AuslanSpell: An Interactive Technology for Improving Auslan Fingerspelling Comprehension

Kalin Stefanov
Monash University
Melbourne, Australia
kalin.stefanov@monash.edu

Andre Pham
Monash University
Melbourne, Australia
andrekypham@gmail.com

Antony Loose
Monash University
Melbourne, Australia
antony.loose@monash.edu

Lucy Robertson-Bell
Monash University
Melbourne, Australia
lucy.robertson-bell@monash.edu

Louisa Willoughby
Monash University
Melbourne, Australia
louisa.willoughby@monash.edu

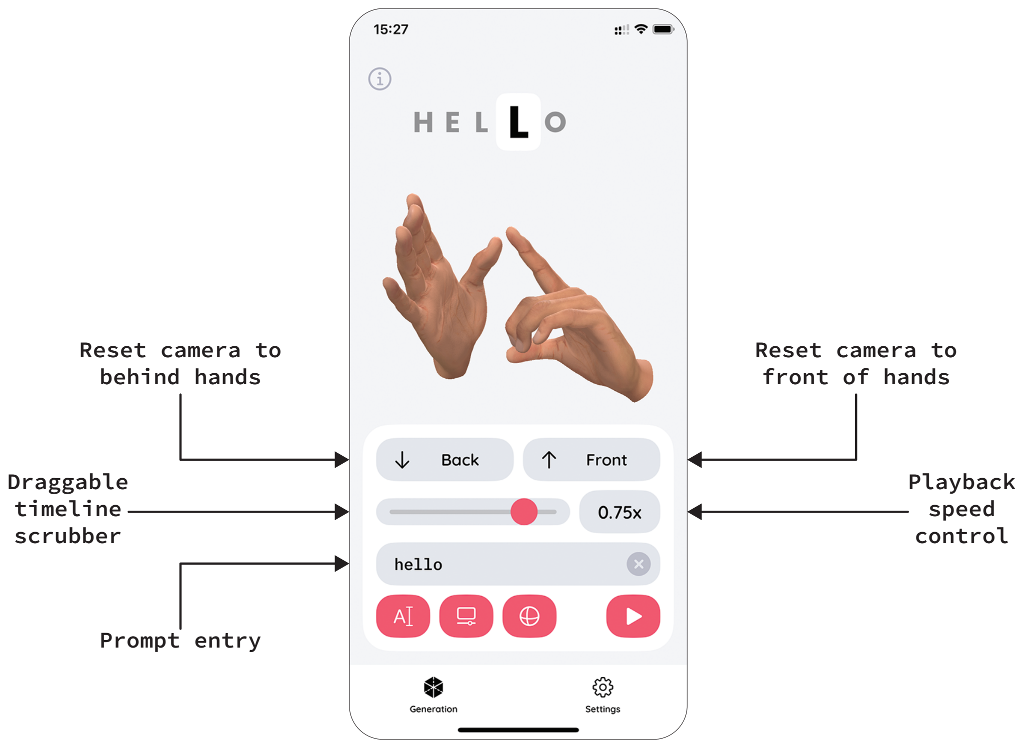


Figure 1: AuslanSpell's core feature allows users to type in any alphabetic prompt and produces animated 3D hands articulating the corresponding Auslan fingerspelling, with the hands seamlessly transitioning between signs.

Abstract

Fingerspelling, the manual representation of alphabetic characters, is a core element of sign languages. Learners report that reading back fingerspelling is among the most challenging aspects of learning sign languages, primarily due to limited practice opportunities. This paper presents AuslanSpell, a novel technology designed to enhance proficiency in Australian Sign Language (Auslan) fingerspelling. AuslanSpell allows learners to input any English word and produces 3D animated models that articulate the corresponding Auslan signs, featuring smooth hand transitions. In tests with 33 novice signers, after brief interaction with AuslanSpell, participants performed above chance on beginner multiple-choice tasks, correctly identified first and last letters more often than middle letters in free-text tasks, and reported high satisfaction – especially with features such as adjustable signing speeds and rotatable views. The results confirm AuslanSpell's potential in enhancing the comprehension and engagement of fingerspelling learners. AuslanSpell will be available through Auslan Signbank and Apple App Store, across iOS, iPadOS, macOS, and the web.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791563

CCS Concepts

• Human-centered computing → Accessibility technologies; **Interactive systems and tools**; **Participatory design**; **• Applied computing → Interactive learning environments**.

Keywords

Accessibility, Fingerspelling education, Participatory design

ACM Reference Format:
Kalin Stefanov, Andre Pham, Antony Loose, Lucy Robertson-Bell, and Louisa Willoughby. 2026. AuslanSpell: An Interactive Technology for Improving Auslan Fingerspelling Comprehension. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 26 pages. https://doi.org/10.1145/3772318.3791563

1 Introduction

Australia is experiencing a nationwide shortage of Australian Sign Language (Auslan) interpreters [77], leaving many Deaf[1] individuals isolated and unable to access essential services, including education and healthcare. Attaining certification as an Auslan interpreter requires years of dedicated study and practice to achieve the necessary proficiency.

Enhancing the quality and accessibility of Auslan training is essential to increasing the number of individuals who successfully pass interpreter certification and join the workforce as capable, inclusive communicators. A core element of Auslan is fingerspelling – the representation of individual alphabetic characters in manual form, which can be combined to spell out words. Many lexicalized Auslan signs incorporate components derived from fingerspelled letters [14], such as the sign for mother, which is an initialized sign based on the M handshape[2]. Fingerspelling is also commonly employed to spell proper nouns and English words that lack a standardized Auslan sign [35], as well as to clarify unknown signs during conversation [26, 42]. Like all signing, successful readback requires learners to process fine movement of the hands, body and any associated facial expressions or mouth movements [68], and to acclimatize to the way these are articulated by signers with different anthropometry.

Auslan learners report that decoding fingerspelling, i.e., reading it back, is among the most challenging aspects of learning the language, primarily due to limited opportunities for practice [88]. Learners of other sign languages also report that fingerspelling readback is a difficult skill to master [49, 51, 56, 75, 87], not least because of the fast pace of delivery, with fluent American Sign Language (ASL) signers capable of production rates in excess of 6 letters/second [61].

Digital platforms offer a promising avenue for accessible, individualized learning and practice. While numerous digital tools exist for learning ASL, Auslan-specific resources for interactive fingerspelling practice are limited. ASL and Auslan are completely unrelated sign languages with very different alphabets (the ASL alphabet is made with one hand, whereas the Auslan alphabet uses both hands), making ASL resources non-transferable for Auslan learning. Currently, the primary digital standard is a webpage[3] on Auslan Signbank, which displays still images of each letter in a typed word. However, this static approach falls short of effectively engaging learners or demonstrating how fingerspelling occurs in natural, fluid communication. Other available resources are typically non-interactive, delivered through static images or videos. Due to the dynamic nature of fingerspelling – which can involve any sequence of letters and complex transitions between them – it is impractical to compile a comprehensive dictionary of fingerspelled words. Learners in Thoryk [81] note their frustration that static resources may not be pitched at the appropriate level and quickly lose their utility once learners know which video is which word.

This paper introduces AuslanSpell, a novel technology designed to improve learner proficiency in fingerspelling – a critical skill for Auslan comprehension. AuslanSpell was developed through a collaborative design process that involved the active participation of Auslan educators and Deaf signers, and marks our initial step toward the development of interactive, engaging tools for Auslan education. This initiative directly addresses the interpreter shortage and the resulting isolation experienced by members of the Deaf community who lack access to communication in professional, social, and recreational contexts. The idea for the project emerged from informal conversations between one of the authors and Deaf teachers of Auslan about ways that current motion capture and machine learning technology might be put to use to create an Auslan educational resource that is functional and acceptable to the Deaf community. While many signing avatars give poor quality attempted translations from written language into a sign language[4] [4], our system avoids these hurdles by simply animating manually-encoded English spelling. It is thus an example of technology that (successfully) addresses a real-world teaching problem rather than a "disability dongle"[5] designed with little knowledge or consultation with the community.

AuslanSpell is available across multiple platforms, including iOS, iPadOS, macOS, and the web. The technology enables users to input any English alphabetic text, which is then converted into an animated 3D fingerspelling sequence. The system is fully interactive, allowing users to rotate the viewpoint in 3D space to observe the fingerspelling from various angles, such as from the signer's or the viewer's perspective. This functionality offers learners the opportunity to practice reading naturalistic fingerspelling without needing a live conversation partner – an unprecedented advancement that has the potential to significantly improve Auslan comprehension. Indeed, in tests with 33 novice signers, after brief interaction with AuslanSpell, participants performed above chance on beginner multiple-choice tasks and correctly identified first and last letters more often than middle letters in free-text tasks.

AuslanSpell will be freely accessible through Auslan Signbank and the Apple App Store. Signbank is the preeminent online Auslan dictionary and serves a broad user base, including trainee interpreters, parents of Deaf children, and late-deafened adults learning Auslan. Our key contributions are:

[1]Throughout this article we follow common convention and capitalise Deaf to refer to people who are signers and members of the Deaf community. We use deaf (no capital) to refer to the audiological condition – particularly as it pertains to people who are not fluent signers.
[2]https://www.auslan.org.au/dictionary/words/mother-1.html

[3]https://www.auslan.org.au/spell/twohanded.html
[4]https://www.mctd.ac.uk/bsl-not-for-sale-deaf-ai-procurement/
[5]https://blog.castac.org/2022/04/disability-dongle/

- The development of the AuslanSpell application, the first technology capable of producing 3D Auslan fingerspelling content from arbitrary English text prompts. The application will be freely available across multiple platforms, significantly increasing the prospective user base.
- The development of the AuslanSpell dataset, the first high-quality, motion-captured 3D Auslan fingerspelling dataset that enables the efficient creation of comprehensive and naturalistic training resources. The dataset will be freely available for research purposes.
- The validation of AuslanSpell as an effective learning technology for novice signers, demonstrating its potential to significantly enhance fingerspelling comprehension and learner engagement.

By contributing the AuslanSpell application and dataset, we hope that this work paves the way toward advanced technologies for accessible, individualized learning and practice of Auslan fingerspelling.

2 Related Work

2.1 Fingerspelling Readback

The past 30 years have seen an explosion of interest from hearing people in learning sign languages. ASL is now the third-most widely taught language on American college campuses [44] and numbers are booming throughout Australia, the UK, and Europe [46, 87]. However, a disproportionately high number of people teaching sign languages in universities are in teaching-only or casual roles [62], hampering the amount of research that is conducted on sign language teaching and learning.

Despite fingerspelling readback being widely acknowledged as one of the most difficult aspects of signing to master as a second language (L2) learner [56, 75, 88], there has been surprisingly little research on how L2 signers develop their fingerspelling skills or how to teach these skills more effectively [25]. Research with Deaf signers suggests that – just as fluent readers do not process words on a page letter-by-letter – fluent Deaf signers process fingerspelling by observing the wider shape of the word and chunking common letter sequences and movements [1, 86]. Learning to read back fingerspelling is thus not simply a matter of identifying each letter, but developing a feel for the movements and transitions that occur between forming the letters [75]. There are also clear rhythmic cues that signers orient to in producing and reading back fingerspelling, for example, the practice of holding the final letter in each word [84].

Empirical studies of fingerspelling readback confirm that accurately reading back decontextualized words is a demanding task. Deaf native signers consistently perform better than those who have learned the language later in childhood, who in turn perform better than hearing adult learners [40, 49]. Leannah et al. [40] also found that their Deaf participants were, on average, more confident that they had correctly identified the fingerspelled word than hearing participants, although it is important to note that the Deaf signers also self-reported higher ASL proficiency, so this may be a proficiency effect more than a true Deaf/hearing comparison. Not all words are equally easy to read back, with Leannah et al. [40] finding statistically different results in accuracy between real and nonce place names, and Shipgood and Pring [75] also observing statistically significant differences in accuracy between words with regular and irregular spelling. Working with adult L2 signers who regularly used British Sign Language (BSL) with family or at work, Shipgood and Pring [75] further found statistically significant increases in accuracy when the signing pace was reduced and when a prompt was provided to cue what the word might be. Under favourable conditions, participants were able to accurately identify over 90% of words presented, but at a faster pace and/or without prompts, performance was around 70% accurate.

An important question in fingerspelling readback studies has been the relative importance of the movements between letters (i.e., transitions) as a cue to what is being signed as opposed to the processing of the fully formed letter signs. Schwarz [69] found that fluent Deaf signers can often accurately read back a word even if only shown transitions, not the fully formed letters. However, Geer and Keane [26] found that third-semester ASL students saw a steep decline in their ability to read back words if shown only transitions (i.e., with masking of the fully formed letters), while videos that masked the transitions and showed only the fully formed letters were read back more accurately than videos with no masking at all. Geer and Keane [26] also found that a short intervention they designed to help students boost their fingerspelling skills by engaging in training on how letters are formed in connected signing led to significantly higher scores on the post-test of fingerspelling ability, suggesting that consciousness raising on the salience of transitions can lead to improved outcomes for L2 learners.

In evaluating this research, it is important to note that transitions in ASL (which is spelled with one hand) operate very differently from the two-handed alphabet used in Auslan and BSL (among other sign languages) [25]. Care needs to be taken in extrapolating findings from ASL to Auslan fingerspelling; however, a system that attempts to mimic transitions between letters – and that allows learners to practice different words at different speeds – seems to have promise as a learning tool based on the above findings.

2.2 Fingerspelling Production

Fingerspelling production (sometimes generation/synthesis) refers to synthesizing and rendering in an avatar or video the manual alphabet articulations that a human signer would produce to spell a word or phrase, ideally with natural timings, co-articulations, and non-manual elements. The two immediate complications are 1) Alphabet variation across languages and 2) Highly co-articulated, fast, and small finger motions [38]. Provided the scarce literature on fingerspelling production, in this section, we expand the scope to include studies on the related topic of sign language production, spanning from rule-based animations of known handshapes to modern data-driven approaches.

Fingerspelling production must ensure that the output is both linguistically accurate (i.e., each letter is correctly formed) and naturally fluent (i.e., timings and motions resemble human signers). However, generating realistic and intelligible fingerspelling is challenging due to several factors. One major challenge is modeling realistic transitions between letters. Unlike static images, which show isolated handshapes, real signing involves continuous movement – the transitions are as important as the static poses for perception [24]. Straight interpolation from one letter's handshape to the next can cause the hands or fingers to collide or look unnatural [6, 7]. McDonald et al. [45] proposed a data-driven collision avoidance algorithm that focuses on the motion of individual fingers during transitions, where instead of treating each letter-to-letter transition as a whole, they predict how to move each finger to avoid collisions.

A closely related issue is sign co-articulation, the fact that the shape or motion of a given letter can be influenced by the preceding and following letters, analogous to how speech sounds are affected by neighboring sounds. In ASL, co-articulation results in certain letter sequences, such as I-L-Y and U-R, being articulated simultaneously (i.e., as a single sign) rather than sequentially [26, 49]. Co-articulation is less studied in Auslan, but Cresdee and Johnston [15] provided preliminary observations about how the articulation of the B handshape varies in their corpus data.

Achieving a high level of realism remains a difficult problem; even when hand movements are technically correct, subtle aspects such as acceleration and deceleration of motions, finger stiffness, slight random variation in each repetition, natural idle motions (such as brief, relaxed handshakes for a pause in spelling), and integration with facial expression can make the difference between a robotic look and a human-like one. Early rule-based approaches were often criticized as robotic or uncanny. Elliott et al. [19] developed the Signing Gesture Markup Language (SiGML), an XML notation closely tied to the Hamburg Sign Language Notation System (HamNoSys) [29]. HamNoSys encodes handshapes, orientations, locations, and movements with high granularity. A typical pipeline would include 1) Authoring that produces HamNoSys, 2) HamNoSys is converted to SiGML, and 3) An animation engine that realizes the motion with avatar-specific kinematics and inverse kinematics. Rule-based approaches offer flexibility (any word could be signed/fingerspelled) but often require fine-tuning for natural transitions and avoiding collisions.

Beyond rule-based systems, there is a growing trend of using data-driven generative models for sign language production. Sign-Diff [20] enhanced the correspondence between textual inputs and dense sign language motions while reducing the occurrence of multiple fingers. Neural Sign Actors [5] was trained on a carefully curated large-scale 3D signing avatar dataset. The model can generate 3D avatar sequences from text through a diffusion process formed on an anatomically informed Graph Neural Network [66] defined on the SMPL-X [52] human body. Sign-IDD [79] provided a new perspective on joint associations by disentangling conventional 3D joints into 4D bone representations. These approaches draw insights from human motion synthesis literature, where natural language descriptions are converted to appropriate 3D human body movements. Initial research on human motion synthesis focused on action generation tasks; ACTOR [54] proposed a Variational Autoencoder (VAE) [39] to embed actions for motion generation, and the DSAG [28] framework ensured diversity of the generated motions by conditioning synthesis on both actions and temporal constraints. However, since the distributions of natural language descriptions and motion sequences differ greatly, learning a probabilistic mapping between them poses a challenge, as noted in previous studies, e.g., MotionCLIP [80]. Joint-latent models such as TEMOS [55] use VAE architectures to learn a joint latent space for text and motion. Inspired by the success of diffusion-based generative models in other fields, diffusion-based approaches adopt conditional Diffusion Models [31] for human motion synthesis. However, diffusion models that operate on raw motion sequences are susceptible to noise and face inefficiencies, leading to artifacts and high computational overhead during training and inference [57, 65, 91]. To address these issues, MLD [12] proposed a latent diffusion model [63] that operates within the latent space of a VAE.

Although these models demonstrate impressive upper body synthesis, they typically struggle with fine-grained hand articulations that fingerspelling demands. Consequently, ensuring that a neural network correctly articulates fingerspelling (with correct handshapes and transitions) remains a challenge. Nevertheless, this trend represents a shift towards data-driven sign language (fingerspelling) production, where the nuances of transitions and timings could be learned automatically from examples of real signers. However, currently, the literature lacks studies on data-driven generative models for fingerspelling production.

Large-scale and high-quality datasets are essential for advancing data-driven generative models for sign language production. They allow models to learn statistical patterns associated with hand shapes, movements, and facial expressions in signing and how people spell words (e.g., speeds, timings, mouthings, hand poses, and letter combinations) in fingerspelling. Although existing datasets (e.g., for American Sign Language [16, 27, 36, 43, 50, 70, 73, 74, 82], British Sign Language [2, 3, 67], German Sign Language [10, 22, 30], Australian Sign Language [71, 72], Sign Language of the Netherlands [58, 59], Chinese Sign Language [32, 33], Polish Sign Language [37], and Turkish Sign Language [76]) have contributed significantly to the development of sign language generation models, they come with several limitations. A common drawback is missing accurate 2D and 3D annotations, making it challenging to develop models that generate precise hand, face, and body movements. Most datasets only consist of videos, without depth information or other 3D details, limiting the ability to capture the full spatial complexity of sign languages. As the field moves toward data-driven approaches, the availability of large and diverse datasets will be increasingly important for developing generative models, and in addition to intelligibility and comprehension studies with Deaf signers [17, 18], to automatically validate the quality of synthesized signing and fingerspelling.

3 AuslanSpell

AuslanSpell draws on motion capture and computer graphics technology to design and develop an automatic and interactive production of Auslan fingerspelling. We recorded Deaf Auslan signers fingerspelling the letters of the English alphabet and the associated transitions, creating the first large-scale, high-quality 3D motion-captured dataset. Through a collaborative design process that involved the active participation of Auslan educators and Deaf signers, a subset of the dataset was analyzed and post-processed, allowing the creation of 3D handshape animations that smoothly and accurately articulate all 26 letters. The 3D handshape animations were then used to develop AuslanSpell, a novel technology designed to improve learner proficiency in Auslan fingerspelling. AuslanSpell allows learners to input any English letter sequence and produces 3D animated hands that accurately articulate the corresponding Auslan fingerspelling, featuring smooth transitions between letters. This approach aims to balance the accuracy of letter articulations with smooth and natural transitions to deliver realistic fingerspelling content.

The application was developed in conjunction with a co-design team of 5 experienced Deaf Auslan teachers, two of whom also served as signing models, and worked as research assistants providing validation support for the animations produced. Co-designers included representatives from major Auslan teaching providers and Deaf Australia (the peak Deaf community body). All members were reimbursed for time spent working on the project. Deaf researchers such as Angelleni [4] and de Meulder [47] have rightly called for the field of sign language technology research to engage much more proactively with Deaf community members to ensure that technology developed is fit-for-purpose and addresses real world needs (see also Huang [34]). In the case of AuslanSpell, it is not technology aimed at Deaf people (who already know how to sign and fingerspell fluently), but rather at Auslan students who wish to improve their Auslan skills. As noted in the introduction, the initial idea for the application came from informal conversations with Auslan teachers, who noted that current signing avatar technology is not high-quality enough to use in educational settings, but felt that fingerspelling is an area that is constrained enough that it could be possible to achieve a fit-for-purpose application. Once funding was secured, initial meetings with the co-design team were used to identify core functionality that the application should have (such as animation of transitions between signs, the ability to show left and right-handed signings, the ability to rotate viewing angle) and to identify suitable signing models for the project (8 signers in total). As each iteration of the application was developed it was taken to co-designers for feedback, which was then integrated into future cycles. As discussed below, changes arising from the co-design process included adding/tweaking the realistic hand models to make clearer how the signs are being made, revising the approach taken to animating transitions between letters to enhance realism and many small corrections to the animation of individual letters to reflect the group's view on how the citation form of each letter should be presented. In addition to the formal co-design process, one of the authors also shared near-final versions of the application with Deaf and hearing signing colleagues and garnered informal feedback from them around the accuracy and fluency of animations that were fed back into the animation cleaning process.

3.1 Dataset

3.1.1 Scope. The complete dataset includes words with prefixes, words with suffixes, names, and places, where the scope was defined to ensure coverage of the most common letter combinations in English[6]. A total of 93 words (see Appendix A.1 and Appendix A.2) were selected and recorded with two data collection techniques to capture handshape and upper body information. Handshapes were recorded with the Manus Prime 3 Mocap gloves[7]. The gloves measure the splay and bend of the fingers and the relative rotation of each joint within the hands. Upper body motions (i.e., place of

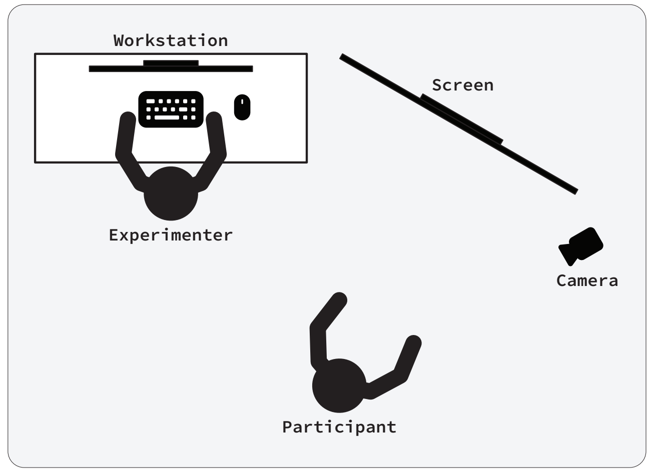


Figure 2: Illustration of the physical setup used during the AuslanSpell dataset recordings.

articulation and movement of the arms) was recorded using a Vicon Motion Capture[8] setup with reflective markers.

3.1.2 Production. We have developed an efficient data recording pipeline that achieves an average capture time of 5 seconds per word, although the time varies between participants. The setup preparations, including the fitting of the gloves, positioning the reflective markers, and calibration, require approximately 1 hour.

Our motion capture space includes a designated detection zone for Vicon cameras, a high-resolution video camera mounted on a tripod, a screen to display words, and a desktop workstation to control the recordings (see Figure 2). The recording of a session (i.e., a session captures 1 signing model fingerspelling all 93 words) consists of the following steps:

- The signing model assumes an upright posture, with the arms relaxed in a neutral position around their body.
- The signing model fingerspells the English alphabet, their name, and suburb as a warm-up.
- The experimenter presents a slide with several words on the dedicated screen.
- The signing model fingerspells the presented words, making sure to return to the neutral stance between words.
- The experimenter moves on to the next slide, once all words on the current slide are correctly fingerspelled.

Signing models were instructed to fingerspell at a normal, natural pace and could repeat the articulation of any word in case there were errors. A total of 8 Deaf signers contributed to 31 sessions. All were reimbursed for their time. All were fluent signers who are experienced Auslan educators, well-respected in the community, and were recommended to us by our co-designers as appropriate language models. More details on the signing models and their contributions to the dataset are provided in Appendix A.3.

3.1.3 Post-Production. A subset of the AuslanSpell dataset went through a thorough iterative validation process involving Deaf

[6]https://homepages.math.uic.edu/~leon/mcs425-s08/handouts/char_freq2.pdf
[7]https://www.manus-meta.com
[8]https://www.vicon.com

signers to ensure error-free letter articulations. This process included detailed feedback on each letter animation given by two Deaf signers working as research assistants on the project, as well as discussion within the wider co-design group at project meetings. As noted by Angelini [4], correctly annotating tokens and cleaning animations produced is a highly labour-intensive process that relies on detailed feedback from Deaf community members to ensure accurate and realistic avatar development. This data subset, used in the AuslanSpell application, includes all 26 letters fingerspelled by 2 Deaf signers (one left- and one right-handed), 2 sessions each. The post-processed subset will be freely available for research purposes under the CC BY-NC-SA 4.0 license[9].

The video recordings were transcribed with the ELAN [90] annotation tool, producing 2 tracks of metadata, one for words and one for letters. The words track provides information about the beginning and end times of each word in the corresponding recording session. Similarly, the letters track logs information about the beginning and end times of each letter. In a final step, the video recordings were temporally aligned with the motion capture recordings.

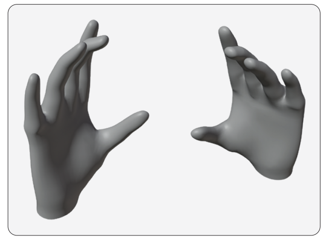
The handshape recordings from the Manus gloves were retargeted onto a humanoid rig (i.e., system of bones, joints, and controllers for a 3D model) using the Rokoko[10] plugin in Blender[11]. The Manus rig mesh is mechanical in structure, necessitating retargeting to transfer the motions on a human-like mesh with arms. Once on the humanoid rig, the motions required cleanup to fix common issues. These included fingers bending too far or not enough, signs that were supposed to have contact points but did not, and others that had contact where they should not. Cleanup was done with reference to the video recordings and Auslan Signbank to ensure accuracy.

After the first round of corrections, using the ELAN transcriptions, the motions were split into individual animation blocks for each letter. This meant copying the section of keyframes for a given letter onto a duplicate rig and aligning both the starting and ending poses with a consistent rest position. Another pass of cleanup was needed to adjust the timing of transitions, since the hands needed to return to rest naturally and not too quickly or too slowly. Each letter went through this process, ensuring it could be used independently.

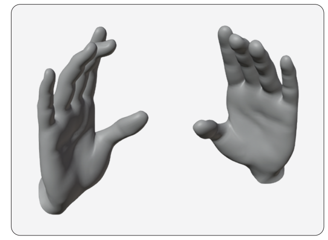
Animation export was handled differently depending on the target platform requirements. For the iOS/iPadOS/macOS version of the application, each letter was exported as a separate file. This resulted in creating sets of 26 .dae[12] files, one set per recording session. For the web version of the application, the animations were exported in session-based bundles where all letters were included as separate animation tracks within a single file. These were exported in .glb[13] files.

Each export was done for two mesh versions. The simple hands (see Figure 3a) version uses a smooth, low-detail mesh, while the realistic hands (see Figure 3b) version has a more detailed surface. The realistic hands mesh is slightly larger, resulting in problems

[9]https://www.creativecommons.org/licenses/by-nc-sa/4.0/
[10]https://www.rokoko.com
[11]https://www.blender.org
[12]Format for COLLADA (Collaborative Design Activity), an XML-based standard for exchanging 3D data
[13]Format for glTF (Graphics Library Transmission Format), a JSON/binary-based standard for exchanging 3D data



(a) Simple hands



(b) Realistic hands

Figure 3: 3D hand models included in the AuslanSpell application. The simple hands (a) use a smooth, low-detail mesh, while the realistic hands (b) use a more detailed surface.

with touch points not aligning correctly when compared to the simple hands version. A further round of adjustments was performed to fix these discrepancies before the final export.

Through this process, the handshape recordings from the Manus gloves were converted into clean, accurate, and platform-ready animations. The process ensured that the animations could be reliably used in the AuslanSpell application, with both simple and realistic hand models available for different display requirements.

3.2 Application

3.2.1 Platforms. The AuslanSpell application is supported on the following platforms:

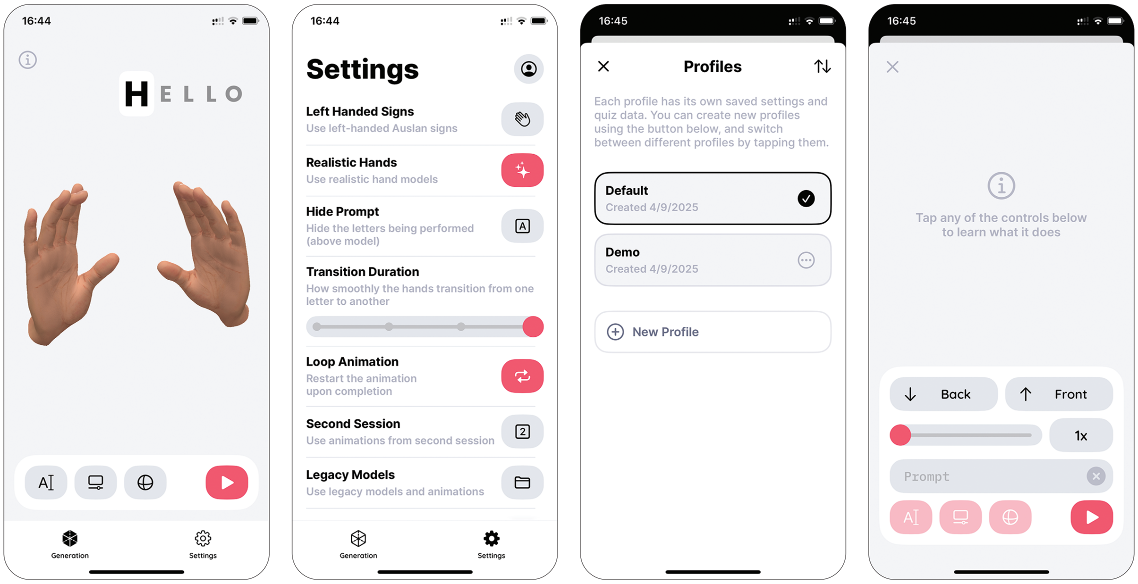
- iPhone (iOS 15.0 or later).
- iPad (iPadOS 15.0 or later).
- Mac (macOS 12.0 or later and Apple M1 or later).
- Web (JS ES2016 or later).

The iOS/iPadOS/macOS version of the application will be distributed free of charge on the Apple App Store[14], while the web version will be distributed on the npm package manager[15]. To experience the exact functionality of the web version used for user evaluation, refer to this webpage[16]. We welcome researchers, developers, and learners to build on our work, and thus the application source code will be publicly shared on GitHub[17][18].

3.2.2 Overview. The application's core functionality allows users to type in any alphabetic prompt and produces animated 3D hands articulating the corresponding Auslan fingerspelling, with the hands seamlessly transitioning between signs. The entered prompt is displayed above the hands and tracks the letter that the hands are actively articulating. The application is composed of 4 pages: Generation, Settings, Profiles, and Information page (see Figure 4), incorporates responsive design, supports light/dark mode, and adjusts its layout in response to the device/window size.

Generation Page. The generation page includes a 3D scene and scene controls. The animated hands are rendered in the 3D scene. The 3D scene allows users to view the hands from any angle and

[14]https://www.apple.com/app-store/
[15]https://www.npmjs.com
[16]https://alpha.auslan-spell-web-package.cloud.edu.au
[17]https://github.com/monash-assistive-tech/auslan-spell-ios
[18]https://github.com/monash-assistive-tech/auslan-spell-web



(a) Generation (b) Settings (c) Profiles (d) Information

Figure 4: The AuslanSpell application is composed of 4 pages – (a) Generation, (b) Settings, (c) Profiles, and (d) Information page.

position, using touch gestures or mouse inputs to control the camera (see Figure 5):

- Panning with one finger or holding the left mouse button to orbit around the hands.
- Panning with two fingers or holding the right mouse button to move orthogonally to the viewing axis.
- Pinching with two fingers or scrolling the middle mouse button to zoom in and out.

The iOS/iPadOS/macOS version supports two additional camera controls: 1) Panning with three fingers to move forward or backward along the viewing axis and 2) Rotating with two fingers to roll around the viewing axis.

The scene controls provide access to the following functionality:

- **Play/Pause button:** This button allows pausing and resuming animation playback. If tapped while a prompt is being entered, the prompt is automatically submitted and playback begins.
- **Prompt input:** This input allows entering the text articulated by the hands. Unsupported characters are automatically sanitized.
- **Timeline scrubber:** The timeline scrubber tracks the animation's progress, and can be manually dragged to any position to update the animation. During dragging, the hands update to match the timeline scrubber's position. The iOS/iPadOS/macOS version supports a haptic trigger to provide tactile feedback when the user drags to a position that changes the letter.
- **Speed controls:** The playback speed controls allow the user to adjust the animation speed. The available preset speeds are 0.25x, 0.5x, 0.75x, 1x (the speed at which the original fingerspelling was articulated), and 1.5x.
- **Reset controls:** The camera reset controls allow the user to reset the camera position and direction to face the front or back of the hands. This allows users to get back to the default viewing angle after changing it through touch gestures/mouse inputs, or to quickly switch between facing the front and back of the hands.

Settings Page. The settings page provides controls over additional features supported by the AuslanSpell application:

- **Left- and right-handed:** Selects whether the hands articulate left- or right-dominant fingerspelling.
- **Realistic and simple hands:** Toggles whether the hand models are realistic or simple.
- **Hide and show prompt:** Toggles whether the prompt display is shown or hidden. The prompt display is rendered above the hands, showing the prompt and the letter that is actively being articulated.
- **Transition duration:** Allows users to select the blending amount between signs. The minimum value indicates no blending, where each letter is articulated one at a time with the hands returning to a neutral pose between letters. The maximum value indicates an optimized blending amount, where the hands transition seamlessly from one letter to the next. Increasing the transition duration attempts to match natural hand transitions between signs. Lower transition durations are intended for users with less experience, who may find it difficult to distinguish the start and end of signs.
- **Loop and end:** Controls whether the animation should restart or stop playing upon completion. If enabled, the animation replays automatically, and if disabled, the animation pauses upon reaching the end.

Profiles Page. In anticipation of a major extension that will offer an Auslan fingerspelling course (similar to Duolingo[19] for Auslan), the AuslanSpell application supports profiles that allow users to store their data and settings. Switching between profiles is as simple as tapping the desired profile on the profiles page.

Information Page. The information page guides each of the scene controls. It is accessible via the info icon button on the top left of the generation page.

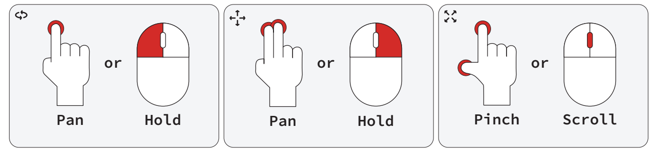
3.2.3 Development. The described application development timeline[20] includes three major milestones – 1) Minimum viable product, 2) Visual realism and fidelity, and 3) Profiles and refinements. These milestones also mark points where Auslan educators and Deaf signers provided major feedback on the state of the application and suggested directions for the next development stage.

AuslanSpell was created as a native application for iOS 15.0 and iPadOS 15.0 and higher. Apple platforms were targeted as requested by Auslan educators collaborating on the project, due to the majority of accessibility-focused learning occurring on Apple platforms. The exact iOS and iPadOS versions were selected to support the vast majority of active devices (at the time, approximately 99.92% of iPhones were compatible with iOS 15.0 or later) based on the relevant Orbital Survey[21] for market share of the iOS versions.

Minimum Viable Product. During this stage, development was focused primarily on the core functionality and key features of the application. The core functionality includes the conversion of arbitrary text prompts to 3D fingerspelling articulations, scene controls, and settings. Entering a text prompt (i.e., any alphabetic character sequence) produces animated, interactive, 3D hand models articulating the prompt in Auslan fingerspelling. Current online resources for individualized learning and practice of Auslan fingerspelling are limited to images and videos. Images are limited to angle and position and lack context, such as transitions between signs, pacing, and movement. Similar to images, videos are limited to angle and position and lack customization, such as the range of words. AuslanSpell's support for custom prompts and a 3D scene resolves these limitations, producing a realistic and accurate articulation for any alphabetic prompt and allowing any viewing angle through interactive gestures in the scene.

Scene controls include a play/pause button, a timeline scrubber to track and control articulation progress, playback speed controls, and front and back camera reset controls. These were required to provide users with full control over the articulated prompt. The play/pause button allows users to pause the animation and inspect the position and shape of the hands at any point. The timeline scrubber allows users to follow the animation's progress and manually move through it to any point in time. Playback speed controls allow users to view the animation at the speed they are comfortable with, as well as inspect the articulations in slow motion. The camera reset controls allow users to flip between the front and back of the hands and to reset the viewing angle.

[19]https://www.duolingo.com
[20]Note that the timeline is not exhaustive and excludes aspects such as research and planning, feature validation, UI/UX revisions and improvements, internal changes and refactors, optimizations, internal tooling and testing, general improvements, changes to features, and bug fixes
[21]https://www.telemetrydeck.com/blog/category/orbital-survey/



(a) Orbit **(b) Move** **(c) Zoom**

Figure 5: The AuslanSpell application supports different interaction modalities to control the camera in the 3D scene – (a) Orbit around the hands, (b) Move orthogonally to the viewing axis, and (c) Zoom in and out.

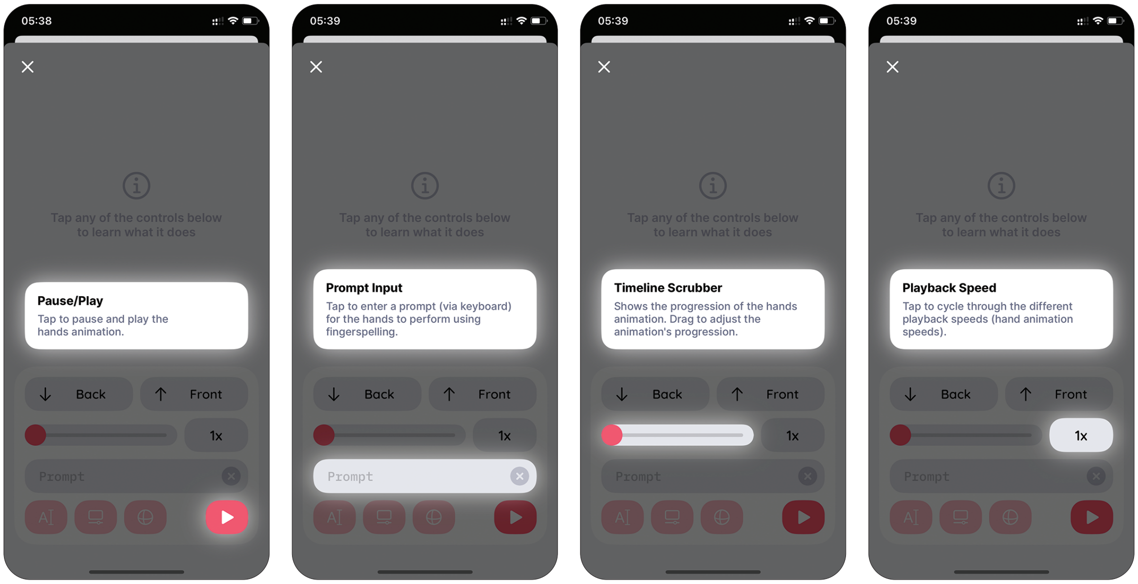
The settings provide toggles for left- and right-handed articulation, playing sequentially instead of interpolating between signs, and hiding the prompt display. The settings provide application preferences and allow customization of the hand models and the way they articulate signs. Toggling between left- and right-handed models was required to support both left- and right-handed users. Toggling sequential and interpolation playback, suggested by Auslan educators, supports novice learners in distinguishing between signs, while experienced signers could view more realistic articulations. Hiding the prompt display allows learners to test their knowledge by interpreting what is being articulated without seeing the answer.

The key features included a new animation engine. The initial version of AuslanSpell generated the transitions between letters using interpolation. Following Auslan educators' feedback through the co-design process, the animation engine was redesigned to blend the animations instead of linearly interpolating between them. This allows for smooth and seamless transitions between the articulated letters. This was required to accomplish realistic articulations, a prerequisite for the application to be widely adopted as an accurate and reputable source for learning Auslan fingerspelling.

This stage marks the introduction of support for Mac (macOS 12.0 or later and Apple M1 or later) and the web (JS ES2016 or later), making the application compatible with iOS, iPadOS, macOS, and modern web browsers, e.g., Google Chrome, Mozilla Firefox, and Safari. Extended support was added to increase the prospective user base.

In addition to user experience, the development of AuslanSpell aims to enhance prospective developers' experience with the application. For example, the web version, an npm package, is built with vanilla TypeScript (compiles to JavaScript) and is highly configurable (see Appendix B.1). Building upon vanilla web technologies results in framework-agnostic code, a crucial feature in modern web development, given the popularity of front-end frameworks [78]. The outcome is a highly extensible and configurable package that is simple to integrate into any web application, regardless of the technology stack. This enables external developers and researchers to extend our work by implementing user-facing tools, such as quiz applications, plug-ins that fingerspell highlighted words, and alternatives to closed captions.

Visual Realism and Fidelity. During this stage, development was focused primarily on providing visual realism, fidelity, and accuracy to the animations and the scene.



(a) Pause/play (b) Prompt (c) Timeline (d) Speed

Figure 6: Information pages for the functionality of some scene controls of the AuslanSpell application – (a) Play/Pause button, (b) Prompt input, (c) Timeline scrubber, and (d) Playback speed page.

Realistic hands were added. The initial version of AuslanSpell used simple hand models (see Figure 3a). The realistic hand models (see Figure 3b) are anatomically accurate, include textures, and interact better with light. Realistic hands were requested by Auslan educators collaborating on the project to reflect how signs actually look in real-world contexts. The simple hands were kept for users who prefer them and became a setting in the application.

Scene lighting was added for more natural scene rendering, including the hands casting shadows. Realistic scene lighting was requested along with the realistic hands by Auslan educators to support the initiative for visual realism.

Auslan educators reviewing the application noticed that in some instances, the core action of a sign (e.g., fingers making contact) was not completed; the fingers would almost touch, but the hands would start transitioning into the next sign before the fingers could make full contact. To resolve this, the animation engine was updated to introduce new animation blending logic which retains the core aspect of each sign (e.g., the index finger and the thumb making contact during A) via a concept called key intervals (see Appendix B.2). Each sign was assigned a key interval that could not be skipped, ensuring the integrity of the sign's articulation. Provided the updated animation engine, the prompt display, which shows the prompt and current letter being articulated above the hands, was updated.

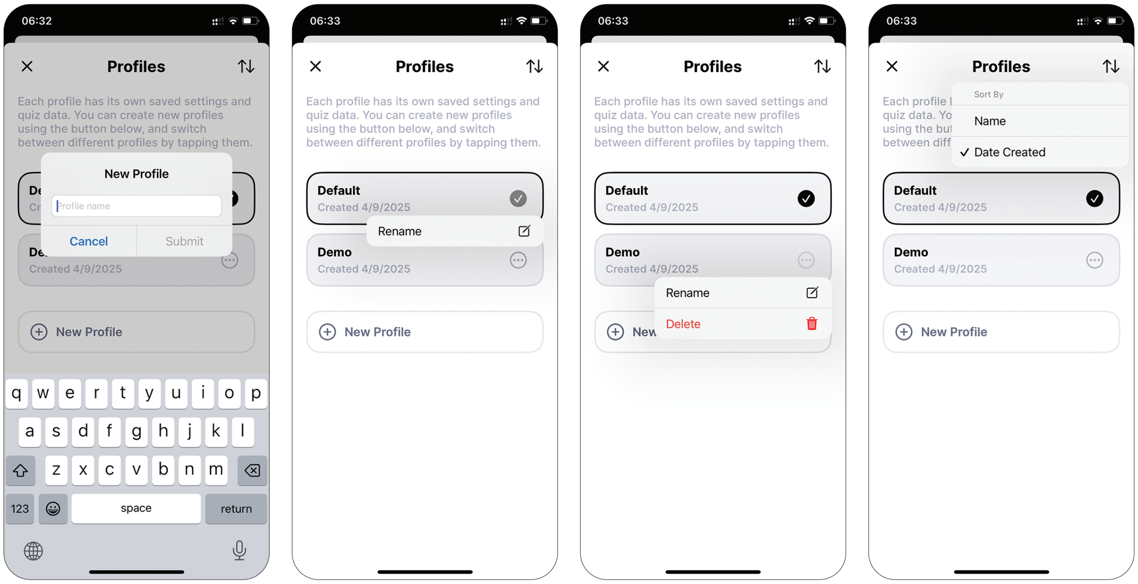
Variable transition durations (see Appendix B.3) were added to allow users to select the amount of blending between letters. The customization of transition durations was requested by Auslan educators for users with different levels of experience. Novice users could start with each letter articulated independently, and as they gain experience, they would increase the blending amount until reaching the maximum, where hands transition seamlessly between signs.

A Deaf signer reviewed the animations of every sign and provided detailed professional guidance on the identified issues and potential improvements. These documented suggestions were then applied to the hand animations and models to ensure the accuracy of sign articulations. This resulted in a new suite of hand models and animations added to the application. The old suite of hand models and animations became a setting in the application.

The information page was added to provide explanations for the function of the scene controls. Figure 6 illustrates the information pages for the functionality of some scene controls. The information page guides users who might need help understanding the controls, making the application more accessible to those with less technical experience.

Profiles and Refinements. During this stage, development was focused primarily on profiles and refinements to the scene controls and the application as a whole.

As a standalone application that does not require sign-up or login, profiles were introduced, whereby users could effectively switch to different local accounts on a shared device. Profiles allow different users to save different settings on the same device, enabling multiple users to share the same device without interfering with each other's preferences. This base functionality will be extended in a future AuslanSpell release that will include an Auslan fingerspelling course. The profiles page provides the ability to rename, delete, and sort profiles (see Figure 7). The design applies UI/UX principles to profiles, including avoiding the case whereby users could make an irreversible change, given they made a mistake in naming or creating a profile.



(a) New (b) Rename (c) Delete (d) Sort

Figure 7: Profiles pages functionality in the AuslanSpell application – (a) New, (b) Rename, (c) Delete, and (d) Sort.

Anonymous user metrics were also introduced to track trends and help make more informed decisions regarding future application improvements. User metrics support both short-term anonymous data collection (e.g., during user studies) as well as monitoring the application usage in production.

The animation engine was reworked from the ground up to provide full access and control over the articulation sequence. This major update resolves limitations of the previous animation engine, such as allowing for smooth scrubbing throughout the entire sequence, including mid-blend, and allowing for the changing of the playback speed without clamping the animation progress while paused. These enhancements improved usability and brought the playback functionality in line with standard media player capabilities, ensuring users are not constrained in how they interact with and review Auslan fingerspelling articulations. Based on feedback from several users, the prompt display letters (which show the prompt and current letter being articulated above the hands) were updated to lowercase. This will become a setting and will be available in a future release.

A new setting to toggle the articulation loop was introduced. Disabling the loop animation setting results in the animation ending upon completion. This optional setting was introduced as part of the reworked animation engine, aligning the application's capabilities with those of other media players.

The layouts were redesigned to be responsive on all screen sizes. Previously, the application was optimized for mobile screens. These changes were introduced to provide a consistent experience on all supported platforms.

4 User Evaluation

4.1 Participants

Participants were university students who were novice signers, that is, they had never attended Auslan classes or learnt to sign from Deaf people. We also excluded students who had studied or were users of other sign languages as a precautionary principle to eliminate any language transference effects. Participants were not pre-tested on their fingerspelling readback ability, since knowledge of any sign language was an exclusion criteria for participation.

The study was actively promoted to students enrolled in Linguistics units, but open to all students currently enrolled at Monash University. We focused on Linguistics students as past experience has shown that they are often keen to participate in language learning experiments because of their inherent interest in language. They are also a reasonable proxy for the intended audience of the app, namely people studying Auslan in formal classrooms. Thirty-three students volunteered to participate: 29 undergraduates, 3 master's, and 1 postgraduate. The core demographics of the participant group are reported in Table 1.

Our sample skews female, but this reflects a wider trend of language studies being a feminized pursuit [11, 85]. Handedness is roughly in line with global norms [53]. Despite Australians often being characterized as having a monolingual mindset [13], our participants were quite multilingual, with most having either intermediate or advanced proficiency in a language other than English (LOTE). This is likely an artifact of the research task being more appealing for students who are inherently interested in learning languages. It should be noted that our participants are more experienced language learners than the average class of Auslan students

Table 1: Core demographics of the participants in the user evaluation.

Gender	**Count**	**Fraction**
Female	20	60.61%
Male	6	18.18%
Male, Non-binary / Third gender	1	3.03%
Non-binary / Third gender	5	15.15%
Prefer not to say	1	3.03%
Age	**Count**	**Fraction**
17-20	12	36.36%
21-25	20	60.61%
31-40	1	3.03%
Handedness	**Count**	**Fraction**
Ambidextrous	1	3.03%
Left-handed	2	6.06%
Right-handed	30	90.91%
Language Learning Experience	**Count**	**Fraction**
I am a native speaker or highly proficient in a LOTE (e.g., completed university major, CEFR C1 or higher)	11	33.33%
I have beginner level proficiency (e.g., studied at primary/ junior secondary school, one or two beginner units at university, CEFR A2 or below)	8	24.24%
I have intermediate proficiency (e.g., completed VCE, university minor, CEFR B1, B2)	13	39.39%
I have never learnt a LOTE	1	3.03%

in Australia [87], and a follow-up study could investigate the degree to which the experience of language learning is an important variable in the way people experience the AuslanSpell application.

4.2 Data Collection Instruments

Prior to undertaking the formal evaluation of the AuslanSpell application reported in this paper, we piloted our proposed system interaction and readback tasks to ensure clarity in instructions, appropriate time allocations to the tasks and to test the appropriateness and difficulty of the readback task. The pilot involved five novice learners and included one upper-primary aged child (one of the authors' daughter), two neurodivergent users and a physically disabled user accessing the system through foot-based controls. All pilot testers were hearing, but the system and tests make no sounds and have already been deemed appropriate by our Deaf co-design group. The pilot allowed us to clarify instructions and caused us to adjust several test items that proved ambiguous or experienced playback issues. The pilot further provided informal evidence of the usability of the system by diverse audiences. This research was approved by the Human Research Ethics Committee of Monash University. The evaluation consisted of four components:

- **Familiarisation with fingerspelling:** Participants began the task by watching a simple video of a right-handed human signer slowly signing each letter of the alphabet[22]. Participants were instructed to watch the video up to 3 times before moving to the next step.
- **Free exploration and use of the AuslanSpell application:** Participants were instructed to explore the application for 15 minutes. They were given basic instructions on how to enter words and encouraged to experiment with different settings, such as adjusting the speed. Participants were free to enter any word at their own pace and to replay any given word as often as they liked. To test the exact functionality used for user evaluation, refer to this webpage[23].
- **Fingerspelling readback test:** Participants completed a custom-built 20-item test of their fingerspelling readback ability. For the exact test refer to this webpage[24] (note that the test no longer collects data). For each item, they received a video of a human signer fingerspelling a word, which could be played back a maximum of 3 times. For 10 items, they were asked to select the correct word from 5 multiple-choice options, and for the other 10 items, they were instructed to enter any letters that they saw, even if they could not work out the full word (see Appendix C.1). The videos used were made available to us by an Auslan provider and had been designed to be used in similar quizzes with beginner learners in their initial weeks of Auslan study, so they were appropriately paced for this cohort. The videos contained a mix of left- and right-handed signing models. Since word length affects the potential opportunities for participants to make errors in their fingerspelling readback, we included words from 5 to 11 letters in length. However, it is important

[22] https://www.youtube.com/watch?v=rV1KfQlRAds

[23] https://alpha.auslan-spell-web-package.cloud.edu.au

[24] http://alpha.auslan-spell-web-package.cloud.edu.au/novice-study

to note that Occhino [49] found no correlation between error rate and word length among ASL users with intermediate and above proficiency in the language.

Data from the pilot suggested that participants would be able to correctly identify the words in most multiple-choice items as well as the first and last letter of most free-text words, but would likely struggle to correctly identify the full word. This is also in line with Geer and Keane [26], who found that third-semester ASL students averaged scores of 37% on tests where they were required to correctly transcribe ASL words.

Our main focus in conducting the user study was to understand what - if any - learning was facilitated by the system. However, in order to facilitate UX improvements in later versions and to better understand user needs with this type of system, we also asked participants to complete a 16-item UX and demographic questionnaire, hosted on the Monash University's Qualtrics[25] account. The UX questionnaire began with the System Usability Scale (SUS) [9] and also included Likert-scale questions around the utility of various features of the application (e.g., speed controls or the ability to rotate views), as well as soliciting feedback on the appropriateness of the various speed settings. Participants could also leave free comments. Demographic variables collected included age, gender, handedness, education level, and a question on proficiency in languages other than English aligned with coding used in Willoughby et al. [87], to allow us to compare the demographics of our sample with that prior study of adult classroom learners of Auslan. As participants were sampled oppotunistically we do not have balanced sub-groups that could be used to investigate the impact of these variables on results. So they will not be considered in this exploratory paper. A copy of the UX and demographic questionnaire is available in Appendix C.2.

4.3 Procedure

Each data collection session lasted approximately 45 minutes. The collection took place on campus in person in a small group setting, with a member of the research team present to keep time and answer any questions participants had on using the AuslanSpell application. At the start of each session, participants were asked to confirm again that they did not know how to sign/fingerspell and then given access to a document with instructions, links, and login information to access each step of the task. The group setting was chosen to manage data collection efficiently given the popularity of certain timeslots and small research team. Participants did not interact during the sessions and completed the tasks on their own devices under test conditions. Participants were reminded of the importance of answering the test honestly and were actively supervised to ensure against cheating.

It is notable that while the study was couched entirely as being about fingerspelling readback, all participants spontaneously began signing along while watching the first video, and most continued to sign along to the words they entered when using the application in the second step. Thus, while our focus was on readback, we note that participants all took the opportunity to learn fingerspelling production as part of the learning process, a synergy also recommended by Geer [25]. When exploring with the AuslanSpell application, many participants chose to enter their own names or "the quick brown fox jumped over the lazy dog" as sample words, or repeatedly played each letter of the alphabet while attempting to master the movements involved in producing the letter.

[25]https://www.qualtrics.com

Table 2: Results for the multiple-choice words test (descriptive statistics). A perfect score is 10.

Mean	Median	Standard Deviation	Min	Max
8.2	8	1.7	3	10

4.4 Data Analysis

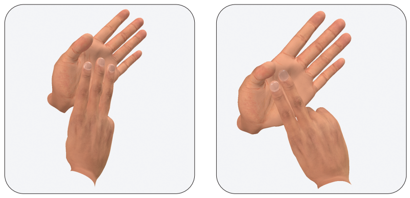
Data from the third and fourth steps was extracted from the respective platforms and processed. Results from the fingerspelling readback test were coded via a two-stage process. Stage 1 involved a simple identification of correct and incorrect answers for each word. Stage 2 involved manual inspection of errors made in both the multiple-choice questions and the transcription of the free-text words to better understand the nature of the errors. There is no agreed-upon standard for coding types of errors that signers make in fingerspelling readback tasks [26, 49, 75]. In this study, for the free-text words, we recorded whether participants produced the correct letter(s) for 1) First letter of the word, 2) Last letter of the word, 3) Middle letters of the word, and 4) Word as a whole. This coding practice reflects the known high salience of the first and last letters of fingerspelled words [84] and facilitates comparison of scores between words that have different lengths.

As this was an exploratory study to test the utility of the AuslanSpell application as a learning aid for Auslan fingerspelling, we primarily report descriptive statistics in this paper. Future research could systematically vary aspects such as the orthography of words chosen for fingerspelling or the pace and handedness of the signing models in the test videos to assess their influence on overall test performance. The usability test scores on the SUS questionnaire were calculated via the standard process outlined in [9]. Descriptive statistics were used to analyze results, and the Pearson's correlation coefficient was used to assess the relationship between the selected variables.

5 Results

5.1 Fingerspelling Readback

The primary focus of our user testing was to evaluate the degree to which interacting with our avatar-based system prepared and enabled novice signers to readback fingerspelling produced by human signers. We begin this discusion by considering the results of the multiple-choice component of the test. For each of the ten multiple-choice items, participants were asked to select the correct word from 5 options. This means that a score above 2 out of 10 on this part of the test is a performance above that of chance. As shown in Table 2, all students performed better than chance, with most recording a score of 8 out of 10 or higher. This clearly demonstrates that after 15 minutes of interacting with the AuslanSpell application, participants had built basic skills in reading back Auslan fingerspelling. Nine (out of 33) participants recorded perfect scores



(a) M (b) N

Figure 8: The similarity of M (uses 3 fingers) and N (uses 2 fingers) in Auslan.

Table 3: Results for the multiple-choice words test (per word).

Word	Length	# Correct	% Correct
healthy	7	32	97
reject	6	32	97
spray	5	32	97
narrow	6	28	85
transport	8	27	82
accessible	10	26	79
modest	6	25	76
noisy	5	25	76
honey	5	24	73
complete	7	19	58

Table 4: Results for the free-text words test (descriptive statistics). A perfect score is 10.

Mean	Median	Standard Deviation	Min	Max
3.4	4	2.3	0	7

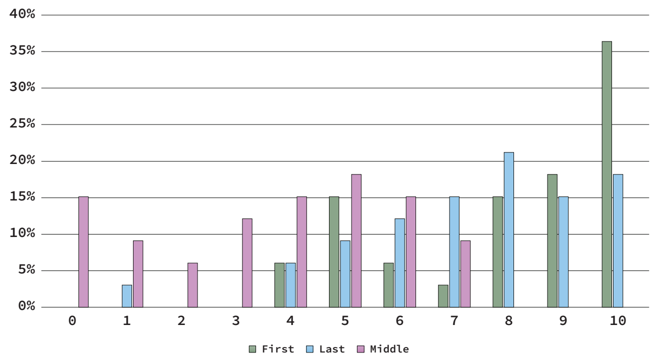


Figure 9: Number of correct letters in the free-text questions. The Y axis is the fraction of participant that got particular letter correctly. The X axis is the number of correct letters. For example, the chart shows that 36% of the participants got all 10 first letters correct, but none got all 10 middle letters.

on this part of the test, and only one participant scored less than 5 (out of 10), further supporting the conclusion that the application has been effective in building readback skills.

Looking at individual words, we see several interesting patterns. There was no relationship between word length and percent correct answers ($r = -0.11$). When looking at the incorrect options chosen by participants, it becomes clear that errors are not random. Rather, they shed light on users' developing, but still imperfect, ability to distinguish between similar letters. Thus, for example, 5 (out of 6) errors made on *narrow* chose the word *marrow*, with M and N being very similar as shown in Figure 8.

Participants were also more likely to make errors when the options included words that varied in their middle letters (see Table 3), which is in line with the finding that these letters are often articulated faster, and are thus perceptually less salient [84]. For example, 7 (out of 8) incorrect answers on *noisy* chose *nosey* (i.e., swapped the order of the vowel and S) and 12 (out of 14) errors on *complete* chose *compete* (i.e., deleted the L). All 9 errors on *honey* chose either *heavy* or *hokey* as options, again showing challenges in identifying middle letters of a word and distinguishing between vowels.

As would be expected, participants did not perform as strongly when asked to transcribe words without a cue to what they were reading back (see Table 4). On average, participants got 3.4 words fully correct, with individuals ranging from 0 to 7. Scores of below 50% are common on write-in readback tasks and underscore the challenging nature of this skill relative to multiple-choice. That students get any words completely correct is promising and suggests that they are making good initial progress in learning to identify fingerspelled words through interaction with the AuslanSpell application.

On closer inspection of the data (see Figure 9), a clear pattern emerges that participants were much more likely to correctly identify the first letter and last letter of the word than they were to correctly identify the middle letters of the word. A paired-samples t-test indicated that there was a significant decrease in correct identification of middle letters ($M = 3.42$, $SD = 2.28$) compared to first letters ($M = 8.06$, $SD = 2.09$), $t(32) = 17.56$, $p < .001$. In addition, a paired-samples t-test indicated that there was a significant decrease in correct identification of middle letters ($M = 3.42$, $SD = 2.28$) compared to last letters ($M = 7.39$, $SD = 2.13$), $t(32) = 16.80$, $p < .001$. Average correct identification of the last letter was slightly lower than average correct identification of the first letter (7.4 vs. 8.0), but a paired-samples t-test indicated that there was no significant difference in correct identification of first letters ($M = 8.06$, $SD = 2.09$) compared to last letters ($M = 7.39$, $SD = 2.13$), $t(32) = 2.13$, $p = .02$. Similarly to the multiple-choice questions, there was no correlation between word lengths and number of correct answers for any of the first, last, or middle letters or whole words. For example, *personally*, *happiness*, and *laughing* are only one letter different in length, but were correctly identified by 1, 25, and 11 students, respectively. Shipgood and Pring [75] observe in their data that words with regular spellings were more accurately read back than those with irregular spellings, which may account for some of the challenge with *laughing*, but does not explain the difficulty seen with *personally*. Further research is required to better understand the influence of word shape and frequency or other factors in mediating beginner students' ability to correctly read back fingerspelling.

5.2 User Experience

While our primary focus in testing AuslanSpell was to understand if it functions as a useful learning aid, we were also interested in gaining participant feedback on the UI and UX that could be incorporated into future iterations of the application. This feedback is reported here for transparency and also to guide UI development of similar systems in future. Participants reported high levels of satisfaction with the AuslanSpell application (see Table 5), rating it highly on both usability and perceived efficacy as a self-paced learning tool (Cronbach's Alpha was $\alpha = .76$). The mean SUS score was 81.2 (standard deviation 9.0). Scores ranged from 54 to 94 with a median of 84. Scores for consistency and speed of learning to use were particularly high, with means of 4.3 and 4.4. It is also notable that by far the lowest score on the SUS is for the first question about using the application frequently (3.5), where participants' lukewarm response may relate more to them not being Auslan learners, rather than finding the application itself problematic.

We then asked participants to rate the usefulness of five specific settings of the application, such as the ability to rotate the view or adjust the playback speed. As can be seen in Figure 10, the ability to change speed was the standout hit, with participants also valuing the ability to change and rotate views (a feature also valued highly by novice learners interacting with Fleming [21]'s Auslan digital narrative). The fact that swapping hands was not more highly regarded is surprising, but it may relate to participants reporting difficulties in the free-text responses with recognizing this feature within the application. For example:

> *"The only odd part was that the ability to change from the left handed to the right handed signing demonstrations was not as obvious as the other features, as a beginner I hadn't noticed that the signing was left handed until I messed around with rotating the hands, and it took me a while to go to the settings to change it and find the other options."*

> *"I had trouble realising I could reset the hands from left to right until I was told. I think a popup at the beginning asking the handed-ness of the user would be good for it to automatically then chose a left or right hand. I also had trouble figuring out if I was signing correctly or mirrored."*

Participants were also asked to comment on the appropriateness of the fastest and slowest speed settings for fingerspelling playback. For reference, the 5 speed options are x0.25 (0.37 letters/second), x0.5 (0.74 letters/second), x0.75 (1.11 letters/second), x1 (1.49 letters/second, the speed at which the original data was signed), and x1.5 (2.23 letters/second). There was strong consensus that the slowest possible speed was appropriate, with only 1 user regarding it as too fast and only 3 regarding it as too slow. For the fastest possible setting, 3 users (9%) found it still too slow, but 14 (42%) found it too fast. Given that these users are completely new to Auslan fingerspelling, it may be that they are not best-placed to identify the upper bounds of useful playback speeds. Further testing of the application with more experienced signers would help us identify the top desired speed, which may be in excess of x2 the recorded pace.

Table 5: Results from the SUS questionnaire (descriptive statistics). A perfect score is 100.

Mean	Median	Standard Deviation	Min	Max
81.2	84	9	54	94

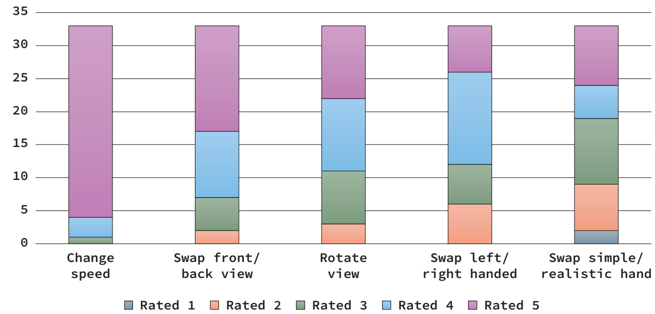


Figure 10: Feature ratings (Likert 1 to 5). The Y axis is the number of participants.

The final closed questions on the user experience questionnaire asked whether users found the animations in the application realistic and whether it helped them learn to fingerspell. These questions scored similarly, with average ratings of 3.9 and 4.1, respectively. There was a weak correlation between participants' answers on these two questions ($r = 0.31$), suggesting that participants are largely judging these features independently. Participants were given the option at the end of the questionnaire to share any concluding thoughts in an open-ended question, which was taken up by 22 participants. Fourteen of these comments included suggestions for improving the UI/UX (63%), three focused on how the signing is presented (14%), and nine included details of the aspects of the application that they particularly liked (41%, note that some comments addressed multiple themes). Representative comments include:

> *"The ability to rotate was really useful but difficult to use."* (UI/ UX comment and comment on aspect of the application that they particularly liked)
>
> *"Maybe [include] notes on if there are variations in the signs people would use."* (comment on how the signing was presented)
>
> *"The app could include a quick-start guide at the beginning, or a link to a video tutorial. First, it should explain that you need to type a letter and press Enter before it plays. Second, it should cover how to swap left/right, rotate, and mirror views, since in sign language, which hand is considered the left or right can be a bit confusing. Overall, this app is very useful, especially with the rotatable hand model and adjustable playback speed, which help people learn sign language quickly. If a vocabulary list and quiz feature could be added in the future, this app could become the Duolingo of sign language. It's fantastic."*(UI/ UX comment and comment on aspect of the application that they particularly liked)

6 Discussion

6.1 Identifying Target Forms

A core challenge in developing sign language applications is the tension around what the target forms of each letter or sign should be. A recent report on AI-based signing solutions[26] identified lack of reflection on this issue as a core problem with much current technology, with the consequence that the signing output produced is incorrect and/or incomprehensible.

A challenge in producing Auslan resources is the relatively high rate of sociolinguistic variation in the language [35]. What this means in practice is that while there is general agreement on the way to form each letter in its citation form (i.e., the form that appears in dictionaries and learning resources) the way individuals sign a string of letters may deviate from these norms [15]. Initially it was our goal to capture these deviations for enhanced realism, but we received feedback from Deaf educators through the co-design process that they did not want students to be taught bad habits.

Where and how to introduce sociolinguistic variation in language learning courses is a perennial point of debate in the language teaching literature, but many courses tend to leave it for upper level units [41]. The desire to respect the wishes of Deaf educators and the fact that our application will likely be of most use to beginner and intermediate students motivated us to clean the signs and ensure that they resembled citation forms at the apex of their articulation.

As noted by Fleming [21], co-designing animated sign language resources with Deaf people is also important because it helps to ensure linguistic accuracy and that users can treat the application as an authoritative source of information on how to form the sign. Beginner signers report that one challenge they have in accessing sign language content on YouTube[27] and similar platforms is that they are not able to tell which creators are good language models or to distinguish what is natural variation in how signs are formed from errors made by content creators who are not themselves fluent signers [89]. This was an important reason why we carefully selected the signing models for this study (rather than creating a crowdsourced application) and conducted iterative cleaning of the animations based on feedback from Deaf educators to ensure students received appropriate animations from the application. Poor quality and limited acceptance of signing avatar technology is an enormous issues in the field currently [4, 47], with our project allocating significant resources to co-design from inception to cleaning to build a product acceptable to Deaf Auslan educators. Given the very small movements involved in fingerspelling and the inherent limitations of current motion capture and computer vision techniques, our experience also reinforces how vital it is that fluent Deaf signers are involved in validating sign language resources (and ideally that these people lead the development of resources) to ensure that correct forms are used [48].

6.2 Technical Contribution

Unlike many existing sign language datasets, the AuslanSpell dataset offers substantial diversity across both signers (multiple signers per sign) and signs (multiple signs per signer). This makes it a valuable resource for building advanced data-driven fingerspelling models, applicable to both recognition and generation tasks. To our knowledge, AuslanSpell will also be the first publicly available high-quality 3D dataset for Auslan fingerspelling. Prior research highlights the importance of depth-awareness and viewing angle in sign language comprehension for both humans [83] and machine learning systems [23, 60]. This suggests that effective representations must incorporate 3D information, as failing to do so risks reduced accuracy in realistic settings where signs are not always produced from a frontal perspective.

The success of data-driven machine learning approaches has fueled a growing demand for data. However, data collection within minority language communities raises particular ethical concerns. The Deaf community is considered sensitive, and the limited size of such populations makes it difficult to ensure anonymity [8]. Some publicly available sign language datasets were compiled without obtaining proper informed consent, especially those derived from online platforms. In contrast, all contributors to the AuslanSpell dataset were informed of how their data will be utilized and were compensated for their participation. Furthermore, anonymity is better preserved compared to conventional video datasets, since motion capture recordings do not reveal signers' visual identities.

The application accounts for users with varying levels of technical access, acknowledging differences in both hardware and software platforms. This includes support for a wide range of devices (e.g., smartphones, tablets, and computers) as well as scenarios with limited or low-bandwidth Internet connectivity. By ensuring broad compatibility, AuslanSpell aims to maximize accessibility and reach the widest possible user base.

6.3 Educational Contribution

AuslanSpell represents the first technology of its kind for Auslan fingerspelling and serves as an important initial step toward creating interactive and engaging tools for Auslan education. After brief interaction, novice signing participants performed above chance on beginner multiple-choice stimuli and correctly identified first and last letters more often than middle letters in free-text tasks. While these results are preliminary, they suggest that the application is a useful tool for helping develop fingerspelling readback ability. As the test materials were videos of humans signing the results also suggest that skills developed reading back fingerspelling from the application's animations are transferable to reading back human fingerspelling. Very limited research has taken place to date on how hearing adult learners develop their fingerspelling abilities and none of it has focussed on Auslan. Data from our user testing with novice learners reinforces the salience of initial letters in word comprehension and that learners are often making non-random errors (i.e., they confuse signs with similar forms). These findings align with the existing literature, where fingerspelling tests have generally been taken by more experienced sign language learners [24, 26, 49]. In this study we chose to focus on learners who are complete beginners, even though our target audience is students who are enrolled in beginner to intermediate programs. This was a deliberate choice, to control variables in what is a small proof-of-concept study. As Thoryk [81] notes, any study that attempts

[26] https://www.mctd.ac.uk/bsl-not-for-sale-deaf-ai-procurement/
[27] https://www.youtube.com

to measure the efficacy of a sign language learning resource with students who are already enrolled in classes will have to contend with complex interactions of student prior knowledge, motivation and study habits that make it challenge to measure the true efficacy of the resource. We now plan to test the resource with students enrolled in sign language programs to assess their learning gains, as discussed in the future work section.

6.4 Limitations

During the development of the AuslanSpell dataset and application, we identified several areas for potential improvement. To our knowledge, AuslanSpell is the first dataset to employ Manus Prime 3 Mocap gloves for capturing fingerspelling handshapes in sign language. These gloves were originally designed for generating animation data, which typically does not demand the same level of precision required for representing the fine, varied, and intricate movements of fingers in sign language. Their use in this context is therefore both novel and experimental. While the gloves show strong potential, they also present challenges. Measurement accuracy is highly dependent on glove fit and duration of use: tighter fits generally yield better accuracy, whereas prolonged wear tends to reduce precision as the gloves shift from their calibrated position.

The current scope of the AuslanSpell dataset is limited to Auslan fingerspelling. Alphabet topologies vary across sign languages – for example, ASL employs a one-handed alphabet, while BSL and Auslan use two hands – introducing differences in collision constraints and co-articulations. Nonetheless, our data collection methodology can be extended to support additional sign languages.

A further limitation arises from our emphasis on letter articulation accuracy. At present, the AuslanSpell application renders pre-recorded and validated animations of individual letters, which constrains the natural diversity and realism of articulations. Data-driven generative models, by contrast, can introduce stochastic variation that better reflects natural signing, though they require careful design to ensure that outputs remain valid representations of each sign. In future work, our aim is to integrate high-quality hand keypoints and meshes (e.g., MANO [64] and SMPL-X [52]) to explore such generative approaches while maintaining precise control over articulation accuracy. Further factors contributing to realism that will be explored in future work include a full-body avatar to include subtle aspects such as facial expressions and mouthing.

The production of each and every letter is formally correct but could be linguistically odd. However, fingerspelling co-articulation effects in Auslan are relatively poorly understood. Anecdotally, common letter sequences such as I-N-G take on their own shape in fluent Auslan fingerspelling. Other letter sequences, such as E-G-G and H-O-W have over time become so-called lexicalised fingerspellings where the sign has its origin in, but no longer closely mirrors the fingerspelling of the word [68]. A future version of the application could specifically record these transitions for a more natural readback experience. Further data recordings from a wider group of signers would also allow us to better understand what is constant and open to variation in Auslan fingerspelling and could potentially lead to an application with different settings along a spectrum from citation forms of each letter to letters that more naturally blend into one another.

In this article we have shown that AuslanSpell is an effective learning resource for well-educated university students from diverse language backgrounds[28]. This proof-of-concept testing now needs to be extended to other user groups to generalise findings on learning outcomes with more diverse learners. As discussed below, our priority is to extend testing to Auslan learners enrolled at a large TAFE (equivalent to US Community College). Willoughby et al. [87] have documented that TAFE learners of Auslan have highly varied educational profiles and include a significant cohort of learners who are deaf themselves or have deaf family or friends. User testing with this cohort will allow us to observe the interaction of these demographic and motivational variables on user experience and test performance, as well as exploring the utility of the application for learners with differing levels of prior knowledge of Auslan.

6.5 Future Work

There is a clear knowledge gap in the field that further work could explore more systematically. Having completed basic user testing, we now intend to conduct experiments where we measure the learning gains of more experienced signers interacting with the application over a number of weeks. Given that Geer and Keane [26] saw strong learning gains that were sustained over time from their ASL fingerspelling training program, we would expect that our application will lead to similar learning gains. Design of this experiment will, however, require careful calibration of participants baseline fingerspelling readback ability, as well as significant investment of resources to retain participants over multiple data collection sessions, both of which were reasons why we did not collect such data as part of this study.

In this study we tested fingerspelling readback ability using a convenience sample of videos created by an Auslan provider for their own novice signers testing. Future work could more systematically vary factors such as word frequency, regularity of spelling, pace or handedness of signing models in the test videos to explore their effects on Auslan learners – as they have been shown to influence learner fingerspelling performance in other studies [26, 75]. This would also provide an opportunity to explore the relative effect of each of these factors, which to date have been studied largely in isolation. Future work could also explore psycholinguistic aspects of learning to readback fingerspelling, such as the effect of variables such as handedness, English spelling ability, educational attainment or prior language learning experience on performance over time. There are many promising avenues for building novel applications upon the core AuslanSpell technology. Our immediate future work includes a major application extension that will offer an Auslan fingerspelling course allowing educators to design custom quizzes to assess students' readback skills. This extension will also incorporate feedback on the UI design to increase the visibility of settings such as changing the handedness of the signing model.

[28] While many students were linguistics majors there was no clear difference in performance from students with/without a linguistics background or between the performance of linguistics students and pilot testers without a linguistics background.

7 Conclusion

This work introduces an iterative collaborative design process and evaluation of AuslanSpell, a novel application that aims to improve the proficiency in Auslan fingerspelling. The work highlights the importance of working with Deaf co-designers from inception to completion to produce signing avatar technology that is high quality and addresses real teaching or learning needs. The application provides innovative modes of interaction with fingerspelling content, and our findings highlight its potential to enhance both comprehension and learners' engagement. In an exploratory study with 33 novice signers, after brief interaction with AuslanSpell, participants performed above chance on beginner multiple-choice tasks, correctly identified first and last letters more often than middle letters in free-text tasks, and reported high satisfaction – especially with features such as adjustable signing speeds and rotatable views. Our next step is to integrate this core functionality into user-facing tools that can be embedded within Auslan training programs across Australia, with the aim of substantially advancing the quality of Auslan education. Additional user evaluations will further guide the development, supporting the creation of a technology that significantly improves learning outcomes.

Acknowledgments

We gratefully acknowledge funding for this work from the Telematics Trust and Australian Government through the Australian Research Council. Dr Kalin Stefanov is the recipient of an Australian Research Council Discovery Early Career Researcher Award (project number DE230100049). A/Prof Louisa Willoughby is the recipient of an Australian Research Council Mid-Career Industry Fellowship (project number IM240100062). Additionally we acknowledge in-kind support for the project from Deaf Connect, Deaf Australia, and Melbourne Polytechnic Auslan Program.

References

[1] Carol T Akamatsu. 1985. Fingerspelling formulae: A word is more or less than the sum of its letters. *SLR* 83 (1985), 126–132.

[2] 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 *Proceedings of the European Conference on Computer Vision*. 35–53.

[3] Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. 2021. BBC-Oxford British Sign Language Dataset. arXiv:2111.03635 [cs.CV]

[4] Robin Angelini. 2023. Contrasting Technologists' and Activists' Positions on Signing Avatars. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 566, 6 pages. doi:10.1145/3544549.3583946

[5] Vasileios Baltatzis, Rolandos Alexandros Potamias, Evangelos Ververas, Guanxiong Sun, Jiankang Deng, and Stefanos Zafeiriou. 2024. Neural Sign Actors: A Diffusion Model for 3D Sign Language Production from Text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1985–1995.

[6] Souad Baowidan. 2021. Improving realism in automated fingerspelling of American Sign Language. *Machine Translation* 35, 3 (2021), 387–404.

[7] Souad Baowidan. 2024. A Real-Time Collision Detection and Avoidance Algorithm for Fingerspelling Animation. In *Sign Language Machine Translation*, Andy Way, Lorraine Leeson, and Dimitar Shterionov (Eds.). Springer Nature Switzerland, 267–290. doi:10.1007/978-3-031-47362-3_11

[8] Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring Collection of Sign Language Datasets: Privacy, Participation, and Model Performance. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*. Association for Computing Machinery, Article 33, 14 pages. doi:10.1145/3373625.3417024

[9] John Brooke. 1996. SUS: A 'Quick and Dirty' Usability Scale. *Usability Evaluation In Industry* 189, 194 (1996), 4–7.

[10] Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural Sign Language Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7784–7793.

[11] Mary Joanna Carr and Anne Pauwels. 2006. *Boys and Foreign Language Learning: Real Boys Don't Do Languages*. Palgrave Macmillan.

[12] Xin Chen, Biao Jiang, Wen Liu, Zilong Huang, Bin Fu, Tao Chen, and Gang Yu. 2023. Executing your Commands via Motion Diffusion in Latent Space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18000–18010.

[13] Michael Clyne. 2005. *Australia's Language Potential*. UNSW Press.

[14] K. Cormier, A. Schembri, and M.E. Tyrone. 2008. One hand or two? Nativisation of fingerspelling in ASL and BANZSL. *Sign Language and Linguistics* 11, 1 (2008), 3–44. doi:10.1075/sl&l.11.1.03cor

[15] Donovan Cresdee and Trevor Johnston. 2014. Using corpus-based research to inform the teaching of Auslan (Australian Sign Language) as a second language. In *Teaching and Learning Signed Languages: International Perspectives and Practices*, David McKee, Russell S. Rosen, and Rachel McKee (Eds.). Palgrave Macmillan, 85–110.

[16] 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*. 2735–2744.

[17] Sarah Ebling, Sarah Johnson, Rosalee Wolfe, Robyn Moncrief, John McDonald, Souad Baowidan, Tobias Haug, Sandra Sidler-Miserez, and Katja Tissi. 2017. Evaluation of Animated Swiss German Sign Language Fingerspelling Sequences and Signs. In *Universal Access in Human-Computer Interaction*, M Antona and C Stephanidis (Eds.). Springer, 3–13. doi:10.1007/978-3-319-58703-5_1

[18] Sarah Ebling, Rosalee Wolfe, Jerry Schnepp, Souad Baowidan, John McDonald, Robyn Moncrief, Sandra Sidler-Miserez, and Katja Tissi. 2015. Synthesizing the finger alphabet of Swiss German Sign Language and evaluating the comprehensibility of the resulting animations. In *Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies*. Association for Computational Linguistics, 10–16. doi:10.18653/v1/W15-5103

[19] R. Elliott, J. R. W. Glauert, V. Jennings, and J. R. Kennaway. 2004. An Overview of the SiGML Notation and SiGMLSigning Software System. In *Proceedings of the 4th International Conference on Language Resources and Evaluation: Sign Language Processing Satellite Workshop*. 98–104.

[20] Sen Fang, Chunyu Sui, Yanghao Zhou, Xuedong Zhang, Hongbin Zhong, Yapeng Tian, and Chen Chen. 2025. SignDiff: Diffusion Model for American Sign Language Production. arXiv:2308.16082 [cs.CV]

[21] Rebecca Jane Fleming. 2024. *Digital Narratives: Design for Motivational Impact in Sign Language Learning*. Ph. D. Dissertation. Monash University.

[22] Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus H Piater, and Hermann Ney. 2012. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. In *Proceedings of the International Conference on Language Resources and Evaluation*, Vol. 9. 3785–3789.

[23] Liqing Gao, Lei Zhu, Senhua Xue, Liang Wan, Ping Li, and Wei Feng. 2023. Multi-View Fusion for Sign Language Recognition through Knowledge Transfer Learning. In *Proceedings of the ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. Association for Computing Machinery, Article 1, 9 pages. doi:10.1145/3574131.3574434

[24] Leah Geer and Jonathan Keane. 2014. Exploring factors that contribute to successful fingerspelling comprehension. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*. European Language Resources Association, 1905–1910.

[25] Leah C Geer. 2019. Teaching L2/Ln sign language fingerspelling. In *The Routledge Handbook of Sign Language Pedagogy*, Russell S Rosen (Ed.). Routledge.

[26] Leah C Geer and Jonathan Keane. 2018. Improving ASL fingerspelling comprehension in L2 learners with explicit phonetic instruction. *Language Teaching Research* 22, 4 (2018), 439–457. doi:10.1177/1362168816686988

[27] Manfred Georg, Garrett Tanzer, Esha Uboweja, Saad Hassan, Maximus Shengelia, Sam Sepah, Sean Forbes, and Thad Starner. 2025. FSboard: Over 3 Million Characters of ASL Fingerspelling Collected via Smartphones. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13897–13906.

[28] Debtanu Gupta, Shubh Maheshwari, Sai Shashank Kalakonda, Manasvi Vaidyula, and Ravi Kiran Sarvadevabhatla. 2023. DSAG: A Scalable Deep Framework for Action-Conditioned Multi-Actor Full Body Motion Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4300–4308.

[29] Thomas Hanke. 2004. HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts. In *Proceedings of the International Conference on Language Resources and Evaluation*. European Language Resources Association, 1–6.

[30] Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. Extending the Public DGS Corpus in Size and Depth. In *Proceedings of the 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. 75–82.
[31] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *Proceedings of the International Conference on Neural Information Processing Systems*. Curran Associates Inc., Article 574, 12 pages.
[32] Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li. 2018. Attention-Based 3D-CNNs for Large-Vocabulary Sign Language Recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 29, 9 (2018), 2822–2832.
[33] Jie Huang, Wengang Zhou, Qilin Zhang, Houqiang Li, and Weiping Li. 2018. Video-based Sign Language Recognition without Temporal Segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
[34] Jeremy Zhengqi Huang, Jaylin Herskovitz, Liang-Yuan Wu, Cecily Morrison, and Dhruv Jain. 2025. Weaving Sound Information to Support Real-Time Sensemaking of Auditory Environments: Co-Designing with a DHH User. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 75, 18 pages. doi:10.1145/3706598.3714268
[35] Trevor Johnston and Adam Schembri. 2007. *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press.
[36] Hamid Reza Vaezi Joze and Oscar Koller. 2019. MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language. arXiv:1812.01053 [cs.CV]
[37] Tomasz Kapuscinski, Mariusz Oszust, Marian Wysocki, and Dawid Warchol. 2015. Recognition of Hand Gestures Observed by Depth Cameras. *International Journal of Advanced Robotic Systems* 12, 4 (2015), 36.
[38] Jonathan Keane and Diane Brentari. 2015. Fingerspelling*: Beyond Handshape Sequences. In *The Oxford Handbook of Deaf Studies in Language*. Oxford University Press. doi:10.1093/oxfordhb/9780190241414.013.10
[39] Diederik P Kingma. 2013. Auto-Encoding Variational Bayes. arXiv:1312.6114 [cs.CV]
[40] Carly Leannah, Athena S. Willis, and Lorna C. Quandt. 2022. Perceiving fingerspelling via point-light displays: The stimulus and the perceiver both matter. *PLOS ONE* 17, 8 (2022), e0272838. doi:10.1371/journal.pone.0272838
[41] Betty Lou Leaver and Boris Shekhtman. 2002. *Principles and practices in teaching Superior-level language skills: not just more of the same.* Cambridge University Press, 3—-33.
[42] Lorraine Leeson, Sarah Sheridan, Katie Cannon, Tina Murphy, Helen Newman, and Heidi Veldheer. 2020. Hands in Motion : Learning to Fingerspell in Irish Sign Language (ISL). *Journal of the Irish Association for Applied Linguistics* 11 (2020), 120–141. doi:10.35903/teanga.v11i1.199
[43] Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. 2020. Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1459–1469.
[44] Natalia Lusin, Terri Peterson, Christine Sulewski, and Rizwana Zafer. 2023. *Enrollments in Languages Other Than English in US Institutions of Higher Education, Fall 2021.* Technical Report. Modern Language Association.
[45] John McDonald, Rosalee Wolfe, Souad Baowidan, Ningshan Guo, Sarah Johnson, and Robyn Moncrief. 2017. Using N-Gram Analytics to Improve Automatic Fingerspelling Generation. *Linguistics and Literature Studies* 5, 3 (2017), 187–197. doi:10.13189/lls.2017.050306
[46] David McKee, Russell S. Rosen, and Rachel McKee. 2014. Introduction. In *Teaching and Learning Signed Languages: International Perspectives and Practices*, David McKee, Russell S. Rosen, and Rachel McKee (Eds.). Palgrave Macmillan, 1–7.
[47] Maartje De Meulder. 2025. Deaf in AI: AI language technologies and the erosion of linguistic rights. arXiv:2505.02519 [cs.CY] https://arxiv.org/abs/2505.02519
[48] Maartje De Meulder. 2025. Deaf in AI: AI language technologies and the erosion of linguistic rights. arXiv:2505.02519 [cs.CY]
[49] Corrine Occhino, Ryan Lidster, Leah C. Geer, Jason Listman, and Peter C. Hauser. 2024. Development of the American Sign Language Fingerspelling and Numbers Comprehension Test (ASL FaN-CT). *Language Testing* 41, 1 (2024), 135–161. doi:10.1177/02655322231179494
[50] Peerawat Pannattee, Wuttipong Kumwilaisak, Chatchawarn Hansakunbuntheung, Nattanun Thatphithakkul, and C.-C. Jay Kuo. 2024. American Sign language fingerspelling recognition in the wild with spatio temporal feature extraction and multi-task learning. *Expert Systems with Applications* 243 (2024), 122901. doi:10.1016/j.eswa.2023.122901
[51] Carol J Patrie and Robert E Johnson. 2011. *Fingerspelled Word Recognition Through Rapid Serial Visual Presentation.* DawnSignPress.
[52] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. 2019. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10975–10985.
[53] Michael Peters, Stian Reimers, and John T. Manning. 2006. Hand preference for writing and associations with selected demographic and behavioral variables in 255,100 subjects: The BBC internet study. *Brain and Cognition* 62, 2 (2006), 177–189. doi:10.1016/j.bandc.2006.04.005
[54] Mathis Petrovich, Michael J Black, and Gül Varol. 2021. Action-Conditioned 3D Human Motion Synthesis with Transformer VAE. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10985–10995.
[55] Mathis Petrovich, Michael J Black, and Gül Varol. 2022. TEMOS: Generating diverse human motions from textual descriptions. In *Proceedings of the European Conference on Computer Vision*. 480–497.
[56] David Quinto-Pozos. 2011. Teaching American Sign Language to hearing adult learners. *Annual Review of Applied Linguistics* 31 (2011), 137–158. doi:10.1017/S0267190511000195
[57] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125 [cs.CV]
[58] Oline Ranum, Gomèr Otterspeer, Jari I. Andersen, Robert G. Belleman, and Floris Roelofsen. 2024. 3D-LEX v1.0 – 3D Lexicons for American Sign Language and Sign Language of the Netherlands. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*. 290–301.
[59] Oline Ranum, David R. Wessels, Gomer Otterspeer, Erik J. Bekkers, Floris Roelofsen, and Jari I. Andersen. 2024. The NGT200 Dataset: Geometric Multi-View Isolated Sign Recognition. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop*, Sharvaree Vadgama, Erik Bekkers, Alison Pouplin, Sekou-Oumar Kaba, Robin Walters, Hannah Lawrence, Tegan Emerson, Henry Kvinge, Jakub Tomczak, and Stephanie Jegelka (Eds.), Vol. 251. Proceedings of Machine Learning Research, 286–302.
[60] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2020. Hand sign language recognition using multi-view hand skeleton. *Expert Systems with Applications* 150 (2020), 113336. doi:10.1016/j.eswa.2020.113336
[61] Charlotte M. Reed, Lorraine A. Delhorne, Nathaniel I. Durlach, and Susan D. Fischer. 1990. A Study of the Tactual and Visual Reception of Fingerspelling. *Journal of Speech and Hearing Research* 33, 4 (1990), 786–797.
[62] Octavian Robinson and Jonathan Henner. 2018. Authentic Voices, Authentic Encounters: Cripping the University Through American Sign Language. *Disability Studies Quarterly* 38, 4 (2018). doi:10.18061/dsq.v38i4.6111
[63] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10684–10695.
[64] Javier Romero, Dimitrios Tzionas, and Michael J. Black. 2017. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *ACM Transactions on Graphics* 36, 6, Article 245 (2017), 17 pages. doi:10.1145/3130800.3130883
[65] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *Advances in Neural Information Processing Systems* 35 (2022), 36479–36494.
[66] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The Graph Neural Network Model. *IEEE Transactions on Neural Networks* 20, 1 (2009), 61–80. doi:10.1109/TNN.2008.2005605
[67] Adam Schembri, Jordan Fenlon, Ramas Rentelis, Sally Reynolds, and Kearsy Cormier. 2013. Building the British Sign Language Corpus. (2013).
[68] Adam Schembri and Trevor Johnston. 2007. Sociolinguistic variation in the use of fingerspelling in Australian Sign Language: A pilot study. *Sign Language Studies* 7, 3 (2007), 319–347.
[69] Amy Louise Schwarz. 2000. The perceptual relevance of transitions between segments in the fingerspelled signal. (2000).
[70] Zed Sevcikova Sehyr, Naomi Caselli, Ariel M Cohen-Goldberg, and Karen Emmorey. 2021. The ASL-LEX 2.0 Project: A Database of Lexical and Phonological Properties for 2,723 Signs in American Sign Language. *The Journal of Deaf Studies and Deaf Education* 26, 2 (2021), 263–277. doi:10.1093/deafed/enaa038
[71] Xin Shen, Heming Du, Hongwei Sheng, Shuyun Wang, Hui Chen, Huiqiang Chen, Zhuojie Wu, Xiaobiao Du, Jiaying Ying, Ruihan Lu, et al. 2024. MM-WLAuslan: Multi-View Multi-Modal Word-Level Australian Sign Language Recognition Dataset. (2024). arXiv:2410.19488 [cs.CV]
[72] Xin Shen, Shaozu Yuan, Hongwei Sheng, Heming Du, and Xin Yu. 2023. Auslan-Daily: Australian Sign Language Translation for Daily Communication and News. In *Proceedings of the International Conference on Neural Information Processing Systems*. Curran Associates Inc., Article 3527, 15 pages.
[73] Bowen Shi, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2021. Fingerspelling Detection in American Sign Language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 4164–4173. doi:10.1109/CVPR46437.2021.00415
[74] Bowen Shi, Aurora Martinez Del Rio, Jonathan Keane, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2019. Fingerspelling Recognition in the Wild With Iterative Visual Attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5399–5408. doi:10.1109/ICCV.2019.00550
[75] L. E. Shipgood and T. R. Pring. 1995. The difficulties of learning fingerspelling: an experimental investigation with hearing adult learners. *International Journal of Language & Communication Disorders* 30, 4 (1995), 401–416. doi:10.3109/

13682829509087241
[76] Ozge Mercanoglu Sincan and Hacer Yalim Keles. 2020. AUTSL: A Large Scale Multi-modal Turkish Sign Language Dataset and Baseline Methods. *IEEE access* 8 (2020), 181340–181355.
[77] The Deaf Society. 2015. *Auslan language services and the NDIS: Ensuring quality in an environment of workplace shortage*. Retrieved Aug 19, 2025 from https://www.aph.gov.au/DocumentStore.ashx?id=a43c6757-a24f-41e5-adc9-9dde04edcfa5
[78] Stack Overflow. 2025. *Stack Overflow Developer Survey 2025 – Technology*. Retrieved Sep 6, 2025 from https://survey.stackoverflow.co/2025/technology#most-popular-technologies-language
[79] Shengeng Tang, Jiayi He, Dan Guo, Yanyan Wei, Feng Li, and Richang Hong. 2025. Sign-IDD: Iconicity Disentangled Diffusion for Sign Language Production. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence and 37th Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI Press, Article 808, 9 pages. doi:10.1609/aaai.v39i7.32781
[80] Guy Tevet, Brian Gordon, Amir Hertz, Amit H Bermano, and Daniel Cohen-Or. 2022. MotionCLIP: Exposing Human Motion Generation to CLIP Space. In *Proceedings of the European Conference on Computer Vision*. Springer, 358–374.
[81] Robertta Thoryk. 2010. A Call for improvement: The need for research-based materials in American Sign Language education. *Sign Language Studies* 11, 1 (2010), 100–120. doi:10.1353/sls.2010.0002
[82] David Uthus, Garrett Tanzer, and Manfred Georg. 2023. YouTube-ASL: A Large-Scale, Open-Domain American Sign Language-English Parallel Corpus. In *Proceedings of the International Conference on Neural Information Processing Systems*. Curran Associates Inc., Article 1264, 19 pages.
[83] Freya Watkins, Diar Abdlkarim, Bodo Winter, and {Robin L.} Thompson. 2024. Viewing angle matters in British Sign Language processing. *Scientific Reports* 14, 1 (2024). doi:10.1038/s41598-024-51330-1
[84] Nkenge Wheatland, Ahsan Abdullah, Michael Neff, Sophie Jörg, and Victor Zordan. 2016. Analysis in support of realistic timing in animated fingerspelling. In *2016 IEEE Virtual Reality (VR)*. 309–310. doi:10.1109/VR.2016.7504777
[85] Merideth Wightman. 2020. Gender Differences in Second Language Learning: Why They Exist and What We Can Do About It. (2020).
[86] Sherman Wilcox. 1992. *The phonetics of fingerspelling*. John Benjamins.
[87] Louisa Willoughby, Adam Schembri, and Jess Kruk. 2025. Sign language for all? Profile and retention of students in a beginner sign language program. *The Modern Language Journal* 109, 1 (2025), 216–233. doi:10.1111/modl.12987
[88] Louisa Willoughby and Cathy Sell. 2019. Studying a sign language: What are hearing adults doing outside of class? *Sign Language Studies* 19, 3 (2019), 453–478. doi:10.1353/sls.2019.0004
[89] Louisa Willoughby and Cathy Sell. 2024. Online videos for self-directed second language learning. *Language Learning & Technology* 28 (2024), 1–15.
[90] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*. European Language Resources Association.
[91] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. 2024. MotionDiffuse: Text-Driven Human Motion Generation With Diffusion Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 6 (2024), 4115–4128. doi:10.1109/TPAMI.2024.3355414

A AuslanSpell Dataset

A.1 Words

Table 6: List of 93 words used during the dataset recordings, including words with prefixes, words with suffixes, names, and places.

Prefixes	Suffixes	Names	Places
antecedent	able	bahhari	beijing
antifascist	additive	burridge	bunnerong
cooperate	breach	chan	cabramatta
describe	brown	garcia	egypt
disdain	caution	khoa	ezra
embolden	contrition	nguyen	iroquois
entrap	culmination	mark	jakarta
extraverted	decal	romanowski	hamburg
foreground	delightful	willoughby	nunawading
implicit	evenhandedness		somalia
indecent	falsifying		tokyo
interlude	ferment		yarra
irrational	fewer		wurundjeri
misnomer	final		
nonevent	fishes		
overeager	freedom		
predetermined	frequency		
reintroduce	friendless		
semiconductor	fusion		
stripey	generate		
subsume	gritty		
superdiverse	ground		
telepathic	happily		
transcontinental	hungriest		
unsure	hour		
underspecify	iterative		
uptown	lox		
	mummy		
	manure		
	nighty		
	odorous		
	phantom		
	quantity		
	radial		
	sample		
	slowly		
	spouting		
	shock		
	spurious		
	suffixes		
	terrific		
	truth		
	warm		
	zoned		

A.2 Letters

Table 7: Letter counts in the 93 words used during the dataset recordings.

Letters	Counts
a	55
b	13
c	23
d	30
e	80
f	17
g	19
h	18
i	60
j	3
k	6
l	23
m	22
n	61
o	43
p	16
q	3
r	62
s	35
t	49
u	37
v	7
w	9
x	3
y	15
z	2

A.3 Signing Models

Table 8: Details for each person who was a signing model, including number of sessions contributed to the dataset and handedness.

ID	Sessions	Handedness
01	2	left-handed
02	5	right-handed
03	6	right-handed
04	3	right-handed
05	6	right-handed
06	3	right-handed
07	3	right-handed
08	3	left-handed

B AuslanSpell Application

This appendix provides technical details on our attempt to increase developer experience with the web version of the application. We are also developing comprehensive documentation for the application.

B.1 Customization

The implementation is not limited to Auslan signs. Through the API, developers can provide new 3D models, allowing additional languages. We welcome external contributions towards deploying 3D models for different sign languages.

The implementation applies a dependency injection pattern[29] to allow developers to inject their own `HandsController` implementations (used for manipulating the hand animations) and `HandsScene` (used for manipulating the Three.js[30] 3D scene containing the hands). We intend to utilize dependency injection and provide separate controllers for data-driven generative models-based production of Auslan fingerspelling animations.

The configuration of a `HandsController` is divided into two levels – global and model. The model configuration, as the name implies, is used on a model-by-model basis, while the global configuration includes all attributes of model configuration and is applied to every model within a `HandsController` (i.e., global configuration provides default values for every attribute, allowing developers to omit definitions). The model configuration takes precedence over the global configuration; the priority is: Model configuration → Global configuration → Default configuration. This priority allows for granular configuration without sacrificing simplicity of use.

Application components (e.g., play/pause buttons, speed multiplier buttons, and animation scrubbers) were designed for 1) Ease of use in modern frameworks, where component initialization and view building (DOM) are decoupled, and 2) Reduction of implementation verbosity.

To address the first goal, the implementation separates initialization, building, and connection of components. However, this solution presents an issue where developers could call methods that manipulate the view before it is initialized. To prevent errors while allowing intuitive use of components, the implementation uses build hooks. Internally, before accessing the view, a component controller checks whether the view has been initialized, and if not, it adds a build hook that performs the desired action immediately after the view has been built, allowing for code like:

```
// Before the DOM is available
button
    .lock()
    .outline()
    .addBuildHook(() => console.log("built!"))
    .addConnectHook(() => console.log("connected!"))

...

// Once the DOM is available
button
    // Locks the button, then outlines the button
    .buildView()
    // Appends the button view to the document body
    .connectTo(document.body)
```

This feature is especially important in frameworks like React[31], where developers do not have access to the DOM until the React component has been mounted. To address the second goal, the implementation extracts components into separate npm packages, simplifying the codebase to include only relevant animation logic.

The implementation allows developers to initialize multiple instances of `HandsController` and `HandsScene`. The centralized model loading applies a singleton pattern[32]. It stores models in JavaScript object memory to allow instant access. Internally, when `loadModel` (public API for accessing models) is called, the singleton handles three cases:

- **The model is already loaded:** Returns immediately a cloned version of the model.
- **The model has not started loading:** Starts loading and returns a promise-wrapped model.
- **The model is currently loading:** Returns the promise-wrapped model.

Handling these cases ensures that a model is loaded only once – dramatically improving performance for applications utilizing numerous controllers and scenes. The reader may wonder why this is important – consider a quiz application that presents dozens of models for the user to select from.

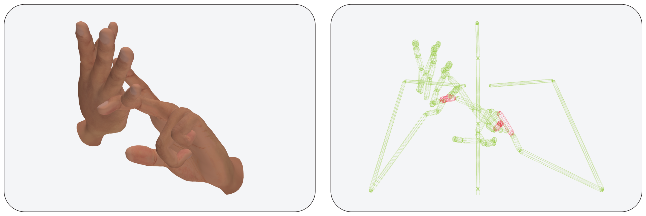
B.2 Key Intervals

The standard approach to transitions in Three.js is to decrease the weight of the outgoing animation and increase the weight of the incoming animation. Weights are the proportion of an animation's influence on the model's final transformation. A transition that starts early will alter a crucial section of the animation, resulting in a significant decrease in articulation accuracy. To rectify this, the implementation introduces the concept of a key interval – a contiguous section of an animation that must be played before a transition. A `KeyInterval` is a class with the following attributes: start time, end time, start frame, and end frame, that is, a `KeyInterval` can be created using seconds or frames. The set of key intervals provided by the developer includes an entry for every letter in the alphabet.

High-precision scrubbing is given an elapsed time to set the exact animation state at that time. In Three.js, each letter has its own `AnimationAction`, i.e., each letter has to be played individually. In addition, during a transition, both actions must be played, and their weights must be set correctly. These factors result in an inability to correctly set the animation state using `mixer.update(delta)`, where delta is the difference between the current and the desired elapsed time. To address this issue, the implementation involves two classes:

- `AnimationTrack:` A data structure containing, for each animation, the following attributes: start time, transition start time, and transition duration. This structure is implemented as a class, with utility methods for accessing animations

[29] https://www.wikipedia.org/wiki/Dependency_injection
[30] https://www.threejs.org
[31] https://www.react.dev
[32] https://en.wikipedia.org/wiki/Singleton_pattern



(a) Mesh **(b) Bones**

Figure 11: Finger collision during an O → N transition.

given a time and checking whether the time falls within a transition.

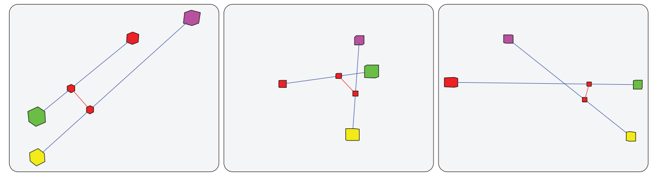
- `AnimationTimer`: A class that consumes `AnimationTrack`, setting the animation state based on an internally managed elapsed time. This class exposes methods for setting, incrementing, and decrementing the elapsed time, as well as methods for playing, pausing, and stopping the animation.

The transition start time (seconds after the key interval end) is calculated by taking the remaining time after the key interval of a letter and multiplying it by one minus the smoothness factor, a value between 0 and 1, representing how soon the animation should begin transitioning after its key interval. The transition lasts until the key interval start of the incoming letter, or the end of the outgoing letter, whichever is first. Developers can provide a smoothness factor in the model configuration.

B.3 Collision Avoidance

Smooth transitioning between letters can cause collisions (see Figure 11). To address this issue, the implementation includes a collision detection and avoidance module that calculates the maximum smoothness factor for a transition that avoids collisions. Collision detection is implemented by 1) Creating 3D vectors between anatomically connected joints (i.e., finger bones), 2) Creating capsules around those vectors to simulate finger thickness, 3) Calculating the shortest distance between all bone vectors (see Figure 12), and 4) Checking whether the distance is smaller than the combined radius of the bone capsules. To simplify collision avoidance, the `AnimationTrack` class was refactored to combine four different actions: `LetterAction`, `MiddleAction`, `TransitionAction`, and `IdleAction`. Each of these actions encapsulates logic for creating a specific type of animation:

- `LetterAction`: Clips an animation to include only the key interval frames.
- `MiddleAction`: Clips an animation slightly before the key interval. This can be inserted between consecutively repeated `LetterAction`, because repeating animations with a single key frame would result in a constant position when played consecutively (consider L → L in HELLO). The key interval for L is a single frame, meaning the `LetterAction` would be a single frame. Without a `MiddleAction`, the consecutive L animations would have no separation, making them appear as one single long animation of the letter L.



(a) View 1 **(b) View 2** **(c) View 3**

Figure 12: Shortest distance between 2 3D vectors as seen from 3 different views.

- `TransitionAction`: Interpolates from the final frame of an outgoing animation to the start frame of an incoming animation. This exposes a static method for generating a collision-free transition. Since interpolation is unnatural, we are converting `TransitionAction` to `BlendedTransitionAction`. `BlendedTransitionAction` will cross fade (increasing incoming animation weight, and decrease outgoing animation weight) using the maximum smoothness that avoids collisions. Interpolation calculates the next position mathematically, which can create synthetic feeling animations. Blending combines two animations, gradually increasing the weight of the incoming animation, while decreasing the weight of the outgoing animation, creating a more natural transition as it uses the real recorded hand movements.
- `IdleAction`: Contains an idle animation inserted at the beginning and end of every `AnimationTrack`.

`TransitionAction` generates a collision-free transition by performing a binary search[33] over smoothness factor between 0 and 1. In case of a collision, it moves the upper bound to the current smoothness factor; in case of no collision, it moves the lower bound to the current smoothness factor. Since 0 → 1 is a continuous range, a tolerance is required – the algorithm stops searching when the change in smoothness factor between search iterations is less than a developer-provided tolerance. The collision avoidance is based on the following heuristic: For a given smoothness factor, a collision will not occur if there exists a greater smoothness factor where there is no collision. This collision avoidance implementation is very slow – it takes approximately 10 seconds for a single transition. To address this run-time bottleneck, a precalculation Node.js[34] script is used to create a dictionary with the maximum smoothness factor for all possible transitions in the alphabet. This dictionary is imported into the application and provides an instant smoothness factor lookup when generating transitions. On an AMD Ryzen 7 3700X 3.6 GHz 8-Core/16 Threads AM4 processor, the script takes approximately 19 minutes to run.

[33]https://en.wikipedia.org/wiki/Binary_search
[34]https://www.nodejs.org

C AuslanSpell User Evaluation

C.1 Words

Table 9: List of words used in during user evaluation (ordered by length).

Free-Text	Multiple-Choice	Distractor Words
smile	noisy	nosey, moist, maybe, missy
yeast	spray	spree, upset, spate, nosey
fresh	honey	hokey, heavy, money, funny
fright	narrow	barrow, marrow, mellow, below
apples	modest	modern, needle, middle, noodle
forever	reject	regret, rugrat, unjust, reached
laughing	complete	compete, contain, campaign, cuddles
happiness	healthy	wealthy, health, healer, happily
personally	transport	transpose, transact, support, supported
coordinated	accessible	accessibility, acceptable, asssesable, acceptance

C.2 Questionnaire

Your experience of the AuslanSpell application

For each of the following statements, please mark one box that best reflects your opinion.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I think that I would like to use this app frequently.	○	○	○	○	○
I found the app unnecessarily complex.	○	○	○	○	○
I thought the app was easy to use.	○	○	○	○	○
I think that I would need the support of a technical person to be able to use this app.	○	○	○	○	○
I found the various functions in the app were well integrated.	○	○	○	○	○
I thought there was too much inconsistency in this app.	○	○	○	○	○
I would imagine that most people would learn to use this app very quickly.	○	○	○	○	○
I found the app very cumbersome to use.	○	○	○	○	○
I felt very confident using the app.	○	○	○	○	○
I need to learn a lot of things before I could get going with this app.	○	○	○	○	○

On a scale of 1-5 how useful did you find...

	1 - Not at all useful	2	3	4	5 - Extremely useful
The ability to swap between left-hand and right-hand signing.	○	○	○	○	○
The ability to swap between front and back view.	○	○	○	○	○
The ability to rotate view.	○	○	○	○	○
The ability to change signing speed.	○	○	○	○	○
The ability to swap between simple and realistic hand animations.	○	○	○	○	○

Do you think the slowest possible speed is:

- ◯ Too slow
- ◯ About right
- ◯ Too fast

Do you think the fastest possible speed is:

- ◯ Too slow
- ◯ About right
- ◯ Too fast

On a scale of 1-5, how realistic did you find the fingerspelling?

- ◯ 1 - not at all realistic
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5 - extremely realistic

On a scale of 1-5, how helpful did you find interacting with the app for developing your own fingerspelling readback skills?

- ◯ 1 - not at all helpful
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5 - extremely helpful

Is there anything else you'd like to tell us about your experience using the app or features you'd like to see included?

About you

What is your gender?

- ◯ Male
- ◯ Female
- ◯ Non-binary / third gender
- ◯ I prefer a different term
- ◯ Prefer not to say

What is your age?

- ◯ 17-20
- ◯ 21-25
- ◯ 26-30
- ◯ 31-40
- ◯ 41-50
- ◯ 51-60
- ◯ 61+

Are you...

- ◯ Left-handed
- ◯ Right-handed
- ◯ Ambidextrous

Which of the following best describes your experience learning languages other than English (LOTE)? If you know multiple LOTEs please answer for your strongest language.

- ◯ I am a native speaker or highly proficient in a LOTE (e.g. completed university major, CEFR C1 or higher).
- ◯ I have intermediate proficiency (e.g. completed VCE, university minor, CEFR B1, B2).
- ◯ I have beginner level proficiency (e.g. studied at primary/ junior secondary school, one or two beginner units at uni, CEFR A2 or below)
- ◯ I have never learnt a LOTE

Which of the following best describes you?

- ◯ I am an undergraduate student
- ◯ I am enrolled in the MAppLing
- ◯ I am a HDR student
- ◯ I am studying a different type of course

In your current or past degrees have you studied linguistics or languages?

- ◯ Yes - major in linguistics
- ◯ Yes - major in a language
- ◯ Yes - minor/ single unit in linguistics
- ◯ Yes - minor/ single unit in a language
- ◯ No study of linguistics or a language

Have you completed any tertiary education courses prior to the course you are enrolled in currently?

- ◯ Yes
- ◯ No

What is the level of the highest tertiary qualification you have completed?

- ◯ TAFE certificate 1-3
- ◯ TAFE Certificate 4 or Diploma
- ◯ Undergraduate degree
- ◯ Postgraduate qualification
- ◯ I have not completed a tertiary course

What is the highest level of schooling you have completed?

- ◯ Year 12
- ◯ Year 11
- ◯ Year 10
- ◯ Year 9 or below

If you are ok with us linking your test to your answers here (we would really like to be able to!) please write your name here.