend on the context, which is not necessarily linked to certain facial expressions. While, the *strong* sign is completely dependent on the speed of execution as well as the facial expression. Also, the *devour* sign depends on the facial expression (e.g., open mouth) to complete the meaning. Our findings indicate that these aspects (i.e., speed of execution and facial expression) should be included in any similar future research in order to add further dimensions, thereby, allowing one to obtain more generalizable results.

6.1 Audio Instructions

One of our experts, who is not DHH, reported knowing two sign languages. One she learned for her family, the other for a friend. Throughout her long learning period (46 yrs), she never came into contact with auditory learning instructions. Nevertheless, since sign language is not only a language for the DHH, but also for whole communities (including friends, family, and people of interest), we believe that audio instructions can prove useful and should not be neglected right from the start.

In terms of the memorability assessment (i.e., number of hints required in the study), in the *audio* condition, participants needed the lowest number of hints. Furthermore, in the interviews, when the participants commented on the audio instructions, they often remembered specific parts of the sign description, such as *"fingers on chest"* [P13]. Participants mentioned that they *"had to think about the audio but none of the others"* [P3]. This indicates that their cognitive involvement was higher while learning the signs via *audio* instruction. This is also in line with the work of Chi and Wylie, who argued that more involvement with the learning material increases the learning performance [6].

However, participants also mentioned in the interviews that the *audio* instructions were the *"most unclear"* instructions [P8] and that they partly needed to *"self-interpret"*[P9] the meaning. These comments were further reflected in the user experience questionnaire rating and in the individual question of the first session (i.e., the clarity of the sign communication), where *audio* was perceived worse than the other conditions. Although we only used signs that did not require a lot of fine detail (i.e., no complex finger movement), the *audio* condition ended up inducing uncertainty and confusion about the exact execution of the sign in our participants. This is supported by the ASL experts evaluation in which *audio* received the lowest ratings and performed significantly worse compared to *combined*. Hence, we think that *audio* can be beneficial in combination with another modality but should not be used to learn sign language as an exclusive modality.

Implication 1: While our findings indicate that audio instructions can prove useful, they should not be used as a standalone instruction modality for teaching signs.

6.2 Visual Instructions

Our experts indicated that video instructions are one of the most common methods for teaching sign languages. In addition, previous work indicated that video-based learning can yield good recall effects [18]. However, our results indicate that it is the least memorable (i.e., the participants needed the maximum number of hints). One of the participants even stated that the *visual* instructions *"just disappeared from my memory"* [P13]. On the one hand, the participants had trouble remembering the shown signs. On the other hand, they executed them in a similar way as in the first session upon being presented with them (see Figure 5c).

Nevertheless, the *visual* instructions were mostly perceived positively by our participants, who described them as *"delivering"* [P8] and *"clear"* [P9]. They reinforced this impression through their high confidence in executing the right sign. One of the participants expressed his confidence in his execution accuracy by describing his performance as the *"same as the one displayed"* [P10]. Similarly, when asked about the clarity of sign communication, participants rated the *visual* condition highest. This observation is in line with previous work that showed that a video-aided approach to teaching a skill leads to a better performance than audio instruction alone [13].

Implication 2: Visual feedback introduces a certainty aspect for the sign execution, as it is unambiguous and easily perceived.

6.3 Electrical Muscle Stimulation Condition

The feeling of experiencing *EMS* was described by the participants as *"different"* [P8], *"unconventional"* [P2], and nothing that one knows from *"daily life"* [P12]. In the user experience questionnaire, it was rated the second highest and was significantly better than *audio* and *visual*. On the contrary, our participants highlighted that they might have executed the signs differently in this condition. One participant said that through *EMS*, one does not receive enough feedback about the sign details. As one participant commented, the sign is *"not obvious"* [P11]. This could be based on the different nature of **EMS**, as it does not necessarily involve human perception [10]. In other words, the conditions containing *audio* and *video*, **EMS** does not require the participant to process the information and then act upon it. Instead, it directly actuates the human to produce the targeted movement. Although the *EMS* condition was least accurate, it achieved the highest correctness in both sessions (see Figure 5a and 5b). That was further indicated by our expert evaluation and participants, as P9 explained that *"EMS is an impulse that indicates the direction"* and P11 stated that *"the feeling of being actuated is easier [...] noticed."* This is in line with previous research that showed the role of proprioception in improving motor skills [47] by storing, updating, and maintaining a *motor skill program* [16].

Implication 3: EMS is suitable for guiding sign execution because it enhances the learning experience by directly actuating the limbs. However, without any additional feedback, the user might not be certain about the sign execution details.

6.4 Combined Visual-EMS Condition

For the participants, the *combined* condition was the most positively perceived. They mentioned it was a *"good way to learn"*[P10], *"better than the others"*[P5], and *"hard to forget"*[P8]. P13 further elaborated that *"seeing and being actuated leads to not forgetting [...] I knew the correct movement because EMS helped me to move and then I could double-check with the visual."*

Furthermore, the signs learned using the *combined* condition were rated the best by our ASL experts and resulted in a better user experience than *audio* and *video*. Additionally, our results show that it was the second best after audio in terms of memorability, which was reflected by the number of hints needed.

Overall, this condition has the potential to combine the benefits of the *visual* and *EMS* conditions. Specifically, it provides feedback and clear instruction via the visual component and a feeling of guidance via the EMS component. On the one side, visual instructions add a high level of certainty as it provides visual cues for how to execute the overall sign. On the other side, the EMS feedback reinforces the visual cue by initiating and stepwise guiding the execution.

Implication 4: Using EMS and visual instruction together provides a good user experience. Moreover, using EMS actuation along with visual instruction might help to prevent errors that can happen if only EMS is used.

7 LIMITATIONS AND FUTURE WORK

We acknowledge the following limitations in our work. Learning ASL requires one to learn hundreds of signs and to perform them in sequence, which takes a long time, and therefore, requires users to stay motivated. In our study, however, we focused on only four signs to be remembered for two weeks. Thus, the generalizability of our results towards sign languages in general remains uncertain. However, our findings can be understood as an upper performance limit because with more signs and higher complexity better learning results are unlikely. Furthermore, we mainly use signs as a part of sign language learning. We do not explore further aspects of these languages, such as hand shapes and facial expressions, which take additional time to learn, and therefore, require users to maintain motivation.

In addition, our focus in this work is to investigate the role of technology in teaching sign languages. Therefore, we started from the beginning by exploring a small aspect before reflecting on the bigger picture. This small aspect is the arm and hand movements used in executing signs (the first stratum of four [37]). Based on the feedback received from our experts, if the technology were to be used in real life, the sign as a whole should be evaluated, not only part of it. However, given the complexity of covering the sign (e.g., knowing the meaning, using facial expressions), we started by exploring the primary potentials of each modality.

In future work, we plan to investigate how the various instruction modalities perform on a larger scale by providing multiple training sessions as well as examining further multimodal approaches. We will achieve this by combining the presented strength of each modality and introducing facial expressions and fingerspells.

8 CONCLUSION

In this paper, we explored the learning effects of various modalities (ie. audio, visual, EMS, as well as a combination of visual and EMS). While deaf or hard-of-hearing people may not benefit from all the presented learning conditions, it was still interesting to carry out a primary investigation to highlight the challenges and potentials for each modality (including audio), especially considering family members and friends of DHH people. A two-week study with 17 participants showed that visual-based instruction is preferred for the most detailed communication and conditions including EMS provide the best overall experience. The evaluation by ten ASL experts indicates a significant difference between the overall performances of the signs learned via the combined condition and those learned via the audio condition, with the combined condition rated as best. We identified the strengths and weaknesses for each modality used, contributing to the understanding of how the choice of modality can impact the learning effects and experiences in technology-based learning. Thus, we believe our work help improving technology-based autodidactic acquisition of sign languages.

ACKNOWLEDGMENTS

This work is funded by the German Federal Ministry of Education and Research (16SV8368).

REFERENCES

[1] João Gabriel Abreu, João Marcelo Teixeira, Lucas Silva Figueiredo, and Veronica Teichrieb. 2016. Evaluating sign language recognition using the myo armband. In *2016 XVIII Symposium on Virtual and Augmented Reality (SVR)*. IEEE, IEEE, Piscataway, New Jersey, United States, 64–70.

[2] Sílvia Grasiella Moreira Almeida, Frederico Gadelha Guimarães, and Jaime Arturo Ramírez. 2014. Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors. *Expert Systems with Applications* 41, 16 (2014), 7259–7271.

[3] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 16–31. https://doi.org/10.1145/3308561.3353774

[4] Jorge Jonathan Cadeñanes Garnica and María Angélica González Arrieta. 2014. Augmented Reality Sign Language Teaching Model for Deaf Children. In *Distributed Computing and Artificial Intelligence, 11th International Conference*, Sigeru Omatu, Hugues Bersini, Juan M. Corchado, Sara Rodríguez, Paweł Pawlewski, and Edgardo Bucciarelli (Eds.). Springer International Publishing, Cham, 351–358.

[5] Paul Chandler and John Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and instruction* 8, 4 (1991), 293–332.

[6] Michelene T. H. Chi and Ruth Wylie. 2014. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist* 49, 4 (2014), 219–243. https://doi.org/10.1080/00461520.2014.965823 arXiv:https://doi.org/10.1080/00461520.2014.965823

[7] Esomonu Chibuike. 2020. The Challenges of Teaching Sign Language to Pupils With Hearing Impairment in Special Education Primary School, Ibom Layout, Calabar. *Ibom Layout, Calabar (April 1, 2020)* (2020).

[8] Sarah Faltaous, Aya Abdulmaksoud, Markus Kempe, Florian Alt, and Stefan Schneegass. 2021. GeniePutt: Augmenting human motor skills through electrical muscle stimulation. *it - Information Technology* 63, 3 (2021), 157–166. https://doi.org/10.1515/itit-2020-0035

[9] Sarah Faltaous, Joshua Neuwirth, Uwe Gruenefeld, and Stefan Schneegass. 2020. SaVR: Increasing Safety in Virtual Reality Environments via Electrical Muscle Stimulation. In *19th International Conference on Mobile and Ubiquitous Multimedia* (Essen, Germany) *(MUM 2020)*. Association for Computing Machinery, New York, NY, USA, 254–258. https://doi.org/10.1145/3428361.3428389

[10] Sarah Faltaous and Stefan Schneegass. 2020. HCI Model: A Proposed Extension to Human-Actuation Technologies. In *19th International Conference on Mobile and Ubiquitous Multimedia* (Essen, Germany) *(MUM 2020)*. Association for Computing Machinery, New York, NY, USA, 306–308. https://doi.org/10.1145/3428361.3432081

[11] Sidney S Fels and Geoffrey E Hinton. 1993. Glove-talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE transactions on Neural Networks* 4, 1 (1993), 2–8.

[12] Susan Goldin-Meadow and Diane Brentari. 2017. Gesture and language: Distinct subsystem of an integrated whole. *Behavioral and Brain Sciences* 40 (2017).

[13] M. Guadagnoli, W. Holcomb, and M. Davis. 2002. The efficacy of video feedback for learning the golf swing. *Journal of Sports Sciences* 20, 8 (2002), 615–622. https://doi.org/10.1080/026404102320183176 PMID: 12190281.

[14] Jan Gugenheimer, Katrin Plaumann, Florian Schaub, Patrizia Di Campli San Vito, Saskia Duck, Melanie Rabus, and Enrico Rukzio. 2017. The Impact of Assistive Technology on Communication Quality Between Deaf and Hearing Individuals. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 669–682. https://doi.org/10.1145/2998181.2998203

[15] Mahmoud Hassan, Florian Daiber, Frederik Wiehr, Felix Kosmalla, and Antonio Krüger. 2017. Footstriker: An EMS-based foot strike assistant for running. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 1 (2017), 1–18.

[16] Marie-Claude Hepp-Reymond, Vihren Chakarov, Jürgen Schulte-Mönting, Frank Huethe, and Rumyana Kristeva. 2009. Role of proprioception and vision in handwriting. *Brain research bulletin* 79, 6 (2009), 365–370.

[17] Kevin Huang, Ellen Yi-Luen Do, and Thad Starner. 2008. PianoTouch: A wearable haptic piano instruction system for passive learning of piano skills. In *2008 12th IEEE international symposium on wearable computers*. IEEE, IEEE, Piscataway, New Jersey, United States, 41–44.

[18] Arthur R Jensen. 1971. Individual differences in visual and auditory memory. *Journal of Educational Psychology* 62, 2 (1971), 123.

[19] Nicholas O Jungheim. 2000. GESTURE AS A COMMUNICATION STRATEGY IN SECOND LANGUAGE DISCOURSE: A STUDY OF LEARNERS OF FRENCH AND SWEDISH. Marianne Gullberg. Lund, Sweden: Lund University Press, 1998. Pp. 253. *Studies in Second Language Acquisition* 22, 1 (2000), 122–123.

[20] Lih-Jen Kau, Wan-Lin Su, Pei-Ju Yu, and Sin-Jhan Wei. 2015. A real-time portable sign language translation system. In *2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, IEEE, Piscataway, New Jersey, United States, 1–4.

[21] Adam Kendon. 2000. Language and gesture: Unity or duality. *Language and gesture* 2 (2000).

[22] Adam Kendon. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

[23] Adam Kendon. 2008. Some reflections on the relationship between 'gesture'and 'sign'. *Gesture* 8, 3 (2008), 348–366.

[24] E.S. Klima and U. Bellugi. 1979. *The Signs of Language*. Harvard University Press. https://books.google.de/books?id=WeBOn6N8PJ8C

[25] Slavko Krapež and Franc Solina. 1999. Synthesis of the sign language of the deaf from the sign video clips. (1999).

[26] Annelies Kusters and Sujit Sahasrabudhe. 2018. Language ideologies on the difference between gesture and sign. *Language & Communication* 60 (2018), 44 – 63. https://doi.org/10.1016/j.langcom.2018.01.008

[27] Alina Kuznetsova, Laura Leal-Taixé, and Bodo Rosenhahn. 2013. Real-time sign language recognition using a consumer depth camera. In *Proceedings of the IEEE international conference on computer vision workshops*. IEEE, Piscataway, New Jersey, United States, 83–90.

[28] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, Springer, Berlin, Germany, 63–76.

[29] David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.

[30] Irit Meir, Wendy Sandler, Carol Padden, Mark Aronoff, et al. 2010. Emerging sign languages. *Oxford handbook of deaf studies, language, and education* 2 (2010), 267–280.

[31] Ross E Mitchell, Travas A Young, Bellamie Bachelda, and Michael A Karchmer. 2006. How many people use ASL in the United States? Why estimates need updating. *Sign Language Studies* 6, 3 (2006), 306–335.

[32] World Health Organization. [n.d.]. Deafness and hearing loss. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss. (Accessed on 01/27/2021).

[33] Prajwal Paudyal, Ayan Banerjee, and Sandeep K.S. Gupta. 2016. SCEPTRE: A Pervasive, Non-Invasive, and Programmable Gesture Recognition Technology. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 282–293. https://doi.org/10.1145/2856767.2856794

[34] Prajwal Paudyal, Junghyo Lee, Azamat Kamzin, Mohamad Soudki, Ayan Banerjee, and Sandeep KS Gupta. 2019. Learn2Sign: Explainable AI for Sign Language Learning.. In *IUI Workshops*.

[35] Max Pfeiffer, Tim Dünte, Stefan Schneegass, Florian Alt, and Michael Rohs. 2015. Cruise Control for Pedestrians: Controlling Walking Direction Using Electrical Muscle Stimulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2505–2514. https://doi.org/10.1145/2702123.2702190

[36] Max Pfeiffer, Tim Dünte, Stefan Schneegass, Florian Alt, and Michael Rohs. 2015. Cruise control for pedestrians: Controlling walking direction using electrical muscle stimulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2505–2514.

[37] Wendy Sandler. 2012. Dedicated gestures and the emergence of sign language. *Gesture* 12, 3 (2012), 265–307.

[38] Stefan Schneegass, Albrecht Schmidt, and Max Pfeiffer. 2016. Creating User Interfaces with Electrical Muscle Stimulation. *Interactions* 24, 1 (Dec. 2016), 74–77. https://doi.org/10.1145/3019606

[39] Caitlyn E Seim, David Quigley, and Thad E Starner. 2014. Passive haptic learning of typing skills facilitated by wearable computers. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM New York, NY, USA, New York, NY, USA, 2203–2208.

[40] Katja Stefan, Leonardo G Cohen, Julie Duque, Riccardo Mazzocchio, Pablo Celnik, Lumy Sawaki, Leslie Ungerleider, and Joseph Classen. 2005. Formation of a motor memory by action observation. *Journal of Neuroscience* 25, 41 (2005), 9339–9346.

[41] Jr. Stokoe, William C. 2005. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *The Journal of Deaf Studies and Deaf Education* 10, 1 (01 2005), 3–37. https://doi.org/10.1093/deafed/eni001

[42] Rachel Sutton-Spence and Bencie Woll. 1999. *The linguistics of British Sign Language: an introduction*. Cambridge University Press, Cambridge, England.

[43] Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. 2011. PossessedHand: techniques for controlling human hands using electrical muscles stimuli. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM New York, NY, USA, New York, NY, USA, 543–552.

[44] Sho Tatsuno, Tomohiko Hayakawa, and Masatoshi Ishikawa. 2017. Supportive training system for sports skill acquisition based on electrical stimulation. In *2017 IEEE World Haptics Conference (WHC)*. IEEE, IEEE, Piscataway, New Jersey, United States, 466–471.

[45] Sarah F Taub. 2001. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press, Cambridge, England.

[46] Sherman Wilcox and Phyllis Perrin Wilcox. 1997. *Learning to see: Teaching American Sign Language as a second language*. Gallaudet University Press, Washington D.C.

[47] Jeremy D Wong, Dinant A Kistemaker, Alvin Chin, and Paul L Gribble. 2012. Can proprioceptive training improve motor learning? *Journal of neurophysiology* 108, 12 (2012), 3313–3321.

[48] Mohammad Zakipour, Ali Meghdari, and Minoo Alemi. 2016. RASA: A Low-Cost Upper-Torso Social Robot Acting as a Sign Language Teaching Assistant. In *Social Robotics*, Arvin Agah, John-John Cabibihan, Ayanna M. Howard, Miguel A. Salichs, and Hongsheng He (Eds.). Springer International Publishing, Cham, 630–639.