Introduction

There is a substantial community of people who are profoundly deaf (2 million in the US in 1989 [2]), of whom perhaps 360,000 speak American Sign Language (ASL) [1] and are referred to as Deaf. ASL is a sophisticated language very different from English in numerous important ways. Deafness complicates many otherwise simple features of everyday life. Only 20-40% of speech is comprehensible via lipreading [20]. Not hearing spoken English makes it difficult to learn to read English, which is in any event a second language for most Deaf students. Thus, an average Deaf US high-school graduate reads poorly [11] and likely lacks sufficient fluency to follow TV closed-captions (in English).

We propose to launch an interdisciplinary project to build devices which translate between written English and American Sign Language (ASL). While some ASL translation projects exist at other universities, none has had any notable successes because ASL has a number of unique features as a language — for example, the natural representation is as video, rather than characters — and no project has involved a sufficiently broad team of investigators. As a result, no existing project has taken even the most basic steps to support machine translation: building very large datasets and aligning them. We have assembled a team consisting of experts in computer animation, computer vision, computational natural language, human computer interfaces, ASL linguistics and ASL. The core activity of our project will involve building very large datasets, including aligned datasets of closed captioning and ASL subtitles, as well as ASL conversation both scripted and unscripted. These datasets will support building our translation devices and they will also provide new insight into ASL linguistics, and into the relationships between gesture, emotion, language and meaning. Our datasets will give UIUC a dominant position in ASL machine translation.

We see our work resulting in broad impact in four important areas. First, such translators would provide better access to auditory information (by being coupled to existing speech recognizers). Second, our evaluation work will result in better understanding of translation needs. Third, we will produce a translation toolkit, in the tradition of HCI research. Finally, we will lay the groundwork for consumer devices, the most important of which will take a closed caption stream associated with a television signal, produce an ASL subtitle, and composite that with the television signal, so that Deaf viewers at home can see television subtitled in their native language without requiring input from broadcasters.

Related Work

Video-based telecommunication and computer technology has exploded in the Deaf Community as consumers place a high value on technologies that support sign language interactions [10]. Existing attempts to render ASL look "robotic" to many deaf individuals because the sign stream is interrupted with hands-down pauses, and it is missing critical grammatical information usually carried through facial expression and body posture. Nonetheless, there are a number of useful signing avatars. Vcom3D's SignAvatar is crude, but still useful. For example, Vcom3D claims in promotional material that their SignAvatar product increases comprehension of classroom stories from 17% to 67%. This general result — rough animation still being useful — is supported by work on mathematics instruction [4] and fingerspelling [3, 9]; and by commercial work on a device called an *icommunicator* (from interactive solutions, www.myicommunicator.com), which produces a sequence of isolated ASL signs in English word order (see also the discussion in [5]).

Two major university projects have looked at machine translation of ASL. One, at the University of Pennsylvania, has produced some papers and ideas but no systems and no datasets (e.g. [22, 12]); the other, at Boston University [16], has produced a markup toolkit and a very small set of videos of scripted signing sessions. A similar small dataset is being produced at Purdue [15]. None of these datasets is large or rich enough to support statistical translation research. A joint project of Radboud University in Nijmegen, Stockholm University and City University, London, has built small corpora for Swedish Sign Language, British Sign Language and Sign Language of the Netherlands (http://www.let.ru.nl/sign-lang/echo/). There has been sustained work on generating British Sign Language (BSL), including a system that renders signed subtitles for television programs using BSL signs in English word order [6, 17] and a form of signed phrasebook for use at post office counters and in weather forecasts [7, 8, 13, 14, 21, 19]. (Another useful phrasebook, for

diagnostic interviews, is described in [18].)

Research Plan

Our project aims to build machine translation systems between English and ASL. We will build a large, aligned dataset of English and ASL. We can then translate English phrases into ASL by identifying streams of video corresponding to particular chunks of English, then translate whole sentences by cutting and pasting phrases. In broad form, this is the standard strategy for statistical machine translation; successfully implementing it for ASL will require innovation in video technologies, language technologies, evaluation procedures and datasets. Building datasets – a crucial first step — is the primary focus of this proposal.

Language Technologies: The core need is an **aligned bitext**, i.e. a body of English text aligned with sections of the ASL that mean the same thing. With an aligned bitext, translation becomes a question of phrase matching and cut-and-paste. However, there are challenges to obtaining an aligned bitext. We will deal with linguistic difficulties by marking up and transforming English to an intermediate form (ESIGN) that more closely mimics ASL syntax and identifies phenomena that require care in translation, e.g. classifiers and pronoun references. Because even large datasets contain only a sampling of possible phrases, we will enrich our dataset by building automatic paraphrases of the English text, thus increasing the chance of finding a match for an incoming sentence.

Video Technologies: There are three core technologies: building representations of video that can be matched, building animations by compositing (splicing together) multiple videos, and matching video to video. We will use feature based representations to align video to text, using statistical models of the appearance of individual ASL signs. We will then build rough output animations by assembling sequences cut from many different videos. Clean animations can then be created by matching rough animations to a large pool of video. We have extensive experience matching video and building animations by cut-and-paste, and the core of the system can exploit established technologies from computer vision.

Evaluation: We will apply both formative and summative evaluation. For formative evaluation, bilingual ASL consultants will transcribe and score automatic translations. We will then perform qualitative failure analysis to identify putative improvements to the ESIGN markup. For summative evaluation, we will compare the use of translated ASL to the use of text captions for experiencing television programming, and test how well our system might function in a system offering bi-directional translation between ASL and spoken/written English.

Datasets

We will collect a variety of types of datasets.

Bitexts: As an initial data set, we will use the numerous children's movies with signed subtitles *and* closed captions, available commercially. These offer the benefit of clean photographic conditions, as well as a clear and careful linguistic style. Models derived from this data will then form the starting point for analysis of captioned video with less-constrained signing (e.g. material from the TV program Deaf Mosaic) and/or covering a wider range of discourse topics. A translation team will (a) mark up video indicating various important ASL features (exact alignment; unusual uses of pronouns; etc.) so as to provide reference datasets for evaluation and (b) prepare a signed translation of some existing closed-caption videos (yielding aligned data).

Observing discourses: Because conversational discourse is typically not polished or scripted, translating it into ASL may reveal algorithmic challenges and, most importantly, will allow us to test our algorithms for translating conversational discourse into ASL. We will capture conversational discourse among hearing users at service-oriented locations (e.g. a library reference desk). We will enlist the services of experts fluent in both ASL and English who will review the transcriptions and will then be video taped signing the discourse. These videos will then result in aligned versions of conversational discourse as opposed to scripted text. We will also collect speeches, such as those recorded at conferences, which were interpreted and/or captioned because some of the audience or presenters were Deaf.

Capturing natural conversation: We will collect natural conversation in ASL, using a variety of strategies. First, we will collect video of chat shows for Deaf people. Second, we will encourage people to talk about family photographs, a strategy that has been successful in other contexts.

Project Context and PI's

Leverage: The proposed project builds on a unique confluence of strengths at UIUC. Our project team has strong activities in computer vision, computer graphics, computational natural language understanding, human computer interaction, ASL linguistics and Deaf community interactions. No other effort in this area has involved so strong and so broad a team; we do not believe that any other institution currently possesses the complete set of skills covered by this team at this level of quality.

Sustainability: The need for ASL machine translation is clear, and the Americans with Disabilities Act would compel many organisations to adopt technologies as they become available. The primary determinant of success in machine translation is the size and richness of the dataset. Because ASL lacks a widely-accepted writing system, there are currently only small corpora. We can exploit this situation to create a competitive advantage for UIUC, because a university group that builds and maintains a significant corpus will dominate ASL translation for the foreseeable future. This dominance will extend to ASL linguistics, because the existing small ASL corpora are inadequate for quantitative rsearch, e.g. deriving frequency characteristics of the language (crucial for psycholinguistics research). Most researchers in ASL linguistics could benefit from having a substantial corpus from which to retrieve data. The primary expenditures for our project will revolve around collecting, administering and disseminating datasets. These datasets will long outlive the CRI funding and the dominance they create will attract other funds in future.

David Forsyth is Professor of Computer Science at UIUC and at UC Berkeley (on leave). He co-authored the standard graduate textbook on computer vision and is also a recognised expert on data-driven animation. He has served on program committees for all major international conferences on computer vision and computer graphics. He has received two best paper awards in computer vision. **Brian Bailey** is Assistant Professor of Computer Science and an expert in human-computer interaction. His research includes measuring effects of interruption and measuring mental workload through the use of pupil size. His multidisciplinary efforts have been recognized with affiliate academic appointments in the Graduate School of Library and Information Science and Aviation Psychology. **Margaret Fleck** is Research Associate Professor of Computer Science at UIUC, and is well known for work on computer vision, computational linguistics, and capturing stories about personal photo collections. **Karrie Karahalios** is Assistant Professor of Computer Science. Her work focuses on the interaction between people and the social cues and signals they perceive and transmit in networked electronic spaces. The goal is to create interfaces that enable users to perceive conversational patterns that are present, but not obvious, in traditional communication interfaces. **David Quinto-Pozos** is Assistant Professor in Speech and Hearing, and works primarily on American Sign Language (ASL) and Mexican Sign Language (LSM). His research has addressed linguistic phenomena that result from contact between users (both bilinguals and monolinguals) of those two languages. He has also worked on tactile ASL as it is used by Deaf-Blind individuals in the U.S. **Dan Roth** is an Associate Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign and the Beckman Institute of Advanced Science and Technology (UIUC). He is a Willett Faculty Scholar of the College of Engineering and was a fellow of the UIUC Center of Advanced Studies. He is well known for his research achievements in theoretical and experimental machine learning and natural language processing and his leadership in these areas. **Jenny Singleton** is Associate Professor in the Department of Educational Psychology. She is a national expert on deafness and sign language research, particularly natural sign language discourse in the classroom and technologies to support deaf children's literacy development. She was appointed by the Governor to serve on the Illinois Deaf and Hard of Hearing Commission, whose mission is to remove barriers faced by deaf and hard of hearing citizens in the state.

Budget

- **Year 1**
    1. Personnel costs $55, 000
    2. Disk $ 3, 000
    3. Translation/Language Consultant Services $ 20, 000
    4. Subject fees $ 7, 000
    5. Kick-off workshop $ 5, 000
    6. ASL Translation and Linguistics Seminar costs $ 6, 000
- **Year 2**
    1. Personnel costs $55, 000
    2. Disk $ 3, 000
    3. Translation/Language Consultant Services $ 20, 000
    4. Subject fees $ 7, 000
    5. ASL Translation and Linguistics Seminar costs $ 6, 000
- **Year 3**
    1. Personnel costs $55, 000
    2. Disk $ 3, 000
    3. Translation/Language Consultant Services $ 20, 000
    4. Subject fees $ 7, 000
    5. Dissemination workshop $ 5, 000
    6. ASL Translation and Linguistics Seminar costs $ 6, 000

1 References Cited

[1] Demographics. Technical report, Gallaudet Research Institute.

[2] National institute on deafness and communication disorders: A report of the task force on the national strategic research plan. *Federal Register*, 57, 1989.

[3] N. Adamo-Villani and G. Beni. Automated fingerspelling by highly realistic 3d animation. *British Journal of Educational Technology*, 35:345–362, 2004.

[4] N. Adamo-Villani, J. Doublestein, and Z. Martin. The mathsigner: an interactive learning tool for american sign language k-3 mathematics. In *IEEE Proceedings of IV04 8th International Conference on Information Visualization*, pages 713–716, 2004.

[5] N. C. R. Association. Comparison of cart to alternative notetaking methodologies. Technical report, National Court Reporter's Association, 2004.

[6] J. Bangham, S. Cox, R. Elliott, J. Glauert, I. Marshall, S. Rankov, and M. Wells. Virtual signing: Capture, animation, storage and transmission - an overview of the visicast project. In *IEE Colloquium on Speech and Language Processing for the Disabled and Elderly*, pages 6/1–6/7, 2000.

[7] S. Cox, M. Lincoln, M. Nakisa, M. Wells, M. Tutt, and S. Abbott. The development and evaluation of a speech to sign translation system to assist transactions. *Int. Journal of Human Computer Interaction*, 16(2):141–161, 2003.

[8] S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, and S. Abbott. Tessa, a system to aid communication with deaf people. In *Proc. ASSETS 2002, Fifth International ACM SIGCAPH Conference on Assistive Technologies*, pages 205–212, 2002.

[9] M. J. Davidson, K. Alkoby, E. Sedgwick, R. Carter, J. Christopher, B. Craft, J. Furst, D. Hinkle, B. Konie, G. Lancaster, S. Luecking, A. Morris, J. McDonald, N. Tomuro, J. Toro, and R. Wolfe. Improved hand animation for american sign language. In *Technology and Persons with Disabilities Conference*, 2001.

[10] J. Harkin and M. Bakke. *Oxford Handbook of Deaf Studies, Language, and Education*, chapter Technologies for communication: Status and trends, pages 406–419. Oxford University Press, 2003.

[11] J. Holt. Stanford achievement test, 8'th ed.: reading comprehension subgroup results. *Am. Ann. Deaf*, 138:172–175, 1994.

[12] M. Huenerfauth. A survey and critique of american sign language natural language generation and machine translation systems. Technical report, University of Pennsylvania, 2003. Technical Report MS-CIS-03-32, Computer and Information Science.

[13] R. Kennaway. Synthetic animation of deaf signing. In *International Gesture Workshop 2001*, 2001.

[14] R. Kennaway. Experience with, and requirements for, a gesture description language for synthetic animation. In *5th International Workshop on Gesture and Sign Language based Humman-Computer Interaction*, 2003.

[15] A. M. Martinez, R. B. Wilbur, R. Shay, and A. C. Kak. Purdue rvl-slll asl database for authomatic recognition of amercian sign language. Proceedings of IEEE International Conference on Multimodal Interfaces, 2002.

[16] C. Neidle, S. Sclaroff, and V. Athitsos. Signstream: A tool for linguistic and computer vision research on visual-gestural language data. *Behavior Research Methods, Instruments, and Computers*, 33(3):311–320, 2001.

[17] F. Pezeshkpour, I. Marshall, R. Elliott, and J. Bangham. Developing of a legible deaf signing virtual human. In *IEEE Multimedia Systems '99 (IEEE ICMCS '99)*, 1999.

[18] A. G. Steinberg, D. S. Lipton, E. A. Eckhardt, M. Goldstein, and V. J. Sullivan. The diagnostic interview schedule for deaf patients on interactive video: A preliminary investigation. *Am J Psychiatry*, 155:1603–1604, 1998.

[19] M. Verlinden, C. Tijsseling, and H. Frowein. A signing avatar on the www. In *International Gesture Workshop*, 2001.

[20] R. Waldstein and A. Boothroyd. Speechreading supplemented by single channel and multichannel tactile displays of voice fundamental frequency. *J. Speech Hear. Res.*, 38(690-705), 1995.

[21] A. Wray, S. Cox, M. Lincoln, and J. Tryggvason. A formulaic approach to translation at the post office: Reading the signs. *Language and Communication*, 24(1):59–75, 2004.

[22] L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, and M. Palmer. Machine translation system from english to american sign language. In *Association for Machine Translation in the Americas*, 2000.