JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-5 E-ISSN: 2347-2693

Asynchronous Multimodal Text Entry Using Speech and Gesture Keyboards

V.Maniraj[1], E.Manohari[2*]

[1,2] *Department of Computer Science, A.V.V.M Sri Pushpam College, Poondi, Thanjavur.*
maniraj_vee@yahoo.co.in; rohiny.mano@gmail.com

www.ijcseonline.org

Received: May/02/2015 Revised: May/10//2015 Accepted: May/24/2015 Published: May/30/ 2015

***Abstract*—** We propose plummeting errors in text entry by uniting speech and gesture keyboard input. We describe a combine model that combines gratitude conorders in an asynchronous and flexible manner. We composed speech and gesture data of manipulators entering both short email sentences and web search queries. By amalgamation gratitude conorders from both modalities, word error rate was abridged by 53% comparative for email sentences and 29% comparative for web searches. For email exclamations with speech errors, we examined providing gesture keyboard alterations of only the mistaken words. Deprived of the user openly indicating the improper words, our model was able to decrease the word error rate by 44% relative.

***Keywords*—** Mobile text entry, Multimodal interfaces

I. INTRODUCTION *(HEADING 1)*

It is difficult to correct speech gratitude errors using speech alone. Therefore speech interfaces frequently provide manipulators with a subordinate response modality. In this paper we propose using a capacitive touch-screen gesture keyboard as a subordinate modality for speech. Capacitive touch-screens are attractive since of their high-quality incessant touch-signals and their increasing ubiquity on mobile phones. Previously sim used a capacitive touch-screen keyboard to agree manipulators to response facts such as word boundaries or the first letter of the envisioned word. Sim found that touch response abridged the decoding time and the error rate on a 5k wsj task.

Here we find that we can considerably decrease error rates by uniting speech with a touch-screen gesture keyboard. A gesture keyboard enables manipulators to quickly write disagreements by swiping a finger over the touch-screen keyboard. For example, to write the word "speech" the manipulators pushes down on the s key and slides to the p, e, c and h keys before lifting up to widespread the gesture (see figure 1). The system then performs a pattern match to find the word whose outline on the keyboard most resembles the gestured shape. Gesture keyboards have been commercialized as shapewriter, swype, t9 trace and flext9.

We combine the speech and gesture keyboard modalities using our recently proposed combine model. This model supports two key features: asynchronicity and *patterned correction*. Asynchronicity means that manipulators do not need to synchronize speech and gestures. Informal testing revealed that manipulators have difficulty speaking and gesturing simultaneously. This is comparable to sim's finding that a user had difficulty speaking though touching the first letter of each word. In addition, the asynchronous combine model also permits manipulators to achieve *post hoc* error correction in which gratitude conorders from a primary modality are later merged with the gratitude conorders from a subordinate modality.

The additional key feature is patterned correction. Our combine model can combine gratitude conorders from a modality that only

Figure 1: *the ideal outline of the word "speech" on a gesture keyboard. The starting point is indicated with a dot.*

Conservative a subset of the disagreements conservative by the other modality. This enables patterned correction in which the user only needs to response the mistaken disagreements in the additional modality to correct the errors in the first modality. This is done deprived of openly specifying the mistaken words. For example, say the user spoke "the cat sat" and the system documented "the *bat* sat". Using patterned correction the user simply gestures "cat" and the system then attempts to locate and replace the mistaken word.

II. ASYNCHRONOUS MULTIMODAL TEXT ENTRY

2.1 Speech Recognition

We used the cmu sphinx speech recognizer, training a usenglish acoustic model on 211 hours of wsj data. We trained cross-word triphones with a 3-state left-to-right hmm topology. We used a 39-dimensional feature vector with 13 melincidence cepstral coefficients, deltas and delta deltas. Our model had 8000 tied-states with 16 incessant gaussians per state and diagonal comodification matrices. We used the cmu pronunciation dictionary (39 phones plus silence). Audio was chronicled at 16khz. We attained cepstral mean normalization based on a prior window of audio. The recognizer was modified to each user's speech using 40 sentences. We modified the model means using maximum likelihood linear regression with 7 regression classes. We used the pocketsphinx decoder and tuned it to near real-time recognition.

Our email language model was trained on text from a usenet quantity (1.8b words), a blog quantity (387m words), and four months of twitter letters (109m words). We composed the twitter data using the free streaming api which provide admittance to 5% of all tweets. We trained our language model only on sentences that were comparable to short email sentences using cross-entropy alteration selection [4]. In this recently proposed method, each verdict is scored by the alteration in perword cross-entropy between an in-dominion language model and a background model. We did this separately for the usenet, blog and twitter datasets. Our in-dominion trigram language model was trained on sentences drawn from the w3c quantity and the non-spam letters in the trec 2006–7 spam track. We only used sentences with six or fewer words. Our background mockups were trained on a comparable quantity of training data as our indominion model but used sentences from usenet, blog or twitter. For each of these three sources we chose the cross-entropy alteration threshold that optimized presentation on held out w3c and trec data. We built a combination model from our three language mockups using linear interpolation with combination weights optimized on the held out data. Our combination model had 43m n-grams. All mockups used interpolated modified kneser-ney smoothing with no count cutoffs and a 64k vocabulary.

2.2. Gesture keyboard recognition

We will now describe the gesture gratitude procedure, which is an adaptation of the standard procedure. If the longitudinal length of the trace is less than a threshold (38 pixels or 6 mm on a 4th group ipod touch), then the system assumes the user envisioned to touch a single key rather than articulate a touch-screen gesture. Unlike preceding work [2] which only returned the nearest key on the keyboard when the user tapped a single key, we provide additional letter hypotheses to the combine model by computing a likelihood p_k for each key k:

$$P_k = \exp\left(-\frac{d_k^2}{\sigma_k^2}\right), \qquad (1)$$

Where d_k is the euclidean distance between the first touchpoint of the user's trace and the key k, and σ_k is a modification estimate. Motivated by preceding empirical work on modeling on-screen keyboard touch-errors [5] we undertake the distance between a touch-point and the center of an envisioned key is approximately normal.
If the trace is not a tap then we need to recognize the $\begin{bmatrix} x & y & 1 \end{bmatrix}_t$ user's incessant gesture as a word.

First describe as a point in homogeneous organizes on a two-dimensional cartesian plane. Then let the orders $u = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $v = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be two ordered orders of n equidistant points. The sequence u represent the sequence of two-dimensional touch-points the user has traced on the touchscreen. The sequence v represents the outline of the ideal traced out word on the keyboard layout (see figure 1). This outline is produced by serially linking the centers of the consistent letter keys for a word. Both orders are resampled to have the same quantity of n example points. Next we describe $\mathbf{t}$ as an affine transform, also in homogeneous coordinates:

$$\mathbf{t} = \begin{bmatrix} s & 0 & dx \\ 0 & s & dy \\ 0 & 0 & 1 \end{bmatrix}. \qquad (2)$$

Here dx and dy are the parallel and vertical translation components, and $s \in [0,1]$ is a scale factor. Dx and dy are set to the respective one-dimensional detachments between the centroids of u and v , and s is set to the maximum ratio of the diagonals the bounding boxes of u and v . Given u and v we calculate a likelihood p_w of a word w as:

$$P_w = \exp\left(-\left[\left(\frac{x_s^2}{\sigma_s^2}\right) + \left(\frac{x_l^2}{\sigma_l^2}\right)\right]\right) \qquad (3)$$

Where σ_s and σ_l are modification estimates, and

$$x_s = \frac{1}{n}\sum_i (||\mathbf{T}\mathbf{u}_i - \mathbf{v}_i||), \qquad (4)$$

And

$$x_l = \frac{1}{n}\sum_i (||\mathbf{u}_i - \mathbf{v}_i||). \qquad (5)$$

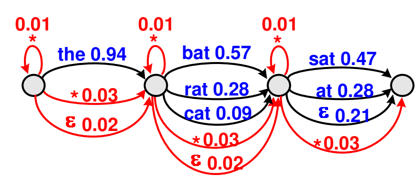
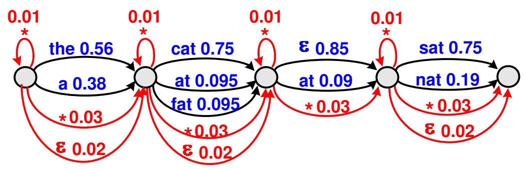


Figure 2: *an example of a speech gratitude* wcn *that has been softened. The additional edges are in dotted red. The * symbol represents wildcard transitions. This figure is modified from.*



\

Figure 3: An example of gesture keyboard recognition WCN With additional dotted red edges. This figure is modified from [3].

X_s is a identical score between u and v which is scale and translation invariant. X_l is a comparable score, but is reliant on on where u and v are positioned on the on-screen keyboard. It is conceivable to make the gratitude scale-translation invariant by setting $x_l = 0$.

We tuned the quantity of example points and the limitations σ_k, σ_s and σ_l to optimal values on a expansion set. The gesture keyboard recognizer used the same 64k terminology as the speech recognizer. Each gesture is documented independently and produces a set of disagreements and likelihoods under the model. We then concept a lattice that attaches each word with every word in the subsequent set. This lattice is then rescored with the speech recognizer's trigram language model. From the rescored lattice we concept a word misperception grid (wcn).

2.3. Combine model

To combine the speech and gesture modalities we use a combine model that we have previously recognized [3]. This model is capable of uniting output from numerous recognizers asynchronously. The model was originally recognized for mishmash of manifold speech signals for a speech-only correction boundary [7]. Here we prove how this model can also be used to fuse speech and gesture keyboard gratitude results. What shadows is a high-level overview of how the model works with illustrative figures modified from the innovative paper [3].

The model operates on wcnS. The innovative wcns are unstiffened by adding three additional changeovers to every cluster. First, an epsilon transition is additional that enables the present cluster to go to the next cluster deprived of generating a word. Second, a wildcard self-loop enables the present cluster to generate any word though remaining in the same cluster. Third, a wildcardnext transition permits a cluster to generate any word and proceed to the next cluster. The likelihood of each of these additional changeovers can be varied between the wcns being cooperative and can also be varied between discomparable clusters within a wcn.

The first wcn is attained from the speech recognizer. The additional wcn is attained from the gesture keyboard recognizer. Figures 2 and 3 show example wcns from the speech and gesture keyboard recognizers.

The model works by searching for a joint path finished the unstiffened wcnS. We explain the search using the token passing model [8]. A token in our model tracks three pieces of information. First, the position in each of the wcnS. Second, the accumulated log probability. Third, the preceding few disagreements of language-model context.

A search is initiated with a token that starts in the first cluster in both wcnS. A token is finished when it reaches the last cluster in both wcnS. At each step of the search, we select a token from the pool of unfinished tokens. From the designated token's position in each wcn, we calculate all conceivable moves that generate a single word (whichever a real word or a wildcard word). We then take the cross-product between candidate moves in each wcn. We consider a mishmash of moves valid only if it obeys two rules. First, at least one of the moves must generate a real word (i.e. Not every wcn can use a wildcard). Second, if manifold wcns generate real words, these disagreements must match.

Every move is assessed a likelihood under a language model. The combine model uses the same language model as both recognizers. Since large wcns have a vast quantity of conceivable combinations, an admissible search is intractable. We apply pruning beams to focus the search on only the most talented possibilities. See the innovative paper [3] for more details.

The free limitations of the combine model, such as the wildcard and epsilon transitions, were tuned on speech and gesture keyboard expansion data chronicled by the authors. As we will describe shortly, the model was tested in three distinct scenarios: amalgamation full speech and gesture results, amalgamation patterned corrections, and amalgamation preventive corrections. Discomparable sets of parameter values were tuned for each scenario.

III. GRATITUDE EXPERIMENTS

A. Mobile email

We tested the entry of brief mobile email sentences using speech and a gesture keyboard. We designated sentences of length 1–6 disagreements typed by enron employees on their blackberry devices [9]. We composed the data for speech and gesture keyboard from separate pools of participants. Tryouts were done offline.

3.1.1. Data collection

Four american english speakers spoke the email sentences. Their audio was chronicled at 16 khz using a plantronics voyager pro wireless microphone. The other four contributors used a gesture keyboard to write the email sentences. These contributors used a 4th group ipod touch with a capacitive touchscreen. The gesture keyboard on-screen display measured 49.9 mm × 22.4 mm (320 × 144 pixels). The dimensions of each different key measured 5.0 × 7.5 mm (32 × 48 pixels). Each sample point conservative from the capacitive touch-screen was displayed as a red dot to provide trace feedback to the participant. The boundary is shown in figure 4. In total we composed 148 paired sentences with each contestant responsibility between 32 and 41 sentences. The sentences had an out-of-terminology rate of 0.7% with respect to our 64k vocabulary.

3.1.2. Results

We first tested the combine model on widespread exclamations with gesture traces for every word in each utterance. The conorders are shown in table 1. Overall speech gratitude (sr) was the least accurate modality with 27% wer. The gesture keyboard (gk) was much better at 14%wer and the scale-translation invariant account of the gesture keyboard (igk) attained about the same (14% wer). We find it interesting that the scale-translation invariant account of the gesture keyboard had comparable presentation as the location-reliant on version. We conjecture this is since position facts can both aid and hinder recognition.

Recognizer(s)	Combo Model	Wer	Ser	Oracle Wer
Sr	-	27.2%	54.7%	8.6%
Gk	-	14.2%	44.6%	8.1%
Igk	-	14.1%	41.9%	8.2%
Sr+gk	Merge	6.6%	25.0%	3.3%
Sr+igk	Merge	6.6%	25.0%	3.5%
Sr+gk	Cnc	10.3%	32.4%	1.0%
Sr+igk	Cnc	7.7%	27.0%	1.2%

Table 1: conorders for a single modality and for uniting modalities in the mobile email domain.

Depending on how carefully the user is gesturing on the keyboard. We also computed the wcn oracle wer which is the path finished the wcn with the lowest error rate. As expected the oracle wer was considerably lower for all modalities.

Uniting the modalities caused in a 53% comparative discount in wer compared to just using the gesture keyboard, the most accurate single modality. Our best result from the combine model was at a lower wer than even the best oracle wer of any of the modalities. This establishes the advantage of balancing modalities which recognize response in discomparable ways. We also compared our combine model to misperception grid mishmash (cnc) [10] as implemented by srilm [11]. Our combine model model providing superior gains to cnc in wer. However, cnc did have a better oracle wer. The oracle wer for cnc is better since nothing gets eliminated throughout the mishmash though our combine model performs pruning throughout its search.

As shown by the verdict error rate (ser) in table 1, uniting both modalities conorders in three out of four sentences being documented totally correct. With speech alone less than half of all sentences were documented totally correct.

Our combine model is also capable of patterned error correction. In patterned correction, one modality only receives gratitude response for a time-ordered subset of the disagreements of the full verdict in the other modality. Our idea is to enable manipulators to see incorrectly documented disagreements in the speech modality and trace only the disagreements that are in error. This is done deprived of providing any position facts about where the improper disagreements are located in the speech result. Table 2 shows the conorders on using patterned correction on the 81 sentences which had at least one improper word in the speech modality. As shown in the table, patterned correction considerably abridged wer by 44% relative.

Last, we also tested preventive error correction. In preventive error correction, manipulators speak an utterance and simultaneously, or shortly thereafter, response a single word using a gesture keyboard. If manipulators can predict the most likely word to be mis-

Recognizer(s)	Combo model	WER	SER
sr	-	48.5%	100.0%
sr+gk	merge	28.8%	79.0%
sr+igk	merge	27.1%	81.5%

Table 2: conorders when amalgamation patterned gesture keyboard alterations with speech gratitude conorders for email sentences where the speech recognizer made at least one word error.

Recognizer(s)	Combo model	Wer	Ser
Sr	-	27.2%	54.7%
Sr+gk	Merge	26.6%	58.8%
Sr+igk	Merge	26.8%	58.8%

Table 3: conorders when amalgamation a one word preventive correction with speech gratitude conorders in the mobile email domain.

Documented by the speech recognizer, a successful combine may preclude the error. To test the viability of this idea, we first recruited two contributors who did not take part in any of the other data assemblage tasks reported in this paper. None of the contributors had any speech gratitude experience. These contributors were shown email sentences and instructed to underline a single word in each verdict which they thought was the most likely to be misrecognized. We then designated these different disagreements from the gesture keyboard data and merged the subsequent wcns in contradiction of the widespread sentences from the speech recognizer. Table 3 shows that this form of preventive error correction had little impact.

B. Mobile web search

We also tested our combine model in the mobile web search domain. We used the system and the data from our past work.

We composed search enquiries from seven contributors recruited from the university campus. Each contestant was shown a search query on the screen and asked to response it using whichever speech or the gesture keyboard though concurrently walking around. The three american english contributors who spoke the search enquiries had their audio chronicled at 16khz using a jabra m5390 wireless microphone. The other four contributors used a gesture keyboard to write the search queries. These contributors used the same ipod touch device and boundary as designated for mobile email. In total, we composed 398 paired search enquiries with each contestant responsibility between 80 and 120 queries.

The subsequent wer for speech gratitude (sr), gesture keyboard (gk), and a mishmash of the two are shown in table 4. Uniting the modalities caused in a 29% comparative discount in wer in comparison to just using the gesture keyboard.

Recognizer(s)	Combo model	Wer	Ser
Sr	-	34.2%	49.3%
Gk	-	15.3%	31.4%
Sr+gk	Merge	10.8%	24.6%

Table 4: *conorders when using a single modality and when uniting modalities in the mobile search domain.*

IV. CONCLUSIONS

We have shown how speech and gesture keyboard response can be cooperative to decrease errors in text entry. We designated a combine model that cooperative gratitude conorders in an asynchronous and flexible manner. We composed speech and gesture data from manipulators entering both email sentences and web search queries. By amalgamation gratitude conorders from both modalities, word error rate was abridged by 53% comparative for emails and 29% comparative for web searches. For email exclamations with speech errors, we examined providing gesture keyboard alterations of only the mistaken words. Deprived of the user needing to openly indicate the improper words, our model was able to decrease word error by 44% relative. Our conorders show that the gesture keyboard is a talented balancing response modality to speech.

REFERENCES

[1] Arif, A.S. ; Dept. of Comput. Sci. & Eng., York Univ., Toronto, ON, Canada, “A survey on mobile text entry handedness: How do users input text on handheld devices while nomadic?”, Published in: Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on Date of Conference: 27-29 Dec. 2012 Page(s): 1 – 6.

[2] Sandnes, F.E. ; Dept. of Comput. Sci., Oslo Univ. Coll., Norway ; Thorkildssen, H.W. ; Arvei, A. ; Buverad, J.O., “Techniques for fast and easy mobile text-entry with three-keys”, Published in: System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on Date of Conference: 5-8 Jan. 2004.

[3] Poirier, F. ; Univ. de Bretagne-Sud, Vannes ; Sad, H.H., “A Platform for Mobile Text Entry Methods Evaluation”, Published in: Advances in Computer-Human Interaction, 2008 First International Conference on Date of Conference: 10-15 Feb. 2008 Page(s): 183 – 188.

[4] Starner, T., “Keyboards redux: fast mobile text entry”, Published in: Pervasive Computing, IEEE (Volume:3 , Issue: 3) Page(s): 97 – 101 ISSN : 1536-1268.

[5] Tantug, A.C. ; Istanbul Tech. Univ., Istanbul, Turkey, “A probabilistic mobile text entry system for agglutinative languages”, Published in: Consumer Electronics, IEEE Transactions on (Volume:56 , Issue: 2) Page(s): 1018 – 1024 ISSN : 0098-3063 DOI: 10.1109/TCE.2010.5506034 Date of Publication : May 2010.

[6] Clawson, J. ; Coll. of Comput., Georgia Inst. of Technol., Atlanta, GA, USA ; Lyons, K. ; Starner, T. ; Clarkson, E., “The impacts of limited visual feedback on mobile text entry for the Twiddler and mini-QWERTY keyboards”, Published in:Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on Date of Conference: 18-21 Oct. 2005 Page(s): 170 – 177.

[7] Ma Yang ; Sch. of Comput. Eng., Nanyang Technol. Univ., Singapore, Singapore ; Leedham, G. ; Higgins, C. ; Htwe, S.M., “On-line recognition of Pitman shorthand for fast mobile text entry”, Published in: Information Technology and Applications, 2005. ICITA 2005. Third International Conference on (Volume:1) Date of Conference: 4-7 July 2005 Page(s): 686 - 691 vol.1.

[8] Ha, R.W.K. ; Dept. of Electr. & Comput. Eng., Waterloo Univ., Ont., Canada ; Pin-Han Ho ; Shen, X.S., “SIMKEYS: an efficient approach in text entry for mobile communications”, Published in: Consumer Communications and Networking Conference, 2004. CCNC 2004. First IEEE Date of Conference: 5-8 Jan. 2004 Page(s): 687 - 689.

[9] Misra, S. ; Covenant Univ., Ota, Nigeria ; Banubakode, S.M. ; Dhawale, C.A., “Novel user interface for text entry on touch screen mobile device for visually impaired users”, Published in: Computer & Information Technology (GSCIT), 2014 Global Summit on Date of Conference: 14-16 June 2014 Page(s): 1 – 5.

[10] Heng-Li Yang ; Deptartment of Manage. Inf. Syst., Nat. Chengchi Univ., Taipei, Taiwan ; Ren-Xiang Lin, “A Chinese predictive text entry method for mobile devices”, Published in: Machine Learning and Cybernetics (ICMLC), 2014 International Conference on (Volume:1) Date of Conference: 13-16 July 2014 Page(s): 309 – 314.

[11] Prince, M.E.H. ; Dept. of Comput. Sci. & Eng., Chittagong Univ. of Eng. & Technol. (CUET), Chittagong, Bangladesh ; Hossain, G. ; Dewan, A.A. ; Debnath, P., "An audible Bangla text-entry method in Mobile phones with intelligent keypad", Published in: Computers and Information Technology, 2009. ICCIT '09. 12th International Conference on Date of Conference: 21-23 Dec. 2009 Page(s): 279 – 284.

[12] McDermott-Wells, P. ; Graduate Sch. of Comput. & Inf. Sci., Nova Southeastern Univ., Dania, FL, "Evaluation of Three Stylus-Based Text Entry Methods on a Pocket PCtm Mobile Device", Published in: SoutheastCon, 2006. Proceedings of the IEEE Date of Conference: March 31 2005-April 2 2005 Page(s): 228 – 234.

[13] Sharma, R. ; Adv. Interface Technol. Inc., State College, PA, USA ; Yeasin, M. ; Krahnstoever, N. ; Rauschert, I., "Speech-gesture driven multimodal interfaces for crisis management", Published in: Proceedings of the IEEE (Volume:91 , Issue: 9) Page(s): 1327 – 1354 Date of Publication : Sept. 2003.

[14] Oviatt, S. ; Dept. of Comput. Sci., Oregon Health & Sci. Univ., Beaverton, OR, USA, "User-centered modeling and evaluation of multimodal interfaces", Published in: Proceedings of the IEEE (Volume:91 , Issue: 9) Page(s): 1457 - 1468 Date of Publication : Sept. 2003.

[15] Pacheco, B.A. ; Comput. & Inf. Fac., Mackenzie Presbyterian Univ., São Paulo, Brazil ; Souza-Concilio, I.A., "Multimodal interfaces: An enactive approach", Published in: Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on Date of Conference: 2-5 Dec. 2013 Page(s): 51 – 58.

[16] Czyzewski, A. ; Gdansk Univ. of Technol., Gdansk, Poland ; Dalka, P. ; Kosikowski, L. ; Kunka, B., "Multimodal human-computer interfaces based on advanced video and audio analysis", Published in: Human System Interaction (HSI), 2013 The 6th International Conference on Date of Conference: 6-8 June 2013 Page(s): 18 – 25.

[17] Karpov, A. ; SPIIRAS, St. Petersburg Inst. for Inf. & Autom., St. Petersburg, Russia ; Carbini, S. ; Ronzhin, A. ; Viallet, J.E., "Comparison of two different similar speech and gestures multimodal interfaces", Published in: Signal Processing Conference, 2008 16th European Date of Conference: 25-29 Aug. 2008 Page(s): 1 – 5.

[18] Czyzewski, A. ; Multimedia Syst. Dept., Gdansk Univ. of Technol., Gdansk, Poland, "New applications of multimodal human-computer interfaces", Published in: New Trends in Audio & Video and Signal Processing: Algorithms, Architectures, Arrangements, and Applications (NTAV/SPA), 2012 Joint Conference Date of Conference: 27-29 Sept. 2012 Page(s): 19 – 24.

[19] Jing Liu ; Dept. of Comput., Macquarie Univ., North Ryde, VIC, Australia ; Kavakli, M., "A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games", Published in: Multimedia and Expo (ICME), 2010 IEEE International Conference on Date of Conference: 19-23 July 2010 Page(s): 1564 – 1569.

[20] Toth, B. ; Budapest Univ. of Technol. & Econ., Budapest ; Nemeth, G., "Challenges of creating multimodal interfaces on mobile devices", Published in: ELMAR, 2007 Date of Conference: 12-14 Sept. 2007 Page(s): 171 – 174.

