Promoting Better Deaf/Hearing Communication Through an Improved Interaction Design for Fingerspelling Practice

Rosalee Wolfe[1(✉)], John McDonald[1], Jorge Toro[2], Souad Baowidan[1],
Robyn Moncrief[1], and Jerry Schnepp[3]

[1] DePaul University, Chicago, USA
{wolfe,jmcdonald}@cs.depaul.edu
{sbaowida,rmoncrie}@mail.depaul.edu
[2] Worchester Polytechnic Institute, Worchester, USA
jatoro@wpi.edu
[3] Bowling Green State University, Bowling Green, USA
schnepp@bgsu.edu

Abstract. Fingerspelling is a manual system used by many signers for producing letters of a written alphabet to spell words from a spoken language. It can function as a link between signed and spoken languages. Fingerspelling is a vital skill for ASL/English interpreters, parents and teachers of deaf children as well as providers of deaf social services. Unfortunately fingerspelling reception can be a particularly difficult skill for hearing adults to acquire. One of the contributing factors to this situation is a lack of adequate technology to facilitate self-study. This paper describes new efforts to create a practice tool that more realistically simulates the use of fingerspelling in the real world.

Keywords: Deaf · Deaf accessibility · American sign language · Fingerspelling · Voice input

1 Introduction

Fingerspelling is a manual system used by many signers for producing letters of a written alphabet to spell words from a spoken language [1]. Members of the Deaf[1] community in the United States use fingerspelling for proper nouns such as person and place names, and for spelling loan words from other languages. An additional use of fingerspelling is to convey technical terminology for which there is no generally accepted sign [2].

---

[1] The term Deaf ("capital-D Deaf") refers to the community that uses American Sign Language (ASL) as their preferred language and share a history, cultural norms, beliefs and values in common.

1.1   Need for Fingerspelling Skill

Fingerspelling skill is important not just for members of the Deaf community, but is essential for interpreters of signed language, parents of deaf children [3], teachers of deaf children [4], and providers of deaf social services. Ninety percent of deaf children have hearing parents [5], but when these children experience increased contact with finger-spelling, the result is a significant positive impact on their reading ability [6]. Further, in a university setting, fingerspelling is useful for linking the instructor's lecture with the readings in the assigned text [7]. Additionally, it has long been considered a highly desir-able skill for vocational rehabilitation counselors working with Deaf clients [8].

1.2   Difficulty of Fingerspelling

Unfortunately fingerspelling reception (recognizing fingerspelled words) is particularly difficult for hearing learners [9]. Due to this fact, many teachers of deaf children are not skilled in fingerspelling [10], and rely on interpreters for this critical skill [11]. This is often not possible due to the scarcity of educational interpreters [12], many of whom are underqualified in fingerspelling [13]. Hearing interpreter training students regularly mention fingerspelling reception as a difficult, if not the most difficult, skill to master [14, 15]. Fingerspelling receptive skills are much harder to acquire than fingerspelling production [16]. Even interpreters, who have already graduated from interpreter training programs and have been hired at interpreter agencies, list fingerspelling as one of their top training needs [17].

Why is fingerspelling so hard? The reasons are myriad but can be grouped into two major categories. The first barrier has to do with the nature of fingerspelling itself. It is not formed as a sequence of static letters, but as a smoothly changing movement where the fingers never stop in their transitions from letter to letter [18]. As a result, the letters in a fingerspelled word are rarely, if ever, perfectly produced. Coarticulation plays a major role since letter handshapes are heavily influenced by preceding and succeeding letters. Simply studying the static positions of the manual letters does not facilitate recognition of a word from the smooth flow of the motion envelope [9].

The second barrier to fingerspelling fluency is the paucity of practice opportunities. Textbooks recommend pair practice [19], but a practice partner will most likely be another classmate. Unfortunately, a fellow student will not be able to produce finger-spelling smoothly or at fluent speeds [20]. Further, due to demanding schedules, it is not always possible to schedule face-to-face practice sessions. These barriers can motivate the typical student to seek options for self-study.

2   Options for Practice

Although technology has provided alternatives to paper-based fingerspelling texts [21], each alternative has drawbacks. VHS video recordings of fingerspelling appeared in the 1980s, and the 1990s witnessed the appearance of DVDs designed for fingerspelling practice [22]. For these media, the fingerspelled words were recorded and thus fixed. It was not possible to create new words without incurring the costs of producing a new

recording. Because the videos were recorded at low frame rates, motion blur was also a problem, as was the lack of variation in the presentation order. As students studied the same recording over and over again, it was not clear if they were improving their skills or memorizing the recording.

In the early 2000s, an alternative to fixed, prerecorded media appeared on web sites such as [23]. On these sites, students can use software to view a word as a succession of snapshots, each displaying a single manual letter. The advantage of these sites is extensibility. The site software can rearrange the snapshots in any order and thus produce new words without incurring any additional cost. However, the static nature of the snapshots is a problem. There is no connective movement between the letters. This limits its utility as a practice tool since most of fingerspelling is comprised of the motion between the letters, not the letters themselves [9].

A third alternative is 3D animation technology, which promises the extensibility for new word formation while producing smoothly flowing motion, but it poses some challenges as well. The lack of physicality in 3D animation complicates the situation. Unless prevented, the thumb and fingers will pass through each other when transitioning between closed handshapes such as M, N, T, S and A in ASL, as demonstrated in Fig. 1. This requires a system to prevent finger collisions. Additionally, 3D animation requires simulating the flexible webbing between the thumb and index finger and mimicking the complex behavior of the base of the thumb [24].
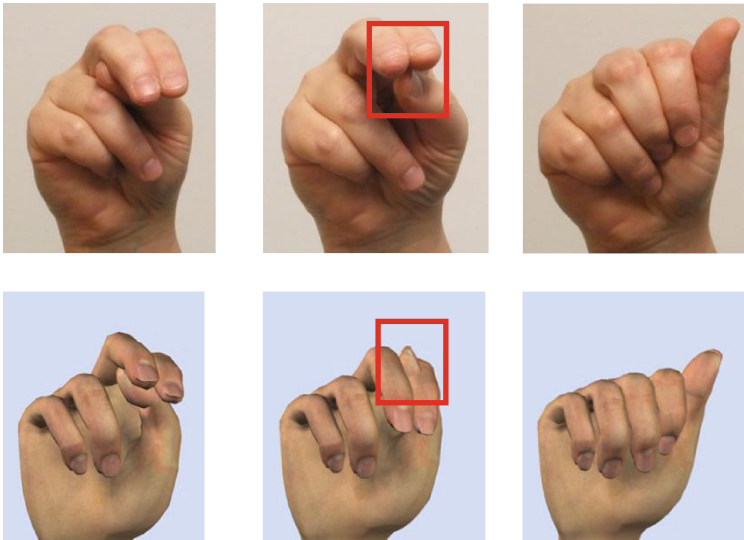

Fig. 1. Comparing physical motion with a naive animation for the transition from N to A

These complexities entail large computational requirements, and require significant resources to render fingerspelling in real time. For this reason, some previous efforts sacrificed realism to gain real-time speeds by using a simplified 3D model that did not accurately portray a human hand and/or did not prevent collisions [25, 26]. Others sacrificed

real-time responsiveness to maintain the realism of the model [27]. To address this, researchers have developed a method to pre-render and organize the transitions in such a way that the software can form new words that display natural motion while maintaining real time responsiveness [28].

However, accurate and realistic fingerspelling movement is only the first step. Practice software needs to offer appropriate user interaction to enhance the learning process. When practicing, students need to be able to respond to questions and receive feedback on their answers.

3   Previous Interaction Designs

In all of the previously-discussed technologies that offer interactive feedback, students view a fingerspelled word and supply their answer by either selecting from a list of choices or by typing. Neither of these interaction options accurately simulates real-life situations where fingerspelling reception skills are needed. Consider the following scenarios:

- When an interpreter is facilitating a conversation between a Deaf and hearing person, the interpreter will be voicing the signing produced by the Deaf conversant. The voicing, of course, will require that the interpreter recognize any fingerspelled words.
- When parents or teachers view a child's fingerspelling, their response will be signed and/or fingerspelled in return.

While it is true that a skilled interpreter will make use of context to eliminate possible choices, this is very different from choosing an answer from a pre-created list of options. Interpreters and other hearing people who converse with members of the Deaf community need to recognize a word in order to voice it, but rarely is there a need to vocally reproduce the word letter-for-letter.

Insisting on text input not only requires users to recognize the word, but forces them to spell it correctly. Thus current software is testing not only a user's receptive capabilities, but their spelling abilities as well.[2] Keyboard input also introduces the possibility of typographical errors [29]. These are not errors in fingerspelling reception but conventional software cannot make this distinction. Further, typing can be slow, especially on mobile devices [30].

Teachers of ASL and interpreter trainers are aware of the shortcomings of evaluating fingerspelling receptive skills through English orthography. An examination of national certification procedures [31] shows that no testing procedure requires applicants to write out words but instead assesses fingerspelling receptive skills through voicing or sign production.

---

[2] It is true that interpreters will need skills in English orthography to *produce* fingerspelled words when signing the voiced discourse of a hearing person, but the skill we are focusing on here is not fingerspelling *production*, but fingerspelling *reception*, which is the area of greater need.

4   Exploring Alternatives for a More Natural Interaction Style

Most modern digital devices provide for speech input, permitting users to voice an answer rather than type it. Researchers [32] have noted that speech is preferable for typing character strings requiring more than a few keystrokes. Further, speech has the potential for generating text more quickly than keyboard typing [33]. Speech input has even more potential benefit when using hand-held devices where keyboards are small [34]. A voice alternative for fingerspelling practice has several potential benefits:

- *More focus on fingerspelling*. The necessity of typing an answer after viewing a fingerspelled word requires a user to shift visual attention away from fingerspelling. The user's mental effort is divided between typing a correct answer and attempting to recognize fingerspelling. Voice input utilizes a separate channel, and the shift of mental modes is much shorter.
- *A shorter distance between user and answer*. In the Keystroke-Level Model used to model complexity of human/machine interactions [35], a vocal operator of *speak* is modeled at 150 ms/syllable, but a manual operator of *Type_in* is modeled at 280 ms/character. Given that each English syllable contains at least one and typically more letters, there should be a shorter distance between a user and the answer when a response is spoken.
- *Closer modeling of real-world usage.* Using speech more closely resembles the real-life scenarios where fingerspelling receptive skills are required. Further, the interaction would more closely match the testing procedure of the national certification agency and could potentially provide better preparation for the examination.

4.1   Design Considerations for Speech Input

Despite the potential benefits, the feasibility of using voice input hinges on the accuracy of the automatic speech recognition engine. In addition to environment [36], major factors affecting accuracy include

- Single speaker/multiple speakers. Speech from a single speaker is easier to recognize because most parametric representations of speech are sensitive to the characteristics of a particular speaker.
- Isolated words/continuous speech. Speech containing isolated words is much easier to recognize than continuous speech because word boundaries can be hard to identify.
- Vocabulary size. Large vocabularies are more likely to contain multiple entries that are difficult to disambiguate.

Because the majority of people will be practicing fingerspelling on a personal device such as a phone, tablet or laptop, they will likely have established a user profile for speech recognition. This facilitates the use of single-speaker recognition strategies. Additionally, a response will consist of a single word, thus the recognition engine will not be forced to identify word boundaries. Finally, the vocabulary size is a single word which means that there will be no ambiguous entries in the vocabulary. Thus we have

what appears to be the perfect confluence of single speaker, isolated word input and highly constrained vocabulary.

However we found that spoken words which are similar to the fingerspelled word were also being recognized as correct. For example, words such as "rendition" and "perdition" were sometimes accepted as matching the fingerspelled word "condition". A vocabulary of a single word opened the door to an unacceptably large number of false positives.

4.2 Evaluating Vocabulary Configurations

To determine the optimal vocabulary size, we set up a software test bed that could simulate errors on the part of the user. The test bed exercised a commercial speech recognition engine via a simple program that displayed a word, and prompted a user to say it. After the user said the word, the program displayed a new word to pronounce. No other feedback was given.

Unbeknownst to the user, sometimes the word displayed was not the word that the speech recognition engine was expecting but was instead a similar word. Two words were deemed similar when they had the same length and matched in initial and final letters. We chose this definition based on coarticulation studies of fingerspelling [9, 18, 37], which indicate that the initial and final letters of a fingerspelled word are the most distinct and most easily recognized. These studies imply that words deemed similar by this definition can be easily confused when reading fingerspelling. This definition was use to simulate the type of false positive discussed in the last section.

Five testers (three male, two female) used the test bed for two sessions each. Each session consisted of 40 trials. Each trial involved viewing and speaking a single word, for a total of 400 trials from the five testers. Half of the trials contained a simulated error. Since the trials were randomized and no feedback given, the testers did not know which words were considered errors.

The outcome is summarized as a confusion matrix [38] in Table 1. There was an unacceptably high number of Type I errors, which corresponds to the test bed accepting a simulated error as being correct. Thus the strategy of using a single-word vocabulary, which is appropriate in a conventional interactive voice response (IVR) system or voice menu, will not be satisfactory for this application, which requires greater specificity. The approach of configuring the recognition engine to accept only one word would not be satisfactory.

Table 1. Confusion matrix for a vocabulary size of one

	Correct word	Simulated error
Accepted as correct	192 (hit)	83 (Type I error)
Rejected as error	8 (Type II error)	117 (correct rejection)

Given the assumption that users will have already trained their device for voice input, a second alternative would be to use the entire dictation vocabulary of the device's speech

recognition engine. The test bed was modified to use the large dictation vocabulary instead of the single-word vocabulary, and the same testers used the new version for a total of 400 trials. The confusion matrix for this second alternative is shown in Table 2. For this configuration, the number of Type I errors has dropped to zero, but the number of Type II errors, which correspond to rejecting a correctly-spoken word, has risen to the point where this configuration is also not an acceptable alternative.

Table 2. Confusion matrix for a large vocabulary

	Correct word	Simulated error
Accepted as correct	56 (hit)	0 (Type I error)
Rejected as error	144 (Type II error)	200 (correct rejection)

The third alternative would be to find a vocabulary size that is somewhere between the two extremes. To evaluate this approach, the test bed was again reconfigured to use progressively larger vocabularies of sizes {1, 6, 11, 21, 41, 81}. Words picked for the vocabularies were again matched with the target word for length, initial letter and final letter. Figures 2 and 3 contain graphs of the summary statistics for each of the six vocabulary sizes. Figure 2 is a plot of the sensitivity (hit rate), and specificity (correct rejection rate) of the confusion matrices and shows the inverse relationship between sensitivity and specificity. The two curves cross near a vocabulary size of 11. The accuracy curve in Fig. 3 clearly exhibits peak accuracy around a vocabulary size of 11 (the target word and 10 distractors).
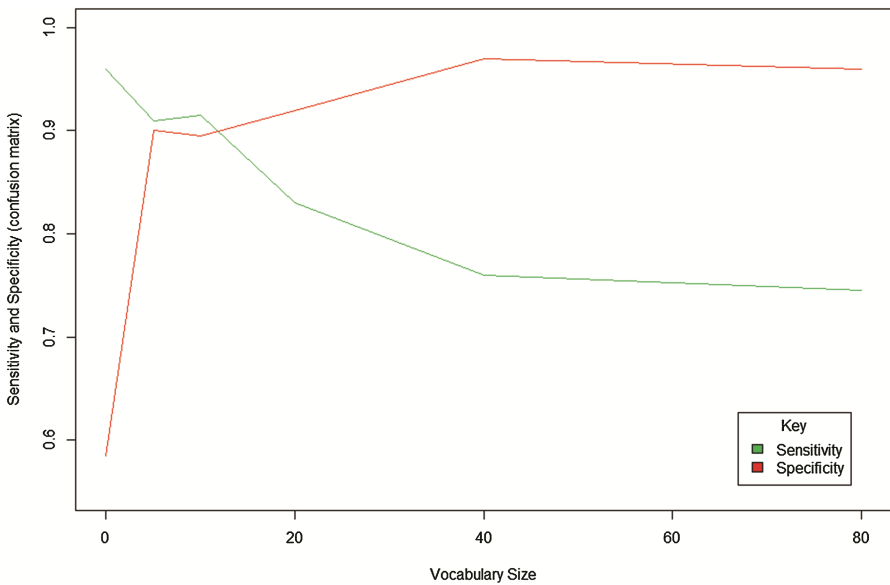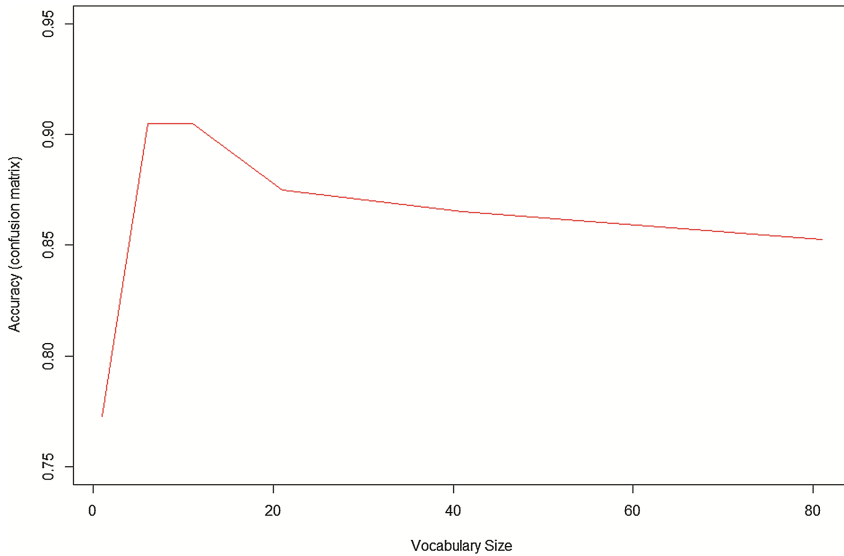

Fig. 2. Sensitivity and specificity plots

Fig. 3. Accuracy plot

These results informed our design decisions for configuring the speech engine vocabulary. The vocabulary for each fingerspelling trial contains:

- The fingerspelled word
- 10 distractor words chosen at random from among a list of words similar to the fingerspelled word.

To maintain quick response times, we pre-computed a list of similar words for over 100,000 entries from the CMU Pronouncing Dictionary [39]. This did increase the software's memory footprint, but since the dictionary size is still dwarfed by the size of the fingerspelling video, overall size was not significantly impacted.

5   Results

We modified the previous version of Fingerspelling Tutor [28], which only had a multiple choice interface, to offer a fill-in-the blank mode where students can either type or speak an answer. We did not add a voice option to the extant multiple choice mode, because its input is already very direct, being only a single tap or click. Also, the multiple choice mode is not one that accurately simulates real-world usage, but instead acts as an intermediate step for acquiring receptive skills [40]. Further, voice input could also introduce speech recognition errors into a response format that is intentionally constrained to help beginners avoid errors.

Speech input is more appropriate for the fill-in-the-blank mode because it more closely resembles real-life usage. However, there are situations when typed input is more appropriate such as:

- in environments with high ambient noise,
- in situations where there is no acoustical privacy,
- or when user decides that keyboard input is preferable.

Thus the modified interface includes both the option to type or to speak the response. Since Fingerspelling Tutor supports both physical and on-screen keyboards, we chose to follow an interaction style that has the microphone attached to the textbox as demonstrated in Fig. 4. This both reduces the distance that a user has to move the mouse in order to activate the microphone and makes it more visible in the interface.

Fig. 4. Screen shot of voice interface

6   Future Work

We are in the process of conducting usability tests to compare user performance and preference of the newly-configured voice interface with the conventional keyboard interface. In addition, we are looking to expand Fingerspelling Tutor for use in signed languages other than ASL.

References

1. Battison, R.: Lexical Borrowing in American Sign Language. Linstok Press, Silver Spring (1978)
2. Padden, C.: The acquisition of fingerspelling by deaf children. In: Siple, P., Fischer, S. (eds.) Theoretical Issues in Sign Language Research, pp. 191–210. University of Chicago, Chicago (1991)
3. Caulderon, R.: Parental involvement in deaf children's education programs as a predictor of child's language, early reading, and social-emotional development. J. Deaf Stud. Deaf Educ. **5**(2), 140–155 (2000)

4. Strong, M., Prinz, P.: A study of the relationship between American sign language and English literacy. J. Deaf Stud. Deaf Educ. **2**(1), 37–46 (1997)
5. Meyers, J., Bartee, J.: Improvements in the signing skills of hearing parents of deaf children. Am. Ann. Deaf **137**(3), 257–260 (1992)
6. Ramsey, C., Padden, C.: Natives and newcomers: gaining access to literacy in a classroom for deaf children. Anthropol. Educ. Q. **29**(1), 5–24 (1998)
7. Napier, J.: University interpreting: linguistic issues for consideration. J. Deaf Stud. Deaf Educ. **7**(4), 281–301 (2002)
8. Quigley, S.: The vocational rehabilitation of deaf people. SRS-72-25037, Washington, DC (1972)
9. Wilcox, S.: The Phonetics of Fingerspelling. John Benjamins Publishing, Amsterdam (1992)
10. Grushkin, D.: Lexidactylophobia: the irrational fear of fingerspelling. Am. Ann. Deaf **143**(5), 404–415 (1998)
11. Jones, T., Ewing, K.: An analysis of teacher preparation in deaf education: programs approved by the council on education of the deaf. Am. Ann. Deaf **147**(5), 71–78 (2002)
12. Hitch, M.: Educational interpreters: certified or uncertified. J. Law Educ. **34**, 161–165 (2005)
13. Schick, B., Williams, K., Kupermintz, H.: Look who's being left behind: educational interpreters and access to education for deaf and hard-of-hearing students. J. Deaf Stud. Deaf Educ. **11**(1), 3–20 (2006)
14. McKee, R., McKee, D.: What's so hard about learning ASL?: students' & teachers' perceptions. Sign Lang. Stud. **75**(1), 129–157 (1992)
15. Shaffer, L., Watson, W.: Peer mentoring: what is that? In: Maroney, E. (ed.) Proceedings of the 15th National Convention Conference of Interpreter Trainers (CIT), CIT, Still shining after 25 years, Washington, DC, pp. 77–92 (2004)
16. Shipgood, L., Pring, T.: The difficulties of learning fingerspelling: an experimental investigation with hearing adult learners. Int. J. Lang. Commun. Disord. **30**(4), 401–416 (1995)
17. Hernandez, R.A.: New ideas of teaching and learning fingerspelling. In: McIntire, M. (ed.) New Dimensions in Interpreter Education: Curriculum and Instruction, pp. 121–124. Registry of Interpreters for the Deaf, Silver Spring (1987)
18. Jerde, T., Soechting, J., Flanders, M.: Coarticulation in fluent fingerspelling. J. Neurosci. **23**(6), 2383–2393 (2003)
19. Smith, C., Lentz, E.M., Mikos, K.: Signing Naturally Student Workbook Level 1. Dawn Sign Press, San Diego (1988)
20. Reed, C., Delhorne, L., Durlach, N., Fischer, S.: A study of the tactual and visual reception of fingerspelling. J. Speech Lang. Hear. Res. **33**, 786–797 (1990)
21. Guillory, L.: Expressive and Receptive Fingerspelling for Hearing Adults. Claitors Publishing Division, Baton Rouge (1966)
22. Jaklic, A., Vodopivec, D., Komac, V.: Learning sign language through multimedia. In: International Conference on Multimedia Computing and Systems, Washington, DC, pp. 282–285 (1995)
23. Vicars, B.: Dr. Bill Vicars' American sign language (ASL) fingerspelling practice site. http://asl.ms. Accessed 2005
24. McDonald, J., Alkoby, K., Carter, R., Christopher, J., Davidson, M.J., Ethridge, D., Furst, J., Hinkle, D., Lancaster, G., Smallwood, L., Ougouag-Tiouririne, N., Toro, J., Xu, S., Wolfe, R.: An improved articulated model of the human hand. Vis. Comput. **17**(3), 158–166 (2001)
25. Su, A.: VRML-based representations of ASL fingerspelling on the world-wide web. In: The Third International ACM SIGCAPH Conference on Assistive Technologies, Marina del Rey, CA (1998)

26. Dickson, S.: Advanced animation in mathematica. Math. J. **15**(2) (2013)
27. Adamo-Villani, N., Beni, G.: Automated finger spelling by highly realistic 3D animation. Br. J. Educ. Technol. **35**(3), 345–362 (2004)
28. Toro, J.A., McDonald, J.C., Wolfe, R.: Fostering better deaf/hearing communication through a novel mobile app for fingerspelling. In: Miesenberger, K., Fels, D., Archambault, D., Peňáz, P., Zagler, W. (eds.) ICCHP 2014, Part II. LNCS, vol. 8548, pp. 559–564. Springer, Heidelberg (2014)
29. Kukich, K.: Techniques for automatically correcting words in text. ACM Comput. Surv. **24**(4), 377–439 (1992)
30. Parhi, P., Karlson, A., Bederson, B.: Target size study for one-handed thumb use on small touchscreen devices. In: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 203–210 (2006)
31. Registry of interpreters for the deaf: NAD-RID National Interpreter Certification. http://rid.org/userfiles/File/NIC2CandidateHBFFeb2014_1.pdf. Accessed 2014
32. Hauptmann, A., Rudnicky, A.: A comparison of speech and typed input. In: Proceedings of the Speech and Natural Language Workshop, Stroudsburg, PA, pp. 219–224 (1990)
33. Rebman Jr., C., Aiken, M., Cegielski, C.: Speech recognition in the human-computer interface. Inf. Manag. **40**(6), 509–519 (2003)
34. Cherubini, M., Anguera, X., Oliver, N., De Oliveira, R.: Text versus speech: a comparison of tagging input modalities for camera phones. In: Proceedings of the 11th International Conference on Human-Computer Interaction (MobileHCI 2009), Bonn, Germany, pp. 1–10 (2009)
35. Kieras, D.: A guide to GOMS model usability evaluation using COMSL and GLEAN3, Ann Arbor (1999)
36. Peacocke, R., Graf, D.: An introduction to speech and speaker recognition. Computer **23**(8), 26–33 (1990)
37. Geer, L., Keane, J.: Exploring factors that contribute to successful fingerspelling comprehension. In: Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, pp. 68–69 (2014)
38. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
39. Carnegie Mellon University: The CMU pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict
40. Jamrozik, D., Davidson, M., McDonald, J., Wolfe, R.: Teaching students to decipher fingerspelling through context: a new pedagogical approach. In: Proceedings of the 17th National Convention Conference of Interpreter Trainers, San Antonio, TX, pp. 35–47 (2010)